Courant Forschungszentrum

Evolution des Sozialverhaltens

Experimental and theoretical investigations of the emergence and sustenance of prosocial behavior in groups

Dissertation zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultäten der
Georg-August-Universität zu Göttingen

vorgelegt von

Katrin Fehl

aus Bad Soden-Salmünster

Göttingen, Juni 2011

Referent: Prof. Dr. Dirk Semmann

Koreferentin: Prof. Dr. Margarete Boos

Tag der mündlichen Prüfung: 11.07.2011

CONTENTS

GENERAL INTRODUCTION

Cooperation is abundant

We, the human species, build houses together, organize ourselves in unions and governments, and as inhabitants of a village we build our own autarkic renewable energy plant – just to name a view out of numerous other examples of cooperation. In fact, we are not the only cooperative species. Throughout the animal kingdom we find various forms of cooperation (for reviews, see e.g. Hammerstein 2003; Pennisi 2009). There is cooperation between such simple organisms as single-cell organisms (e.g. Santorelli *et al.* 2008). Viruses cooperate together to infect cells (Turner & Chao 1999). Eusocial insects such as wasps, ants, and honeybees show very high levels of cooperation and possess a detailed division of labor (Wilson & Hölldobler 2005; Reeve & Hölldobler 2007). Furthermore, pairs of fish inspect predators together (Milinski *et al.* 1990). Social non-human primates support each other in coalitions, share food and groom each other (Barrett *et al.* 1999; de Waal & Brosnan 2006; Cheney *et al.* 2010; Schino & Aureli 2010).

As diverse as cooperative behavior is in nature, so are the fields studying cooperation, which highlights the general interest and broad scope needed to understand the ultimate and proximate mechanisms underlying cooperative behavior. Fields investigating cooperation include anthropology, biology, ecology, sociology, psychology, political sciences, mathematics, and even physics. While studying cooperation one fascinating question arises: why is human cooperation so fundamentally different from that of other species? As reviewed by van Schaik and Kappeler (2006; also see Melis & Semmann 2010) human cooperation stands out, because (i) we much more often cooperate within groups (whereas most animals only engage in dyadic cooperation); (ii) we engage in extremely high-risk cooperation (e.g. sophisticated forms of warfare); (iii) we tend to cooperate more with non-kin than other primates; (iv) we are willing to punish non-cooperating individuals even when this is costly (especially striking is the use of third-party punishment); (v) we rely on the use of reputation to establish cooperation much more than non-human primates; and (vi) we trade goods and services using a token-based exchange. To reach these high levels of cooperation we use language, we show remarkably emotional responses associated with cooperative behavior, and we use culturally transmitted social norms.

Cooperation is an evolutionary puzzle

In evolutionary biology the focus lies in general on behavior, i.e. the particular processes within the psychological "black box" are usually set aside, and the behavioral consequences in terms of fitness benefits or losses are investigated. Hence, *cooperation* is defined as the "outcome of an interaction (or repeated interactions) where all participants on average increase their direct fitness" (Bshary 2010, p. 215). This means that even though one individual could engage in cooperative behavior, the overall outcome of the interaction with another individual does not have to be cooperation. The cooperative act itself is a behavior that provides on average a direct survival benefit to other individuals, but at the same time incurs on average survival costs to the actor him- or herself (costs are somewhat lower than the produced benefits). Logically one would not expect to find cooperative behavior in nature, because of its negative selection pressures. Thus, in light of Darwin's theory of evolution the abundance of cooperation constitutes a puzzle. In fact it has been termed "one of the great puzzles in evolutionary biology" (Rand *et al.* 2010, p. 624) and "one of the most fundamental challenges to date" (Santos *et al.* 2006c, p. 51). In more general terms, the problem that cooperation faces lies in the threat of exploitation of mainly cooperative individuals (also termed cooperators, contributors, or helpers) by mainly uncooperative individuals (also termed defectors or non-contributors). There are two ways in which exploitation can take place. First, in an on-going exchange of helping one can simply refrain from returning the next favor. Within a dyadic context, this problem is captured by the iterated *prisoner's dilemma* paradigm. Second, in a group endeavor one can enjoy the benefits of a public good produced by the group without contributing to its provision and maintenance. Such behavior is called free-riding and can be investigated by the *public goods game* (for more details see below).

Definition of common concepts

Before moving on to mechanisms leading to cooperation some common concepts shall be defined (especially as they are not always clearly defined in the associated research fields; for attempts, see Lehmann & Keller 2006; West *et al.* 2007b; Bshary & Bergmüller 2008).

One of the most important and influential concepts was introduced by Hamilton (1964): the *inclusive fitness*. Inclusive fitness is composed of *direct fitness*, i.e. fitness gained through a behavior that affects the production of own offspring, and *indirect fitness*, i.e. fitness gained through a behavior that affects the reproductive success of related individuals.

Social behavior can be defined on the basis of inclusive fitness[1] (definitions are adopted from Bshary & Bergmüller 2008; Bshary 2010). Social acts can either increase or decrease the direct fitness of the actor and affect the social counterpart (or recipient) by increasing or decreasing his or her direct fitness (see summary in Tab. 1). The act is understood as helping or *prosocial behavior* if it increases the direct fitness of the recipient. This can occur in three forms: (i) in *mutualistic behavior* both parties have an immediate direct fitness benefit and both gain a higher benefit from acting together than from acting alone (here actor and recipient are from the same species; whereas *mutualism* refers to such behavior between species); (ii) in *cooperative behavior* the actor first places a costly investment (and decrease his or her immediate direct fitness), which on average increases the direct fitness of the recipient. The actor can only receive direct fitness benefits in the long-run, for instance via reciprocity. Cooperative behavior is at least in part selected because of the benefits towards the recipient (West *et al.* 2007a); (iii) in *altruistic behavior*[2] the direct fitness of the recipient is on average increased by the actor's costly investment, which decreases the actor's direct fitness. This behavior is under positive selection only if the actor obtains indirect fitness benefits in the long-run. Apart from prosocial behavior there is *selfish behavior*, which increases the actor's direct fitness, but decreases the recipient's direct fitness. *Spiteful behavior* decreases the direct fitness of both the actor and the recipient.

Furthermore, cooperative behavior needs to be distinguished from two other forms of behavior. Even though these forms can also produce a beneficial outcome, cheating is narrowed down. First, in *by-product mutualism* (Brown 1983) both the actor and the recipient benefit. In contrast to mutualistic behavior, the actor has immediate benefits from his or her behavior independent of the recipient's behavior. Thus, the behavior is self-serving and the beneficial outcome for the recipient is merely a by-product. Second, in *pseudo-reciprocity* (Connor 1986) the investment of the actor enables the recipient to perform a self-serving behavior that benefits the actor as a by-product in return. As the actor's first needs to place the investment to receive the by-product benefits, there exists a time delay between the two acts (this contrasts by-product mutualism where both benefits are immediately received).

---

[1] Note that using the concept of inclusive fitness to define cooperative behavior is not meant to imply that direct and indirect fitness benefits of an individual are only affected by selection pressures on the individual itself (individual level), but that also pressures at higher levels of selection can operate.

[2] This "biological altruism" is to be distinguished from "psychological altruism", which does not rely on cost-benefit analyses, but is a prosocial behavior that is defined by its underlying psychological mechanisms, like perspective taking or empathy.

Table 1 Forms of social behavior based on direct fitness consequences (definitions are adopted from Bshary & Bergmüller 2008; Bshary 2010). If the recipient yields positive direct fitness benefits this behavior is termed helping or prosocial behavior.

		Direct fitness effect on the recipient	
		+	-
Direct fitness effect on the actor	+	mutualistic behaviour[1] / cooperative behaviour[2]	selfish behavior
	-	altruistic behaviour[3]	spiteful behavior

[1] The actor and the recipient have an immediate direct fitness benefit and both gain a higher benefit from acting together than from acting alone.

[2] The actor needs to place an investment, which is costly in the beginning, and receives long-term direct benefits.

[3] The actor has a costly investment and benefits the recipient, but to be positively selected the behavior needs to yield (long-term) indirect benefits.

Mechanisms to solve the puzzle

With the constant risk of being suckered cooperative behavior should not evolve under natural selection. Despite this challenge, in the last decades much effort has been devoted to understanding the mechanisms behind the evolution of cooperation. It was demonstrated, that cooperative behavior can evolve under the condition that the inclusive fitness of the actor is increased relative to the average fitness in the population (which is already implied by the above definitions). The following mechanisms cause a higher inclusive fitness by either increasing the direct or indirect fitness of the actor.

Several theoretical mechanisms leading to the evolution of cooperation were reviewed by Nowak (2006b; but see criticism in West *et al.* 2007a). The theory of *kin selection* (Hamilton 1964) focuses on cooperation among closely related individuals, i.e. through cooperating with kin, individuals can increase their indirect fitness. *Network reciprocity* can sustain cooperation via the impact of spatial structures (Nowak & May 1992; Lieberman *et al.* 2005; Ohtsuki *et al.* 2006; but see Hauert & Doebeli 2004). Due to the spatial distribution of individuals (e.g. lattices, cycles, or scale-free networks) only certain individuals or neighbors interact with each other, which then can promote cooperation. Mechanisms of *group/multi-level selection* support the evolution of cooperation (Wilson 1975, 1983; Sober & Wilson 1998; Traulsen & Nowak 2006). Here the selection forces do not only act on the individual level but also on the group level (a group of cooperators might be more successful than a group of defectors). The theories of *direct reciprocity* (Trivers 1971; Axelrod & Hamilton

1981; Axelrod 1984; Nowak & Sigmund 1992, 1993) and *indirect reciprocity* (Nowak & Sigmund 1998, 2005; Ohtsuki & Iwasa 2006) rely on dyadic and triadic long-term interactions to foster cooperation. In direct reciprocity recipients return favors received directly to the actor based on "you scratch my back and I'll scratch yours". In indirect reciprocity the actor provides a benefit to the recipient, but the beneficial return-act is carried out by a third party. In reciprocal interactions, behavior is, for example, influenced by conditional strategies (Wedekind & Milinski 1996; Milinski & Wedekind 1998), reputational effects (Milinski *et al.* 2002; Semmann *et al.* 2005), and rewards and punishments (Fehr & Gächter 2002; Sefton *et al.* 2007; Rand *et al.* 2009a). Apart from mechanisms of natural selection, cultural selection is a strong force in the evolution of human behavior (Richerson *et al.* 2003). All the named mechanisms do not necessarily exclude each other and needless to say interactions between them can arise.

That the puzzle of cooperation has not been completely resolved yet becomes clear with the currently hotly debated value of the concept of kin selection to explain cooperation in eusocial insects (e.g. Nowak *et al.* 2010; Abbot *et al.* 2011; Herre & Wcislo 2011). Hence, despite the theoretical advances much more work, especially empirical and experimental results supporting theoretical assumptions, is needed to fully understand the mechanisms leading to cooperation.

Evolutionary game theory

Evolutionary game theory (Maynard Smith & Price 1973; Maynard Smith 1982) provides a framework to study the evolution of cooperation. The theory looks at (behavioral) phenotypes and how these are distributed in a given population due to individuals' fitness (Nowak 2006a). But fitness is not an absolute parameter, it depends on what other kinds of phenotypes are present in the population, i.e. fitness is frequency dependent. To study cooperation, a population of individuals with different (and usually fixed) strategies is considered. A strategy is the individual's phenotype or in more general terms the strategy specifies what the individual will do in a given situation (Maynard Smith 1982). Within the population individuals interact in evolutionary games with one another; usually at random. Each interaction results in a certain payoff for an individual with a given strategy (cf. the payoff matrix of the two strategies in the prisoner's dilemma in Box 1 of Chapter 2, p. 36). An individual's payoff does not only depend on its own strategy, but also on the strategy of the opponent. Payoffs are understood as the fitness of individuals and fitness is positively correlated with reproductive success. Hence, strategies that do well in evolutionary games reproduce faster and outcompete other strategies that do less well. During frequency-

dependent selection dynamics of two strategies the following outcomes are possible (Nowak 2006a): (i) one strategy dominates the other, meaning that eventually the whole population adopts the dominate strategy; (ii) the strategies are bistable (here, the outcome depends on the initial conditions, leading either to an unstable equilibrium or the convergence to one or the other strategy); (iii) both strategies coexist; and (iv) both strategies are neutral to each other, so that selection will not change the composition of strategies within the population. Another important concept of evolutionary game theory is the *evolutionary stable strategy*. A strategy is thought to be evolutionary stable, if it yields the highest payoff of all strategies within the population and a mutant strategy cannot invade the population.

Evolutionary game theory offers models like the *prisoner's dilemma* (Rapoport & Chammah 1965; Axelrod 1984) and the *public goods game* (Hardin 1968; Ledyard 1995) to illustrate the conflict between selfish and selfless behavior. The prisoner's dilemma is set aside here, as concise descriptions will be provided in Chapters 1 and 2 (see p. 20 and p. 36).

The classic public goods game is made up of four players (e.g. Fehr & Gächter 2002; Milinski 2006). Each receives the same amount of money with the opportunity to contribute this money into a common public good. Whatever amount entered the public good will be doubled, divided by the total number of players and evenly paid to everyone. Thereby, it does not matter whether a player contributed or not. Now, the group does best if all players contribute into the public good. However, a rational player should never contribute at all, because each money unit paid into the public good yields only a return of a half-unit to the contributor. Hence, a social dilemma arises between the conflict of the individual's self-interest and the group's social-interest. A player cannot direct his or her cooperative (defective) behavior towards specific individuals, like in the dyadic interactions of the prisoner's dilemma, but only towards the group as a whole. Usually, players start off quite cooperative, but cooperation soon collapses to almost full defection (Milinski *et al.* 2002; Milinski 2006). Examples of public goods include the overuse of fish stock, leaving public toilets in a clean way, the protection of the environment, and the compliance to pay taxes.

The theoretical background of my thesis rests on the assumptions of evolutionary game theory and its associated concepts. Evolutionary dynamics provide a useful tool to study the conditions for the emergence and maintenance of cooperation. Here, I am interested in which "cooperative strategies" are found in humans, and thereby predictions are derived from evolutionary models.

Contents of the thesis

As outlined at the beginning, human cooperation stands out from all other forms of animal cooperation. Therefore humans provide an extremely interesting study species, but up to now many aspects of human cooperation are not fully understood. Profound conceptual overviews on the evolution of cooperation are provided by Hammerstein (2003) and Kappeler and van Schaik (2006). In general, the aim of this thesis is to investigate the conditions (and their interactions) that help humans to solve cooperation problems. Three topics will be presented. On the one hand, the impact of the social environment on cooperative behavior is addressed. Here, the questions are raised how varying numbers of social partners affect cooperativity and how the structure of social networks influences cooperative behavior of individuals. On the other hand, the impact of punishment as a process to stabilize cooperative behavior and how punishment possibly triggers backlashes is addressed.

Generally, cooperative behavior will be examined in systematic experimental investigations using the prisoner's dilemma and the public goods game. Naturally, experiments only provide a rather limited way to investigate social behavior. However, at the same time experiments provide a useful and necessary opportunity to reduce the complexity of social interactions and to place these interactions in a more controllable environment. For instance, the degree of anonymity is a relevant factor in social settings (Kurzban *et al.* 2007). However, in order to avoid contextual effects caused by anonymity, for instance reputational concerns of participants towards the experimenters, one has to provide full anonymity in experiments. Additionally, measurement errors can be avoided by a computerized set-up. How these and other confounding variables are controlled will be described in more detail in the method sections of Chapters 1 to 3. Needless to say laboratory results need to be treated carefully and need not to be overgeneralized.

In **Chapter 1**, I will investigate how the number of social counterparts affects cooperative behavior. Our everyday lives are made up of a countless number of social encounters, in which we interact with a variety of partners. For example, a campaigning mayor of a town interacts with nearly all inhabitants, whereas others may have only limited interactions (e.g. those with their closest neighbors). The fact that their number of partners varies greatly has been widely ignored in evolutionary games.

In this first experimental investigation, each participant will be involved in dyadic interactions and play iterated prisoner's dilemma. Half of the participants will play a single iterated prisoner's dilemma, thus they have one partner. As a new feature, the other participants will interact in three iterated prisoner's dilemma at a time; meaning that each

participant will have three partners. However, the three games are not linked to each other and different decisions can be made for each partner. Traditional evolutionary game theory assumes independence of games, i.e. one game is played after the other and payoffs are added up (Maynard Smith 1982; Nowak 2006a). Thus, no difference is expected between the two social settings. Nevertheless, individuals are constantly involved in several relationships, which take place at the same time. Thus, in principle experiences can be carried over from one relationship to another. So far there is hardly any experimental evidence of how humans behave when they interact with several partners in numerous cooperative dilemmas (for an exception based on groups, see Falk *et al.* 2010).

Here, I will examine whether the assumption of independent games holds and whether participants behave similar to multiple partners. Overall, I expect reciprocal cooperation in the iterated prisoner's dilemma of both settings, as the exact endpoints of relationships are unknown to participants. Thus, direct reciprocity is expected to operate (Trivers 1971; Axelrod & Hamilton 1981). This puts individuals in a position to use conditional strategies like tit-for-tat (Axelrod 1984) or win-stay lose-shift (Nowak & Sigmund 1993). Hence, the nature of strategic behavior will be investigated.

Chapter 2 focuses on the impact of social structure on cooperativity, as social networks are an essential feature of human societies (Kossinets & Watts 2006). Most theoretical analyses focus on investigating cooperative behavior in well-mixed populations, i.e. each individual is equally likely to interact with everybody else in the population. However, due to spatial conditions this assumption does not always hold and individuals primarily interact with neighbors close in proximity. Recently research has started to focus on structured populations. It has been demonstrated that certain structures of static networks can support the evolution of cooperation (Nowak & May 1992; Lieberman *et al.* 2005; Ohtsuki *et al.* 2006; but see Hauert & Doebeli 2004); for instance, cooperation prevails in spatial lattices, circles and scale-free networks. By assorting (i.e. clusters of neighboring individuals performing the same behavioral strategy) cooperators can avoid interactions with defectors, reducing the chance of being exploited (Nowak & May 1992; Brauchli *et al.* 1999; Ifti *et al.* 2004; see also Fletcher & Doebeli 2009). However, social relationships are flexible generating dynamic networks. Here, not only behavior evolves but also the network structure is under evolutionary pressure. This co-evolutionary process favors the evolution of cooperation (for reviews, see Gross & Blasius 2008; Perc & Szolnoki 2010). Despite theoretical advances in the last two decades, experimental evidence is scare or completely absent in the case of dynamic networks.

Here, I will investigate cooperative behavior in static and dynamic social networks. Participants will play iterated prisoner's dilemma with an unknown endpoint. In dynamic networks participants have the possibility to influence their social relationships based on an active-link-breaking mechanism (Pacheco *et al.* 2006a, 2006b, 2008). Thus, in dynamic networks an interaction can arise between behavior and the network structure, whereas in static networks cooperation can only be influenced by direct reciprocity within the prisoner's dilemma. As theory predicts, I expect higher levels of cooperation in dynamic networks (Perc & Szolnoki 2010). Additionally, as assortment of individuals and also clustering have been suggested to be important factors to favor cooperation, topological changes in the dynamic networks will be investigated.

In **Chapter 3**, the impact of punishment as a process to stabilize cooperation will be assessed. Punishment is a widely spread behavior among humans and animals (for reviews, see Clutton-Brock & Parker 1995; Sigmund 2007; Jensen 2010) and it is very effective in promoting cooperation in humans (e.g. Ostrom *et al.* 1992; Fehr & Gächter 2002; Gächter *et al.* 2008; but see Wu *et al.* 2009). However, punishment is not only costly for the recipient but also for the actor (though costs to assign punishment are somewhat lower than the actual punishment fine). Now, the following problem arises: as punishment is costly individuals should avoid to punish (Dreber *et al.* 2008) and thus punishment constitutes a second-order dilemma (Boyd & Richerson 1992). The consequence is that punishment cannot be evolutionary stable without additional mechanisms (e.g. Henrich & Boyd 2001; Brandt *et al.* 2003; Hauert *et al.* 2007).

Previous research in the area of costly punishment has mainly concentrated on situations where punishment cannot be retaliated. However, under most natural conditions this is not true; usually punishment can be avenged by victims. Thus, the possibility that punishment can escalate into vendettas where "I punish you, because you punished me; but you already punished me, because I punished you before" and so on becomes relevant. Theoretical research shows that vendettas of punishment are not an evolutionary stable behavior (Janssen & Bushman 2008; Rand *et al.* 2009b; Wolff 2009).

In this study, I will allow for vendettas by combining the public goods game with multiple rounds of costly punishment. Studies of punishment – where vendettas are impossible – show that people do indeed engage in costly punishment, which then stabilizes public goods contributions (e.g. Fehr & Gächter 2002). Therefore, albeit the high costs of punishment and the threat of being counter-punished, I expect participants to engage in punishment. Additionally, I also anticipate the occurrence of vendettas as they are observed in the real

world (Ericksen & Horton 1992; İçli 1994; Gould 2000). Subsequently, it will be highly interesting to see how cooperative behavior in the public goods game will be affected.

Overall, the aim of this thesis is to evaluate conditions which affect cooperative behavior in dyadic and group interactions. By doing so, this thesis will contribute a piece of knowledge which eventually helps to achieve a better understanding of the evolution of (human) cooperation.

CHAPTER I INTERACTING IN MULTIPLE PRISONER'S DILEMMA

MANY FRIENDS TEMPT YOU TO CHEAT: DECREASED COOPERATION IN MULTIPLE PRISONER'S DILEMMA

with Dirk Semmann[1]

[1] Courant Research Center Evolution of Social Behavior, University of Göttingen, Germany.

Submitted to *Proceedings of the Royal Society B: Biological Sciences*

Abstract

Humans are an extraordinarily social species. Throughout our day-to-day lives we interact with a variety of counterparts; some interact with many, others only with a few. In an experiment with human participants, we investigate how the number of interaction partners impacts cooperative behavior in the iterated prisoner's dilemma (IPD) with an unknown ending. Half of the participants played a single IPD, which is the common set-up. As a new feature, the other participants interacted in three IPDs at a time. Traditional evolutionary game theory assumes independence of games and thus no difference would be expected in the two social settings. Contrary to this assumption, we find that overall cooperation is lower in the multiple-games setting. In fact, these participants could only establish one cooperative relationship similar to the relationship of the single-game setting, where cooperativity increased over time. Moreover, in one of the two remaining relationships cooperation could not gain a foothold, although cooperative behavior is expected when direct reciprocity can operate. In addition, contradictory to previous findings participants did not rely on a win-stay lose-shift strategy; they used reactive strategies that close to generous tit-for-tat.

Keywords

Cooperation, evolutionary game theory, iterated prisoner's dilemma, multiple games, reciprocity, tit-for-tat

Introduction

Many daily activities, in which humans engage in, are profoundly social and throughout these humans encounter a variety of counterparts. However, within these relationships cooperative behavior constitutes an evolutionary puzzle (see Box 1). Despite this challenge, nature abounds with many examples of cooperativity among humans as well as animals (for recent reviews, see e.g. Hammerstein 2003; Pennisi 2009; Melis & Semmann 2010). Here, we are interested in how the number of social partners impacts cooperative behavior. In doing so, we use the framework of evolutionary game theory and the iterated prisoner's dilemma with an unknown endpoint.

Evolutionary game theory has concentrated on interactions where games are independent, or where one game is played after the other and payoffs are added up (Maynard Smith 1982; Nowak 2006a). Nevertheless, it is plausible to assume that individuals are constantly involved in more than one relationship. Thus, in principle experiences can be carried over from one relationship to another. This scenario is for instance important in structured populations, which have become a favorite topic for studying the evolution of cooperation (e.g. Nowak & May 1992; Brauchli *et al.* 1999; Hauert & Doebeli 2004; Szabó & Fáth 2007; Lion *et al.* 2011). In many biological and social structured systems the interactions between individuals can be characterized as heterogeneous, scale-free networks (Amaral *et al.* 2000; Dorogotsev & Mendes 2003). Furthermore, recent studies show that within heterogeneous networks cooperation evolves (Santos & Pacheco 2005; Santos *et al.* 2006b; Fu *et al.* 2007; Assenza *et al.* 2008; Szolnoki *et al.* 2008; but see Konno 2011). The essential characteristic within these networks is that some individuals have many more contacts than others. Consequently, the number of interactions or social dilemmas per individual varies greatly.

Theory provides only limited predictions for the effect of varying partner numbers on human cooperation. Within structured populations theory predicts that the number of interactions (this equals the number of social partners) is a central feature for natural selection to favor cooperation. In a number of network structures the general rule that the cooperative-benefits-to-costs ratio should be larger than the average number of partners has been identified (Ohtsuki *et al.* 2006; Ohtsuki & Nowak 2007). Hence, the fewer partners one

has the easier cooperation can evolve (Ifti *et al.* 2004). However, this view has now been challenged (Szolnoki *et al.* 2008; Chen *et al.* 2011; Konno 2011; Yamauchi *et al.* 2011) and more research along these lines is needed. In general, a common assumption is that individuals play pairwise games, but can only adopt one strategy to all their partners, i.e. they behave unconditionally. This, however, intensifies the problem of the prisoner's dilemma and rather constitutes a public-goods situation. This differs from our experimental setting (see *Methods*) where individuals play pairwise games, but can still choose independently for different partners – they can use conditional strategies. Therefore, it is not clear how these theoretical predictions relate to our setting, and we rely on the traditional assumption of evolutionary game theory that games are independent.

So far there is hardly any experimental evidence of how humans behave when they interact with several partners in numerous cooperative dilemmas. However, this is central for understanding how diverse social settings influence the evolution of cooperation. There has been an increasing awareness of this issue (Hauk 2003; Ahn *et al.* 2009); nonetheless, Falk and colleagues (2010) seem to provide the only experimental comparison of cooperation in an one-game setting with a multiple-games setting (i.e. individuals participate in two, simultaneous public goods games; a group game, whereas we investigate a dyadic game). They find no difference between the settings and in both public-goods contributions follow the usual pattern. Additionally, the two simultaneous games do not influence each other. There are also studies investigating how different kinds of games influence each other, which show effects of behavioral spillover from one type of game to the other (in alternating games: Milinski *et al.* 2002; Barclay 2004; Semmann *et al.* 2004; in simultaneous games: Bednar *et al.* 2010; Cason *et al.* 2010; Savikhin & Sheremeta 2010).

In this experimental study, we examine whether human participants are affected in their cooperativity when placed in a setting of a single iterated prisoner's dilemma (IPD; see Box 1) in comparison to a setting of three, simultaneously played IPDs. In both settings the precise number of rounds is unknown to participants. Games are understood as independent, we therefore expect no difference in cooperative behavior within the two settings. Derived from this assumption, participants in the multiple-games setting should also treat all three partners alike. In addition, our set-up allows us to further fill the gap on long-term interactions with an uncertain ending (cf. Box 1). In accordance with previous theoretical and experimental literature (Trivers 1971; Axelrod & Hamilton 1981; Dal Bó 2005; Duffy & Ochs 2009), we expect a cooperative outcome in both settings as direct reciprocity can operate. In line with previous findings (Wedekind & Milinski 1996), we conjecture that participants use strategies similar to win-stay lose-shift (this is true for both settings, as we overcome constraints associated with working-memory load, cf. *Methods*).

Apart from this, it is still unclear how specific relationships develop over time, as previous research has provided evidence that within a relationship cooperation can increase (Dal Bó 2005) as well as decrease (Duffy & Ochs 2009, though cooperation increases over several relationships played one after the other). In brief, (i) we address the impact of different numbers of social partners in IPDs, (ii) ask whether multiple interactions are independent, and (iii) examine the reciprocal nature of the game's outcome.

Box 1 ***The evolution of cooperation and the prisoner's dilemma***

The abundance of human cooperation is an evolutionary puzzle when defectors benefit from cooperative interactions without bearing the associated costs, because under natural selection and without any other mechanisms one expects the emergence and persistence of defective behavior. The evolution of cooperation can be studied by the mathematical approach of evolutionary game theory (Maynard Smith 1982) and the prisoner's dilemma (PD; Rapoport & Chammah 1965; Axelrod 1984). In the PD two individuals decide simultaneously whether to cooperate or to defect. If both cooperate, they each receive the reward payoff (R). If one defects and the other cooperates, the defector gets the temptation payoff (T) and the cooperator obtains the sucker's payoff (S). However, if both defect, they each receive the punishment payoff (P). The assumption $T > R > P > S$ must hold. If the individuals cooperate, both do better than if they both would have defected. But for a single individual it is always better to defect no matter what the partner does. Thus, defection is the evolutionary stable strategy in one-shot interactions.

If the PD is played repeatedly, the assumption $2R > T + S$ must hold, because then the payoff of two individuals is higher when both cooperate than if they would alternately choose cooperation and defection. Next, the distinction of finitely or infinitely repeated games becomes important. If the individuals are aware of the PD's ending, there is no incentive to cooperate in the last round as the partner has no opportunity to reciprocate this defection, and no future gains will be lost if both would then drive into mutual defection. However, anticipating that one's counterpart has the same understanding and by using backwards induction it is then best to defect in the second last round as both individuals assume defection in the very last round. Following this line of thought, the individuals should end up in mutual defection in all rounds. A large amount of experimental research partially supports this assumption, as players start off cooperating but turn to the predicted mutual defection towards the end of the game (e.g. Selten & Stoecker 1986; Andreoni & Miller 1993; Cooper & Ross 1996).

On the contrary, in infinitely iterated PD reciprocal cooperation can be an evolutionary

stable strategy when the probability for a continuous interaction is large enough (Trivers 1971; Axelrod & Hamilton 1981). Among strategies frequently discussed in the theoretical literature are tit-for-tat (Axelrod 1984), generous tit-for-tat (Nowak & Sigmund 1992), and win-stay lose-shift (Nowak & Sigmund 1993). Humans primarily use win-stay lose-shift like strategies (Wedekind & Milinski 1996; Milinski & Wedekind 1998). However, they turn to the simpler generous tit-for-tat like strategies when the game is interfered by a second task (Milinski & Wedekind 1998). Experimental evidence on iterated PD with unknown endings (this resembles the infinite character of the game), where humans continuously play with the same partner, and not against pseudo-partners or computers with pre-programmed strategies, and where real and adequate amounts of money are at stake, is scarce. Nonetheless, studies confirm the predicted cooperative outcome (Dal Bó 2005; Aoyagi & Fréchette 2009; Duffy & Ochs 2009; Fehl *et al.* 2011), except when additional competitive incentives are provided (West *et al.* 2006). Infinite relationships presumably constitute a more realistic setting of human and animal interactions, because individuals rarely can foresee the precise endpoint of a relationship.

Method

We recruited a total of 200 students from the University of Göttingen via the online recruitment system ORSEE (Greiner 2004) in fall 2009 and 2010. The students (49% females) came from various disciplines and were on average 23.36 ± 2.79 (mean ± SD) years old. Upon arrival participants were randomly seated in front of touch-screen computers; they were visually separated by partitions, and received written instructions. Participants interacted by means of a computer software; no other form of communication was permitted. Through assignment of aliases, i.e. names of moons of our solar system (e.g. Kallisto, Leda, Metis), participants were ensured that their decisions were made completely anonymously towards other participants and the experimenters. Aliases were also used to carry out anonymous payment (as described in Semmann *et al.* 2005; participants knew this procedure from written instructions before playing).

We ran two treatments and each consisted of 10 sessions with 10 participants in each session. The first treatment consisted of the standard IPD with one partner. Pairs of participants were randomly assigned at the beginning of the experiment. The game was played for 30 rounds, which was unknown to participants. For each round participants were asked to choose between two options (called *orange*- or *blue*-option). In half of the sessions orange mimicked cooperation and blue defection, in the other half the reversed pattern was used; hence, prefixed moral pressure to choose "cooperation" due to wording was excluded.

There was no impact of color coding on cooperativity (both treatments, Mann-Whitney test: $U = 35$, $n_{1,2} = 10$, $p = 0.28$). A list of the possible outcomes with respective partners was presented to the participants all along the experiment. The respective payoffs were 0.40€ for the temptation, 0.25€ for the reward, 0.00€ for the punishment and -0.10€ for the sucker's payoff (cf. Box 1). After each interaction participants were shown their own decision and their partner's decision as well as the corresponding payoffs. Overall, while making 30 decisions participants attended the lab for about 60 minutes and were given a 5.00€-starting amount.

In the second treatment participants played the same IPD, but instead of having only one partner they had three; randomly assigned at the beginning of the experiment. Given that 10 participants attended one session the connections between them can be visualized as a structured population (see Fig. S1, in *Electronic Supplementary Material [ESM]*). For each round, participants had to decide for each partner independently whether to play orange or blue. The information of each interaction (i.e. partner's decision and the respective payoffs of a pair) was displayed on the same screen. Though the three IPDs were played at a time, they were completely independent of each other, i.e. a decision in one game did not change the payoffs in another. Participants had to make 90 decisions and they attended the lab for approximately 90 minutes. They received a starting amount of 3.00€ (different starting amounts were chosen to achieve similar earnings over time, see *ESM* for further details).

Furthermore, working-memory loads can affect the behavior in IPD (Milinski & Wedekind 1998; see also *ESM*). We reduced the influences of memory effects to a minimum when playing in a single-game versus a multiple-games setting. In both treatments, we set no time limit for the decisions to be made and the feedback information of the IPD outcomes could be viewed at individually preferred durations. In addition, all participants were provided with a blank piece of paper in order to make notes, if they wished to do so. To maintain anonymity all papers were destroyed at the end of the experiment, and neither other participants nor the experimenters could access these notes.

For statistical analyses SPSS 18.0.3 and R 2.12.1 were used. Probabilities are reported as two tailed and a 5%-level of significance is used. Furthermore, analyses were done on the group level to account for session effects (especially in the multiple-games treatment).

Results

In the treatment of a single IPD we observed an average cooperation level of 71.70% ± 7.73; whereas the average cooperation level in the treatment of multiple IPDs was 54.91% ± 6.16. The difference in cooperation levels was significant (Mann-Whitney test: $U = 4$, $n_{1,2} = 10$,

$p < 0.001$). To assess the overall difference between the treatments, we further examined the behavior within the dyads of the multiple-games treatment. We assigned participants a "cooperation score" for each partner by giving them one point for every cooperative move towards a partner (theoretically taking values from 0 to 30). We found that only 3% of participants had equal cooperation scores in all relationships, i.e. they treated all partners alike. Eighteen per cent of participants had two similar cooperation scores. The great majority (79%) of participants, however, had three different cooperation scores. Based on this observation, we extracted the most-, middle-, and least-cooperative relationship of each participant (see Fig. 1). The average cooperation level of the most-cooperative relationship (71.27% ± 6.37) did not differ from the single-game treatment (Mann-Whitney test: $U = 48$, $n_{1,2} = 10$, $p = 0.91$). However, the other two relationships showed a significantly lower average cooperation level compared to the single-game treatment (middle-cooperative relationship: 56.90% ± 7.13, $U = 6$, $n_{1,2} = 10$, $p < 0.001$; least-cooperative relationship: 36.57% ± 7.16, $U = 0$, $n_{1,2} = 10$, $p < 0.001$). In addition, a numerical analysis shows that different cooperative stationary states are obtained (see Fig. S3, *ESM*).

In the treatment of a single IPD, and in the most- and middle-cooperative relationships of the multiple-games treatment we observed an increase in cooperativity over time when comparing the cooperation levels of round 1 and round 30 (see Fig. 1; Wilcoxon signed-rank test; single-game IPD: $Z = 2.69$, $n = 10$, $p < 0.01$; most-cooperative relationship: $Z = 2.46$, $n = 10$, $p < 0.05$; middle-cooperative relationship: $Z = 2.53$, $n = 10$, $p < 0.01$). Remarkably, there was no difference between cooperation levels in round 1 and round 30 of the least-cooperative relationship of the multiple-games treatment ($Z = 0.60$, $n = 10$, $p = 0.59$).

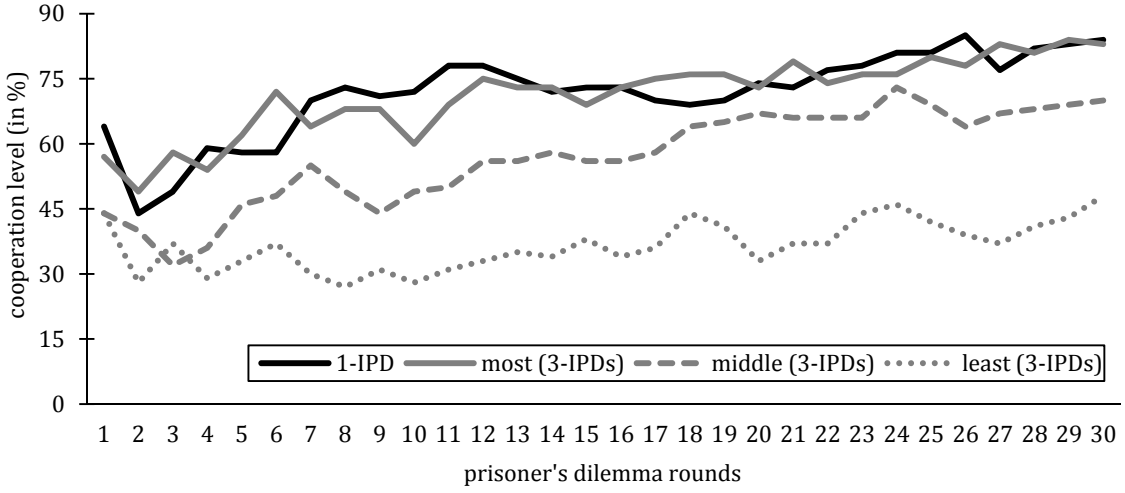

Figure 1 Cooperation levels of iterated prisoner's dilemma (IPD) for 30 rounds (the endpoint was unknown to participants). Participants either played a single game (1-IPD; average SD = 7.73) or they played with three partners simultaneously, though independently (3-IPDs). These three games are ranked from the *most-*, *middle-*, to *least*-cooperative relationship (average SD = 6.37, 7.13, 7.16, respectively).

Theory assumes that cooperation is reached via reciprocating the behavior of the partner. The relative frequencies of cooperative behavior following the four possible outcomes of the previous round reveal the behavioral strategies of participants (see Fig. 2; see also Fig. S2, *ESM*). The cooperative choices did not significantly differ between treatments for all cases except for mutual defection (Mann-Whitney test; mutual cooperation: $U = 41$, $n_{1,2} = 10$, $p = 0.53$; the participant was exploited: $U = 42$, $n_{1,2} = 10$, $p = 0.58$; the participant exploited his or her partner: $U = 40$, $n_{1,2} = 10$, $p = 0.48$). After mutual defection marginal significantly more participants defected in the multiple-games treatment ($U = 28$, $n_{1,2} = 10$, $p = 0.10$). Furthermore, to cooperate after mutual cooperation and to once in a while cooperate after defection by the partner fits a reactive strategy of generous tit-for-tat. After the exploitation of a partner the cooperative response is well below the expected 1 for generous tit-for-tat, but nevertheless the majority cooperated (see Fig. 2). Thus, in both treatments these relative frequencies resemble reactive strategies close to generous tit-for-tat rather than the expected win-stay lose-shift.

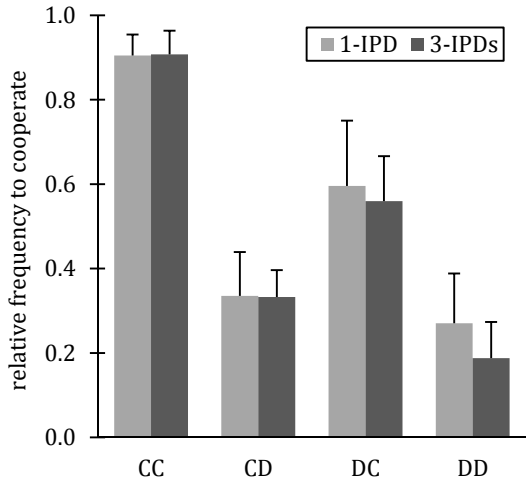

Figure 2 Relative frequency (pooled over all rounds) of cooperative behavior (+ SD) after mutual cooperation (CC), when the participant cooperated and the partner defected (CD), when the participant exploited his or her partner (DC), or after mutual defection (DD). Participants either played one iterated prisoner's dilemma (1-IPD; average occurrence of outcomes per round at an individual level [games = 100]: CC = 58, CD = 13, DC = 13, DD = 16) or three independent games at a time (3-IPDs; average n per round [games = 300]: CC = 116, CD = 47, DC = 47, DD = 90).

Discussion

We used the iterated prisoner's dilemma with an unknown ending in an experimental setting with human participants to study the emergence of cooperation. For the first time, to our

knowledge, we compare the traditional IPD with one partner to a setting of multiple IPDs played with different partners at a time. In general, independent decisions had to be made and identical payoffs were used in each dyadic relationship. Hence, we can examine the impact of the number of social partners on cooperation. We show that having more partners leads to a significantly lower average cooperation level. Additionally, participants in the multiple-games treatment tended to cooperate less often after mutual defection than participants who only played a single IPD. Further analysis of the three relationships in the multiple-games setting showed that participants were, nevertheless, able to establish high levels of cooperation in *one* dyadic interaction (i.e. there was no difference in the average cooperation level compared to the single-game cooperation level and cooperativity increased over time); whereas the two remaining relationships exhibited lower on average cooperation levels. Especially in the least-cooperative relationships participants were not able to raise cooperativity over time. Here, cooperation levels remained as low as 30% to 40%. Considering these observations in the multiple-games treatment and results of the numerical analysis of behavior in the experiment, we cannot support recent experimental findings on groups (Falk *et al.* 2010) and the traditional assumption of evolutionary game theory of game independence.

Several mechanisms could be responsible for the differences between the three types of relationships in the multiple-games treatment. First, cooperative individuals face an increased risk of exploitation and uncertainty in the setting of multiple IPDs. Now in order to avoid possible losses individuals would have to decrease their overall cooperativity (Kahneman & Tversky 1979; but see Harinck *et al.* 2007). Our results show that participants established at least one cooperative and trustworthy relationship, in which uncertainty was most likely reduced. However, high rates of defection and thus a greater risk of exploitation was found in one relationship. A second explanation is that behavioral spillovers impact the relationships. This has been demonstrated in other contexts (Milinski *et al.* 2002; Barclay 2004; Semmann *et al.* 2004; Bednar *et al.* 2010; Cason *et al.* 2010; Savikhin & Sheremeta 2010). In our experiment the relationships of the multiple-games treatment are different; this makes at least a consistent influence of one relationship on another unlikely. Third, in the setting of multiple-games the "temptation" to exploit others is enhanced. Participants had one partner with whom cooperativity was high. This, however, seem to be enough of a secure income and participants were tempted to exploit another partner or participants reacted more likely with defection to defectors, thereby resulting in low cooperation in one relationship. This explanation appears to hold best.

We were interested in whether participants would establish cooperation via direct reciprocity and win-stay lose-shift behavior (see Box 1; Trivers 1971; Axelrod & Hamilton

1981; Nowak & Sigmund 1993; cf. also results of a post-questionnaire in *ESM*) and given that results are so far inconclusive for the cooperativity development over time (Dal Bó 2005; Duffy & Ochs 2009). In both treatments, most participants cooperated (i.e. average cooperation levels are above 50%). In addition, our results show that humans interacting in IPD without clear endpoints can stabilize cooperative behavior and that the majority can raise cooperation levels over time, which is consistent with previous research (Dal Bó 2005; but contrasts Duffy & Ochs 2009). There are several results indicating that participants of both treatments behaved reciprocal. That is participants reacted selectively to previous outcomes and cooperated most when both members of a relationship cooperated in the previous round; though participants who played multiple IPDs tended to be more likely to defect after mutual defection, overall resulting in different cooperative states. This indicates that these participants were more likely to write off a relationship. Furthermore, we found that their actions in the IPD followed reactive strategies close to generous tit-for-tat (Nowak & Sigmund 1992). Our results do not replicate earlier findings where humans preferred to use strategies similar to win-stay lose-shift in a IPD (Wedekind & Milinski 1996). The simpler generous tit-for-tat like strategy was preferred when the game was interfered by an additional task (Milinski & Wedekind 1998). Here, we reduced such effects of working-memory load (see *Methods*), and participants, nevertheless, applied strategies similar to generous tit-for-tat in both treatments. It is worthwhile to mention that both previous studies used pseudo-partners, who used predetermined strategies, to effectively test for conditional behavior. In contrast, our results rest on free-play behavior between real participants. Overall, results correspond with the classic understanding in evolutionary game theory of the emergence of human cooperation via direct reciprocity.

An interesting field to which our results relate is the research on the evolution of cooperation in structured populations, for instance in scale-free networks. Here, the number of interaction partners varies greatly. Theory shows that an increase in partners and thus in interactions can hinder cooperation in structured populations (Ifti *et al.* 2004; Ohtsuki *et al.* 2006; Ohtsuki & Nowak 2007). This result applies to unconditional behavior where individuals react with one strategy to all partners. We now provide experimental support for a setting where individuals can adjust their behavior conditionally to each partner and showed that with more partners overall cooperativity is lower. The problem of reduced cooperation with multiple partners can be overcome by allowing individuals to reject interaction partners. Thereby a dynamic network is generated. Both theory (Perc & Szolnoki 2010) and recent experimental results (Fehl *et al.* 2011) support this assumption by showing that cooperation is enhanced in dynamic networks. A second possibility, which is not necessarily exclusive to the first one, is to let the number of partners vary among

individuals. Theory shows that in static-heterogeneous networks where so-call hubs exist, i.e. individuals that have many more relationships than others, cooperation can evolve (Santos *et al.* 2006b; Santos & Pacheco 2005, 2006), but experimental results are still missing. Especially these hub individuals are found to be cooperative; even though our results show that to cooperate with many partners might be difficult. Essentially, this calls for further studies on varying the partner numbers in evolutionary games within experiments, especially in heterogeneous, large-scale social networks to further validate theoretical assumptions on the evolution of cooperation in complex social settings.

In summary, we would like to emphasize that our results are in contrast to the traditional assumption of evolutionary game theory that multiple games are independent. We find that there is an impact of playing several games that results in overall less cooperation. Participants in the multiple-games setting only established one cooperative relationship, which contrasts with another of their relationships where cooperativity remained low. Thus, even though an identical game structure was provided, participants behaved differently. In conclusion, a new type of models is required which account for behavioural differences within the relationships of a single individual. Theorists have begun to study cooperation within heterogeneous networks, where the numbers of partners vary. The standard within these models is that individuals cannot strategically differentiate between partners. Nevertheless, our results show that this is essential to better understand cooperation in human relationships.

Acknowledgements

We are grateful to Daniel van der Post and Arne Traulsen, who also did the numerical analysis, for stimulating discussions. We thank the students at the University of Göttingen for their participation and special thanks go to Johannes Pritz, Frederic Nowak and Christine Wittge for support. The research is funded by the German Initiative of Excellence of the German Science Foundation (DFG).

APPENDIX TO CHAPTER I

Electronic Supplementary Information

Experimental set-up: Structure of interactions

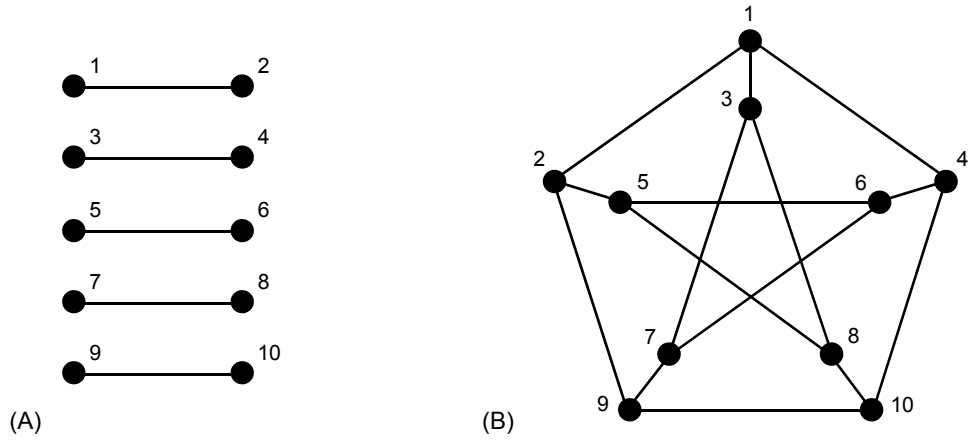

Figure S1 Overview of interacting partners in the treatment of one iterated prisoner's dilemma (A) and in the treatment of three, simultaneous iterated prisoner's dilemma (B).

Experimental set-up: Starting amounts

We chose different starting amounts for the two treatments in order to control for different amounts of time spent in the lab (60 minutes versus 90 minutes), and the fact that 90 decisions in the multiple-games treatment provided more opportunities to earn money compared to 30 decisions of the single-game treatment. The goal was that on average amounts would be earned that resemble the same students' hourly wages; thereby participants would make careful decision, because meaningful amounts were at stake. Nevertheless, all participants played with the same prisoner's dilemma payoff matrix and thus the same formal incentive structure per interaction. To test whether all participants earned comparable amounts of money (average total payoff: 13.31€ ± 2.99 [mean ± SD]) independent of the treatment, we calculated the average payoff per minute. In the treatment of a single IPD participants earned on average 0.18€ ± 0.01 and in the treatment of multiple games participants earned on average 0.18€ ± 0.02. There was no significant difference in earnings per minute (Mann-Whitney test: $U = 45$, $n_{1,2} = 10$, $p = 0.74$).

Effects of working-memory load

The current load of working memory is known to effect cooperative behavior (Milinski & Wedekind 1998). In our experimental set-up this could interfere with the effect of interacting with different numbers of social partners, however, we reduced the influence of the working-memory effect (i) by setting no time limit for the decisions to be made, (ii) by individually preferred durations of the feedback information of the iterated prisoner's dilemma (IPD) outcomes, and (iii) by providing participants with blank pieces of paper in order to make notes if they wished to do so.

In addition, from another experiment (Fehl *et al.* 2011) conducted in our lab under the very same conditions we know that participants are good at paying attention to their partners and at recalling social information. In one of the treatments of this experiment, participants also interacted in three simultaneous IPDs each lasting 30 rounds. However, they could change partners and thereby interact with up to nine different partners. Here, the working-memory load should be even higher and the social setting is more complex. In a computerized post-questionnaire participants (n = 100) were asked whether they remembered to have played with the different participants from their session. In total, 89.9% ± 0.05 could correctly identify all their partners (by aliases) and all the participants they have not interacted with. In addition, 27.5% ± 0.11 stated the exact number of rounds they interacted with their partners (59.0% ± 0.09 correctly stated the number of rounds within a range of ± 2). These percentages are remarkable since the number of rounds was of no relevance to participants, i.e. they did not know how many rounds would be played, and during the experiment the current round number was not presented to participants. In addition, to use reciprocal strategies only the outcome of the previous round is of interest. Nevertheless, many participants could recall this information. Moreover, participants' statements of how many rounds their partners had cooperated deviated only by 11.3% ± 0.06 from the actual numbers of rounds their partner had cooperated (all interactions where the participants' guessed numbers of rounds was equal to the actual numbers of rounds). In sum, as we used the exact same method in the present experiment, where the social complexity is lower, we conclude to have reliably reduced the impact of working-memory load in the present experiment by providing additional tools. Therefore, the impact of working-memory load when interacting with one partner or three partners is of no or little relevance for cooperative decisions.

Cooperative behavior: Numerical analysis

The relative frequencies of cooperative behavior following the outcomes of the previous round reveal the behavioral strategies of participants, who either played a single IPD or played with different partners simultaneously three IPDs. Possible outcomes within the IPDs are: mutual cooperation (CC), the participant cooperated and the partner defected (CD), the participant exploited his or her partner (DC), or mutual defection (DD; see Fig. S2, cf. also Fig. 2).

Based on the strategy choice parameters of the experiment the temporal dynamics in the IPD were simulated. The relative frequencies to cooperate given the prisoner's dilemma outcome of the previous round (see Fig. S2) and the initial distribution of prisoner's dilemma outcomes in round 1 from the experiment were used. The probabilities to cooperate after the four different outcomes (CC, DC, CD, and DD) define a stochastic strategy in the IPD (Nowak & Sigmund 1990). For this strategy, one can construct a transition matrix between the different states, e.g. the probability to go from CC to CD is $p_{cc}(1-p_{cc})$, where p_{cc} is the probability to cooperate after a round of mutual cooperation. The level of cooperation shown in the numerical analysis is the fraction of cooperative moves, starting from the initial condition of the experiment (see Fig. S3; cf. also Fig. 1). In the long-run, a stationary state is reached, in which the probability of moves is given by the first eigenvector of the stochastic transition matrix.

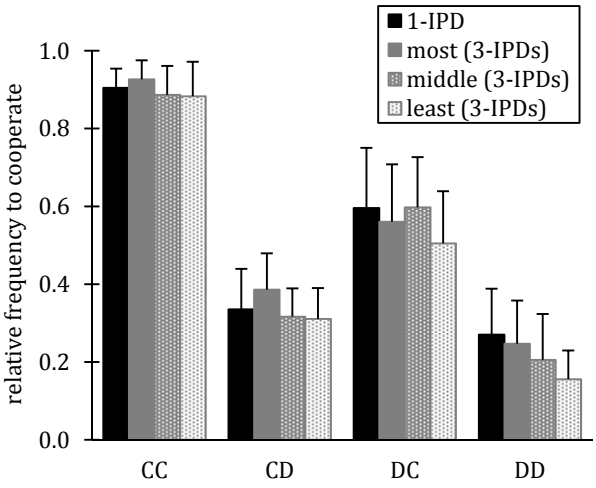

Figure S2 Relative frequency (pooled over all rounds) of cooperative behavior (+ SD) after mutual cooperation (CC), when the participant cooperated and the partner defected (CD), when the participant exploited his or her partner (DC), or after mutual defection (DD). Participants either played one iterated prisoner's dilemma (1-IPD), or three independent games at a time (3-IPDs). These three games are ranked from the *most-*, *middle-*, to *least-*cooperative relationship.

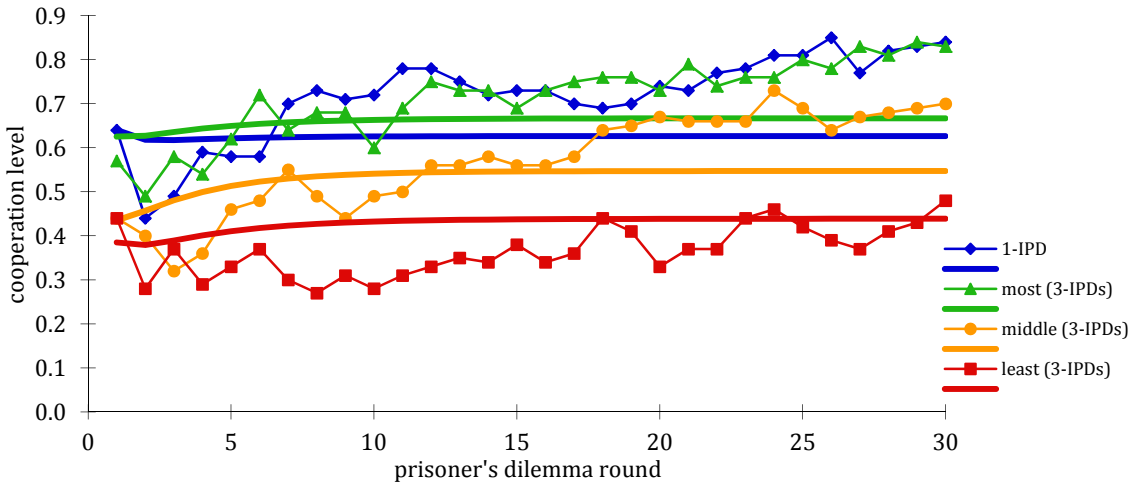

Figure S3 Cooperation levels of iterated prisoner's dilemma (IPD) for 30 rounds. Participants either played a single game (1-IPD), or they played with three partners simultaneously, though independently (3-IPDs; lines with symbols). These three games are ranked from the *most-*, *middle-*, to *least*-cooperative relationship. Continuous, bold lines are cooperation states from the numerical analysis.

Results of the post-questionnaire

At the end of the experiment participants were asked to fill out a computerized questionnaire, which provided us with self-reported experiences apart from the behavioral responses in the IPDs (individual-level analysis). Their answers concerning their motivation to participate showed that the majority ("yes" = 67%) wanted to earn money (see Fig. 3a). This supports our choice of a payoff-oriented set-up to measure costly, but cooperative incentives. Participants in both treatments reported to have focused on reciprocal decisions. In the multiple-games treatment more participants ("yes" = 76%) reported to have applied reciprocal strategies than in the single-game treatment (59%; Chi-square test: $\chi^2 = 7.79$, $df = 2$, $p < 0.05$; see Fig. 3b). Our indirect approach (to reduce the social desirability bias) to reveal whether participants exploited others via the statement "before others could exploit me, I rather did it", shows that most participants answered "no" or "in parts" (85%; see Fig. 3c). This shows a general tendency to engage in costly cooperation, as they refrained from exploiting others.

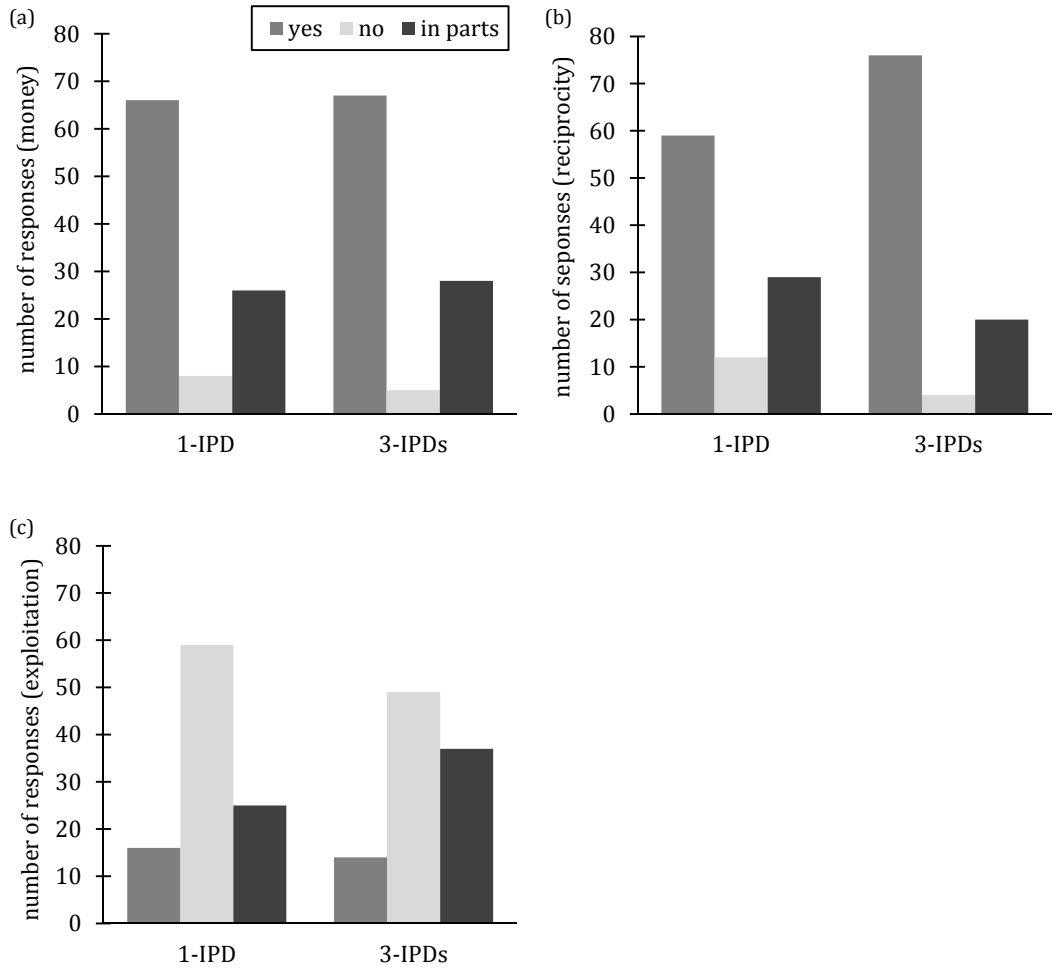

Figure S4 Results of the computerized post-questionnaire in the treatment of a single iterated prisoner's dilemma (1-IPD; n = 100) or of three games (3-IPDs; n = 100). (a) "I tried to earn as much money as possible" (Chi-square test: $\chi^2 = 0.77$, $df = 2$, $p = 0.68$). (b) "My decision (blue/orange) depended a lot on my partners previous decisions" ($\chi^2 = 7.79$, $df = 2$, $p < 0.05$). (c) "Before others could exploit me, I rather did it" ($\chi^2 = 3.38$, $df = 2$, $p = 0.18$).

CHAPTER II COOPERATION AND SELF-ORGANIZATION IN NETWORKS

CO-EVOLUTION OF BEHAVIOUR AND SOCIAL STRUCTURE PROMOTES HUMAN COOPERATION *

with Daniel J. van der Post[1] and Dirk Semmann[1]

[1] Courant Research Centre Evolution of Social Behaviour, University of Göttingen, Germany.

Published in *Ecology Letters* (2011), 14, 546-551

Abstract

The ubiquity of cooperation in nature is puzzling because cooperators can be exploited by defectors. Recent theoretical work shows that if dynamic networks define interactions between individuals, cooperation is favoured by natural selection. To address this, we compare cooperative behaviour in multiple but independent repeated games between participants in static and dynamic networks. In the latter, participants could break their links after each social interaction. As predicted, we find higher levels of cooperation in dynamic networks. Through biased link breaking (i.e. to defectors) participants affected their social environment. We show that this link-breaking behaviour leads to substantial network clustering and we find primarily cooperators within these clusters. This assortment is remarkable because it occurred on top of behavioural assortment through direct reciprocity and beyond the perception of participants, and represents a self-organized pattern. Our results highlight the importance of the interaction between ecological context and selective pressures on cooperation.

* This article has been published in British English.

Keywords

Assortment, co-evolution, cooperation, dynamic network, game theory, prisoner's dilemma, self-organization, social behaviour.

Introduction

Cooperative behaviour is widespread throughout the animal kingdom (for recent reviews, see Pennisi 2009; Melis & Semmann 2010). Such cooperation occurs within social animals, which naturally interact in networks, for instance in guppies where pairs more likely inspect predators when they have strong social associations with the partner (Croft *et al.* 2006). In primates and social insects, network structures affect the environment in which the individuals socially interact and also cooperate (Fewell 2003; Voelkl & Kasper 2009). In addition, in humans, social networks are an essential feature of social behaviour (Kossinets & Watts 2006). However, from an evolutionary perspective, cooperative behaviour is puzzling. This is because given that cooperative behaviour benefits others and produces costs for the actor, there is the potential for exploitation of cooperative individuals by "cheaters". Thus, those individuals enjoying cooperative benefits without performing cooperative acts themselves should be favoured by natural selection. To understand the evolution of cooperation, particularly in relation to the structure of animal social networks, is therefore a challenge.

Network reciprocity has been put forward as a mechanism to explain how the structure of static networks can support the evolution of cooperation (Nowak & May 1992; Lieberman *et al.* 2005; Ohtsuki *et al.* 2006; but see Hauert & Doebeli 2004). Cooperation can prevail in spatial lattices, because by assorting (i.e. clusters of neighbouring individuals performing the same behavioural strategy) cooperators can avoid interactions with defectors, reducing the chance of being exploited (Nowak & May 1992; Brauchli *et al.* 1999; Ifti *et al.* 2004; see also Fletcher & Doebeli 2009). In line with this theoretical work, evolutionary simulations based on social networks of non-human primates show that these have the appropriate static structure to support cooperation (Voelkl & Kasper 2009).

However, in relation to more extensive theoretical work, it is somewhat surprising that so far, experiments with humans could not show that network structure promotes cooperation. Both spatial lattices and other network topologies either caused cooperation to decline over time (Grujić *et al.* 2010; Traulsen *et al.* 2010) or could not convincingly reveal differences in levels of cooperation between network structures (Cassar 2007; Kirchkamp & Nagel 2007).

A potentially very important network property has, however, been neglected in these studies: network dynamics. In dynamic networks, not only do strategies evolve but also the network topology is under evolutionary selection pressure. Recent theoretical work shows that such co-evolution of behaviour and network structure favours the evolution of cooperation (for reviews, see Gross & Blasius 2008; Perc & Szolnoki 2010). In particular, the "active-linking" models of Pacheco *et al.* (2006a, 2006b, 2008) show that when individuals playing prisoner's dilemma (PD; see Box 1) are allowed to control their interactions, i.e. to break existing links and to form new links with random partners, cooperation evolves.

The defining feature of dynamic networks is the interaction between behaviour and network structure. Such interactions allow feedback to arise allowing individuals to assort on the network and to alter their social environment. This in turn can have an impact on individual fitness and hence selection pressures on behavioural strategies at the individual level. In general, such ecological interactions and the self-organizing, or self-structuring processes that they generate, have been suggested as fundamental to understanding evolution, in particular that of cooperation (Hauert *et al.* 2006; Lion & van Baalen 2008).

In general, in models of dynamic linking individuals can only react unconditionally (same reaction to all partners). Such models have been used to show that network reciprocity can be sufficient to support cooperation (Pacheco *et al.* 2006a, 2006b). Cooperation is favoured if the link-breaking rate to defectors is high (Fu *et al.* 2009; Wu *et al.* 2010) and if links between cooperators are long-lived (Pacheco *et al.* 2006a, 2006b; Santos *et al.* 2006a; Fu *et al.* 2008, 2009; Wu *et al.* 2010). In addition, in dynamic networks the formation of clusters has been suggested to support cooperation (Jun & Sethi 2009; but see Hanaki *et al.* 2007). Other dynamic network models include the possibility for reacting conditionally to different partners, allowing the feedback between conditional behaviour and network structure to be studied (Pacheco *et al.* 2008). In this way, Pacheco *et al.* (2008) show that the prediction of breaking links to defectors may not hold. Instead it might be better to maintain links to avoid repeated exploitation by the same individual.

Here, we focus on this second setting. In this way we do not constrain the solution of a social dilemma purely to network reciprocity, but study the impact of network dynamics in light of repeated interactions and the possibility of cooperating via direct reciprocity. This likely constitutes a more natural setting for humans. In our analysis, we focus on assortment (Fletcher & Doebeli 2009) and clustering (Nowak & May 1992) as these are thought to be the most important factors in the evolution of cooperation. From this perspective, we address empirically the question: *does the co-evolution* (in the broad sense of the word) *of cooperative or defective behaviour and network structure really make a difference?* Participants play iterated PDs (see Box 1), and only for dynamic networks they have the

possibility to influence their social relationships based on an active-link-breaking mechanism (Pacheco *et al.* 2006a, 2006b, 2008). Thus, only in dynamic networks can an interaction arise between behaviour and the network, whereas in the static network, cooperation can only be influenced by direct reciprocity. Within this framework, we address the impact of the interaction between behaviour and network by focussing on the following. (1) In relation to theoretical work (see Perc & Szolnoki 2010), we expect rates of cooperative behaviour to be greater in dynamic than in static networks. Moreover, given that our experiment allows conditional behaviour, with respect to link breaking we assess the prediction that individuals should keep links to defectors and reciprocate defection (based on models with conditional behaviour, Pacheco *et al.* 2008), rather than breaking links to defectors (as predicted by models with unconditional behaviour; Fu *et al.* 2009; Wu *et al.* 2010). (2) We characterize topological changes in the dynamic network in terms of cluster formation. (3) We examine the interrelation of individual behaviour and network topology, namely whether participants, not only start to match each other's behaviour within relationships (behavioural assortment), but also assort on the network into clusters (network assortment).

Box 1 ***The prisoner's dilemma***

Within pairwise interactions, reciprocity has been put forward as a mechanism to maintain cooperation. The prisoner's dilemma (PD; Rapoport & Chammah 1965; Axelrod 1984) has been widely used to study the evolution of cooperation (for a recent review, see Doebeli & Hauert 2005). In the PD two individuals simultaneously decide whether to cooperate or to defect. If both cooperate, they each receive a reward (R). If one defects and the other cooperates, the defector gets the temptation payoff (T) and the cooperator obtains the sucker's payoff (S). However, if both defect, they each receive a punishment (P). Furthermore, the assumption T > R > P > S must hold (and in addition, if the game is repeated 2R > T + S). This is summarized by the payoff matrix which we applied in the experiment:

$$\begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} 0.25\,€ & -0.10\,€ \\ 0.40\,€ & 0.00\,€ \end{pmatrix} \end{array}$$

If the individuals cooperate, both do better than if they both would have defected. But for a single individual it is always better to defect no matter what the opponent does. Thus, a social dilemma arises and mutual defection is the dominant outcome in a one-shot PD. However, if the PD is played repeatedly, direct reciprocity (Trivers 1971; Axelrod & Hamilton 1981; Nowak & Sigmund 1992, 1993) is a mechanism for cooperation to be evolutionary stable and supported by experimental evidence (reviewed in Dal Bó 2005).

Materials and Methods

The participants

We tested 200 participants who were recruited from the University of Göttingen via the online recruitment system ORSEE (Greiner 2004) in fall 2009. The students (45% males and 55% females) came from various disciplines and were on average 23.0 ± 2.9 years (mean ± SD) old. Participants were ensured that their decisions were made completely anonymously towards other participants and the experimenters as well as an anonymous payment at the end of the experiment. Throughout the experiment, which lasted c. 90 min, they earned on average 17.64 € ± 4.67. The interaction took place via computers and no other form of communication was permitted.

Static and dynamic network treatment

We ran two treatments: a static network and a dynamic network treatment. Each treatment included 10 sessions (randomly assigned but corrected for sequential and time effects) with 10 participants in each session. The game was played for 30 rounds; however, participants did not know the total number of rounds in order to avoid end-round effects. The static network treatment only consisted of the iterated PD and was played with fixed partners. The dynamic network treatment consisted of a PD stage as well as an active-link breaking stage (cf. Pacheco *et al.* 2006a, 2006b, 2008; see Appendix S1 for more details).

Each participant was linked to three partners and played independently with each partner. In the PD stage the participants were asked to choose between two options (called ORANGE or BLUE option). In half of the sessions orange mimicked cooperation and blue defection, in the other half the reversed pattern was used. Hence, wording like “cooperate”, “defect”, or “collaborate” was avoided to exclude prefixed moral pressure to choose cooperation. The participants were shown the payoff matrix accordingly (see Box 1). After each PD stage the participants were shown their payoffs and the payoffs of their current partners. Thus, the participants knew their total payoff per round. However, they would not receive any information on their partners’ total payoffs just their partner’s payoff with respect to their own interaction with that partner.

In the dynamic network treatment a second stage followed. The participants were asked whether they wanted to continue to play with a partner (indicated by YES or NO decisions). Afterwards, information was given to the participants whether one’s partners wished to continue the relationship or not. If a linked pair agreed to do so, they were also paired in the following round. If, however, at least one of them refused to keep playing, the link was broken off and both received new partners, randomly chosen from all players looking for

partners at that time point. There was a chance of being linked to the same partner again, which was higher if only few players had to be re-linked. The participants were given aliases to ensure anonymity. Thus, they were able to recognize other players by aliases and when meeting a player again, the participants were in the position to recall previous interactions with this partner.

Network topology

For both treatments we used an initial network topology in which all the players had three links. We limited the maximum number of links per player to three because it is reasonable to assume limited resources (e.g. time) for individuals. In the initial network two linked players never share a partner (i.e. there are no clusters) nor does a player have two partners who share another partner (see Fig. 1a). The network remained the same in the static network treatment. The dynamic network treatment started with this initial network, but from thereon links would be determined according to the active-link-breaking stage (see Fig. 1b). The initial position of participants, i.e. the node in the network, was randomly assigned. Moreover, at no point in time did participants have any knowledge of the overall network topology.

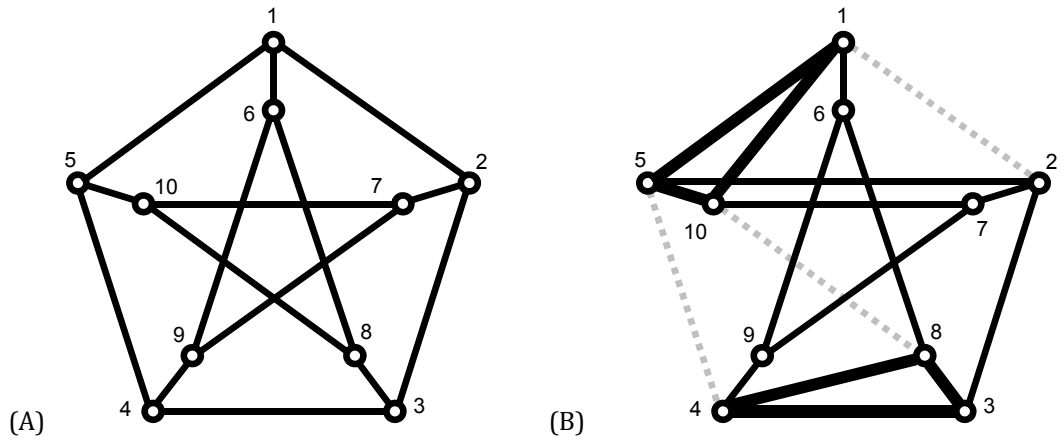

Figure 1 Network topology. Circles represent individuals and lines are links between individuals (i.e. connections between individuals that play iterated prisoner's dilemmas). There are 15 links in total. (a) Graph of the static network treatment and initial configuration of the dynamic network treatment. (b) Example of active-link breaking in the dynamic network treatment (grey dotted lines: former links; bold triangle: cluster).

Statistical analyses

For statistical analyses SPSS 17.0.3 and R 2.10.1 were used. Probabilities are reported as two tailed and a 5% level of significance is used. Furthermore, analyses were done on the group

level, except in the case of the generalized linear mixed models where session effects are considered in terms of random factors. In addition, we developed an agent-based model and ran simulations to assess emergent properties in the dynamic networks.

Results

Participants' game behaviour

Our primary result is that the average cooperation level was significantly greater in dynamic than static networks (see Fig. 2; Mann-Whitney U-test: $U = 4$, $n_{1,2} = 10$, $p < 0.001$; for further analyses see Appendix S2). A difference was already present in the very first round of the PD (average cooperation level, dynamic network treatment: 59.67 ± 9.36%; static network treatment: 48.33 ± 8.64%; Mann-Whitney U-test: $U = 21$, $n_{1,2} = 10$, $p < 0.05$).

In terms of link breaking, we find that participants, irrespective whether they were more cooperative or defective, broke links to defectors, and hence newly established links lasted longer when both participants were cooperative. Although the average break rate of links was 22.90 ± 8.76%, we observed a significant decrease of link breaking over rounds [comparing average link-breaking rates in the first (50.67 ± 8.43%) and last round (10.67 ± 13.16%); Wilcoxon sign-rank test: $T = 0$, $n = 10$, $p < 0.01$]. We used a generalized linear mixed effect model to model the participant's decision to break a link as a function of his or her partner's decision in the PD stage: we included session as well as participant identity nested within sessions as random factors; we assumed binomial-distributed errors; possible time effects were disregarded with all 30 rounds weighted equally. The model revealed a significant impact on the participant's decision to cut the link when his or her partner defected in the previous PD round ($\beta = 3.47$, $SE = 0.10$, $p < 0.001$; see also Fig. S6 in Appendix S2). Finally, we find that if participants met a new partner the link duration was significantly longer if both players cooperated in the first round of a new link than if either of them defected in that round (see Fig. 3; sign test, CC link vs. CD link: $n = 10$, $p < 0.01$; CC link vs. DD link: $n = 10$, $p < 0.01$). The link duration did not differ significantly when one of them defected from when both defected (sign test: $n = 10$, $p = 0.11$).

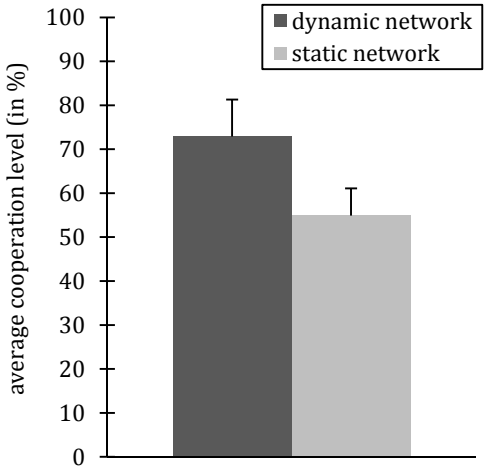

Figure 2 Average cooperation levels (± SD) of 30 rounds of prisoner's dilemmas played either with fixed partners on a static network or with possibly changing partners through an active-linking-breaking mechanism on a dynamic network (Mann–Whitney U-test: $U = 4$, $n_{1,2} = 10$, $p < 0.001$).

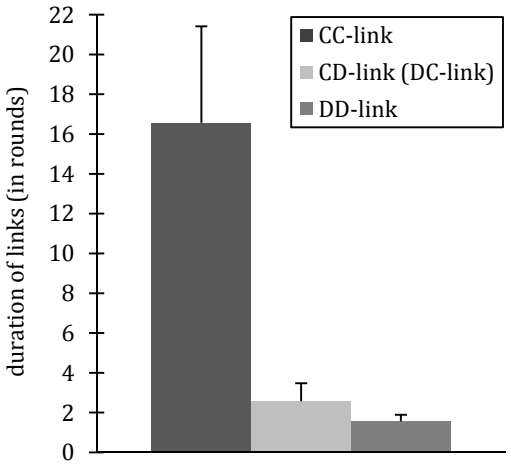

Figure 3 Duration of links in the dynamic network treatment. Bars represent average duration of links (± SD) when paired participants could decide to cooperate, C, or defect, D, in their first prisoner's dilemma round. Accordingly, they either form a CC link, a CD link (DC link, respectively), or a DD link (Sign test; CC link vs. CD link: $n = 10$, $p < 0.01$; CC link vs. DD link: $n = 10$, $p < 0.01$; CD link vs. DD link: $n = 10$, $p = 0.11$).

Assortment on the dynamic networks

To reveal network assortment on top of behavioural assortment within links (cf. the static network treatment with an average cooperation level of 48%), we needed to show assortment into clusters that are beyond the pair level (i.e. we cannot distinguish between behavioural and network assortment at the pair level). Moreover, we needed to use a clustering measure that is independent of cooperative behaviour measures. In this way we could relate cooperation and clustering and reveal assortment of cooperators into clusters.

Clustering in the dynamic networks

We find a greater degree of clustering in the dynamic networks than would be expected at random. To determine this, we devised a clustering score to capture the degree to which individuals were clustered into "cliques" (i.e. clusters, where "your friends are each others friends"; from here on "Friends of Friends" or FoF) and how stable this is over time (see Appendix S3 for details). Next, we compared whether the average FoF score achieved in the experimental sessions (11.01 ± 4.24) differed from the FoF score under random link breaking. To generate an expectation for "random" network clustering, we developed an agent-based model in which links where broken randomly (i.e. not conditional on the

partner's decision in the PD). We ran agent-based simulations based on our experimental sessions (where we used round specific breaking rates measured in the experiment, which accounted for the effect that breaking rates decreased over time; see Appendix S3). We then compared the FoF mean from the experiment to the distribution of FoF scores from the simulations. The FoF mean from the experiment was beyond the top 5% of the distribution of FoF-means obtained from random link breaking (11.01 > 5% threshold of 7.92), demonstrating that the dynamic networks in the experiment indeed became significantly more clustered than would be expected for random link breaking.

Interrelation between game behaviour and network topology

When analysing participants' behaviour in the PD in relation to the cluster formation, we found that it was cooperative participants in particular, who ended up in clusters (for details on the cluster score see Appendix S3). We assigned participants "cooperation" scores by giving them one positive point for every cooperative move towards any partner and one negative point for every defection (theoretically taking values from -90 to 90). Participants' average "cooperation" score was 39.36 ± 43.46 (range: -78–90). We used a generalized linear mix effect model, in which we included sessions as random factors and assumed Poisson-distributed errors, to model cluster scores as a function of the participants' "cooperation" scores. We find that the higher the participants' "cooperation" scores the higher their cluster scores were (intercept = 1.96, $SE = 0.16$, $p < 0.001$; $\beta = 0.0076$, $SE = 0.0008$, $p < 0.001$).

Discussion

In this study we show that for human participants cooperating on social networks, the interrelatedness of behaviour and network structure matters. The level of cooperation in the iterated prisoner's dilemma was significantly increased on dynamic networks relative to static networks. Thus, relative to reciprocity in static relationships, the ability to change partners enhances cooperation.

Theory predicts two possible link-breaking behaviours: (i) keeping links to defectors to keep track of them in a model with conditional PD strategies (Pacheco *et al.* 2008), and (ii) breaking links to defectors, mainly for models with unconditional PD strategies (Fu *et al.* 2009; Wu *et al.* 2010). We find that although our experiment allows conditional behaviour, our link-breaking results more closely match the predictions of unconditional models. Participants broke links to partners who defected much more likely than to partners who cooperated. Hence, links with two cooperative participants lasted much longer on average.

Thus, our results provide experimental evidence for general conditions established in dynamic network models (Pacheco *et al.* 2006a, 2006b; Santos *et al.* 2006a; Fu *et al.* 2009; Wu *et al.* 2010). The most likely reason that our results do not match the prediction of "keeping links to defectors" is that in our experimental setting the number of links per individual was limited, in contrast to Pacheco *et al.* (2008). Thus in our experiment maintaining links to defectors implies a loss of opportunity to be connected to a more cooperative player. It is likely that such opportunity limitations play an important role in structuring the payoffs of behavioural choices in natural settings.

The link breaking in relation to the PD is crucial for network dynamics, because it generates the interaction between behaviour and the network structure. In our experiment, we find that the more cooperative a participant is the greater its cluster score is likely to be. This happens because cooperative links are maintained while links with defectors are broken. Through random re-linking, eventually two cooperative participants are linked and thus became assorted. In fact the assortment occurred in the form of "cooperative cliques", which means that individuals over time become linked to the "friends of their friends". Thus the link breaking and link keeping feeds back on the network structure and thereby defines the social ecology in which individuals find themselves. As a consequence, social structures are generated in which behaviour and network positions are interdependent.

The formation of cooperative clusters is remarkable if one considers that (i) it requires the appropriate type of link-breaking behaviour (see theoretical prediction for keeping links to defectors), (ii) a participant could also assort behaviourally through direct reciprocity (see cooperative outcome in the static networks) and (iii) our participants could never at any moment observe who the neighbours of their neighbours were. Thus, even if people use higher cognitive reasoning within the PD games, such reasoning would not include information on assortment and clustering because these processes occurred beyond the perception of participants. We can therefore only understand the formation of "cooperative cliques" in terms of a self-organized assortment process generated by the interaction between PD behaviour and link-breaking decisions.

The cooperation-enhancing effect of the interaction between behaviour and network structure possibly works at multiple levels. At the behavioural level, we can see that cooperation already increases in the first round. Whether this is because of "a threat of link breaking", "the possibility to get rid of defectors", or "the possibility to stay with like-minded partners" is beyond the scope of this experiment to determine. On the network level, we observe the assortment processes, which allowed cooperative participants to find each other and form clusters. Whether the formation of these "cooperative cliques" then enhances cooperation on top of the assortment in general (i.e. assortment does not necessarily imply

cliques) is impossible to disentangle here. Theoretical work done on static graphs indicates that with clustering, higher levels of cooperation can be reached (Santos *et al.* 2006c). Here we cannot tease apart these different levels of explanations because they are all integrated within the same process. Future work will have to determine how these processes interact in more detail.

Our results could explain why experiments conducted on spatially structured and non-structured static networks have not found a cooperation-enhancing effect of network structures (Cassar 2007; Kirchkamp & Nagel 2007; Grujić *et al.* 2010; Traulsen *et al.* 2010). In our dynamic network treatment, the structure is generated by behaviour of participants, and the participant's position in the network then stands in relation to his or her behavioural tendencies. Hence, the fact that previous experiments impose a network structure may play a role. In such static networks, an individual's position on the network and its behavioural traits do not necessarily have a meaningful relationship and network assortment does not occur. A possible explanation is that people do not simply imitate each other, which is a mechanism that allows assortment in models with static networks (Ohtsuki *et al.* 2006). Another difference is that in our experiment, we do not use the scenario often used in evolutionary game theory on networks (but see Pacheco *et al.* 2008; Do *et al.* 2010) that social interaction decisions are fixed across all links: one has to play the same with all one's partners, which creates a harsher social dilemma. This was the set-up used in the experimental studies of cooperation on static networks (Cassar 2007; Kirchkamp & Nagel 2007; Grujić *et al.* 2010; Traulsen *et al.* 2010). Our result of a cooperation-enhancing effect of network structure is therefore specific for a reciprocal setting. However, given our, and theoretical results (Perc & Szolnoki 2010) we would predict that even if we used the "one strategy to all partners" scenario, it is likely only to find cooperation-enhancing effects of network structure in experiments with humans on networks with dynamism.

In conclusion, we emphasize that the interaction between behaviour and network structure can significantly increase the level of cooperative behaviour in human social networks beyond that of direct reciprocity by itself. Crucial is the biased link breaking, which defines the interaction between behaviour and network. We show that even when individuals could establish cooperation via direct reciprocity (behavioural assortment), there is assortment of individuals on the social network. Such assorted social environments are similar to those suggested to be important for the evolution of cooperation (Nowak & May 1992; Fletcher & Doebeli 2009; Jun & Sethi 2009). Thus, our results strongly support theory that includes co-evolutionary processes and their cooperation-enhancing effects. This fits in a larger tendency to give ecological interactions and feedback, and the self-organizing processes and emergent properties they generate, a more central role in our attempts to

understand evolution (e.g. Boerlijst & Hogeweg 1991; Lion & van Baalen 2008; Nowak *et al.* 2010), in particular that of cooperative behaviour (Fewell 2003; Hauert *et al.* 2006). In addition, our findings may provide a new perspective with which to analyse the vast amount of observational data on cooperative behaviour in social animals and also other behavioural traits that coevolve with network structures and thereby show an ecological interdependence.

Acknowledgements

Discussions with Mathias Franz, Arne Traulsen and Margarete Boos are gratefully acknowledged. We thank the students at the University of Göttingen for their participation. Special thanks to Johannes Pritz and Frederic Nowak for technical support. We thank our three referees for insightful comments. The research is funded by the German Initiative of Excellence of the German Science Foundation (DFG).

APPENDIX TO CHAPTER II

Appendix S1 Experimental set-up

General experimental procedure

Upon arrival participants were randomly seated in front of touch screen computers; they were visually separated by partitions and received written instructions (original German version available from authors upon request). Participants interacted by means of a network computer software developed in our research group (in Java). Through assignment of aliases, i.e. names of moons of our solar system (e.g. Kallisto, Leda, Metis) anonymity was ensured. The aliases could not be connected with the participants' real identities. Initially participants received an endowment of 3.00€. Payment was carried out by using envelopes with the participants' aliases to ensure anonymity (as described in Semmann *et al.* 2005; participants knew this procedure from the written instructions before playing).

In the prisoner's dilemma stage participants could choose between the ORANGE and BLUE option. In half of the experimental sessions orange represented cooperation and blue defection, in the other half the reversed pattern was used. There was no obvious difference in the average total cooperation level on whether cooperation was represented by orange or blue (Mann-Whitney U-test; dynamic network treatment: $U = 8$, $n_{1,2} = 5$, $p = 0.42$; static network treatment: $U = 9$, $n_{1,2} = 5$, $p = 0.55$).

Active link breaking rules

In the prisoner's dilemma stage as well as during link breaking participants could make independent decisions, i.e. to make different decisions for different partners. In the prisoner's dilemma this contrasts theoretical model assumptions where individuals play one strategy against all partners (Pacheco *et al.* 2006a, 2006b, 2008).

In the active-link-breaking stage we assumed for simplicity that all individuals would have the same propensity to look for new links (as in e.g. Pacheco *et al.* 2006a). Furthermore, due to the randomness of receiving new partners, it was possible that some participants were left with only two or one partner. Participants knew these conditions from the written instructions. If a link was not occupied participants received no payoff, which equals the payoff if a linked pair mutually defects (0.00 € each).

Screenshots: decision making during the experiment

During the experiment participants were confronted with different decisions. In the static network treatment participants saw Fig. S1 and Fig. S2 (however, no decisions could be made here). In the dynamic network treatment participants were provided with Fig. S1-S3.

Figure S1 In the prisoner's dilemma stage participants were asked whether to play "orange" (orange, in this particular case cooperation) or "blau" (blue, defection) and had to make one decision for every linked partner.

Figure S2 Participants were provided the outcome of the prisoner's dilemma stage (the third row indicates the participant's payoff and the fifth row the partner's payoff). In the active link breaking stage of the dynamic network treatment participants were asked whether they wanted to keep play with a partner and could answer "ja" (yes) or "nein" (no). They had to make one decision for every linked partner.

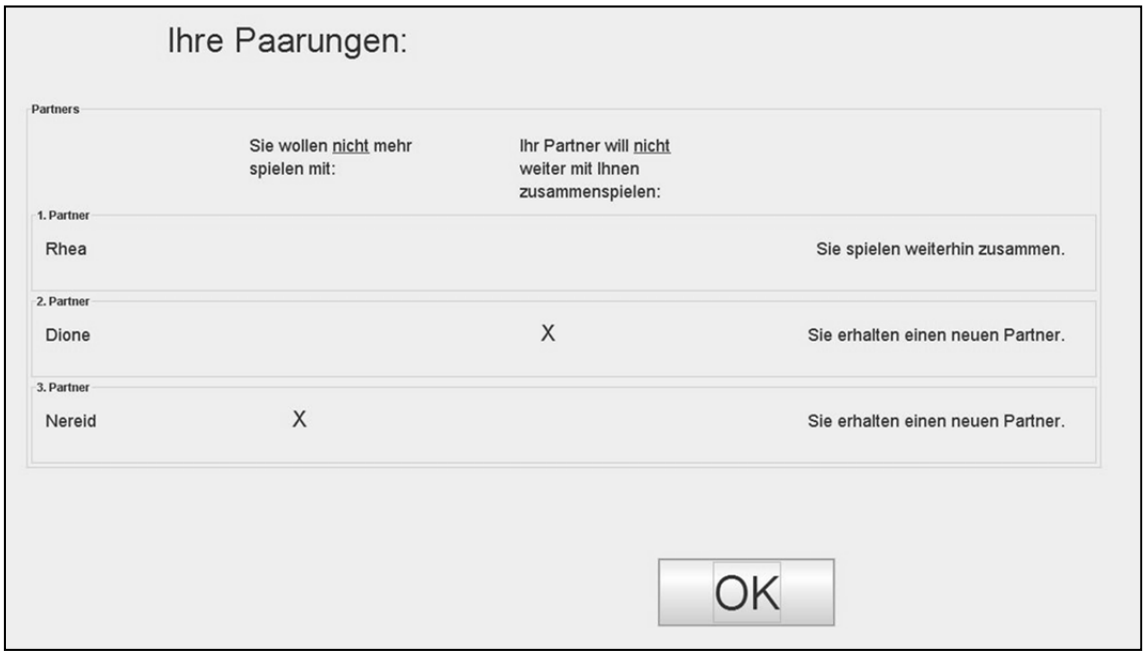

Figure S3 In the dynamic network treatment participants were provided with a summary of the active link breaking stage. Here, the participant continues to play with Rhea (alias); Dione declined to keep playing with the participant; and in the case of Nereid the participant declined to continue the relationship. Thus, the participant would receive two new, randomly chosen partners.

Appendix S2 Additional analyses

The evolution of cooperation

If the frequency of cooperative pairs is found to be high, cooperation is favoured (Pacheco *et al.* 2006a, 2006b, 2008). Thus, the number of links between cooperators should be high in comparison to links including defectors. Our analysis revealed a significantly higher average number of cooperative pairs (8.92 ± 1.43 [mean ± SD]) than pairs with one defector (3.15 ± 0.73; Sign test: $n = 10$, $P < 0.01$) and than defective pairs (2.36 ± 1.01; $n = 10$, $p < 0.01$). There was no difference in the average number of pairs when one of them defected and when both defected (Sign test: $n = 10$, $p = 0.11$).

Additionally, earnings were significantly greater in dynamic networks (average final payoff, dynamic network treatment: 19.22€ ± 1.73; static network treatment: 16.05€ ± 1.39; Mann-Whitney U-test: $U = 8$, $n_{1,2} = 10$, $p < 0.001$) where the cooperative norm was more prevalent.

Results of the computerized post-questionnaire

After participants played the iterated prisoner's dilemma arranged on a static or dynamic network, they completed a short computerized questionnaire concerning their motivation to

participate and their decisions during the game (see Fig. S4 – S7). Pooled answers of both treatments are presented, unless the question was only asked in the dynamic network treatment.

Our participants stated to be strongly motivated to earn money during the experiment (see Fig. S4). This goal can best be achieved by obtaining the temptation payoff of 0.40 € (cf. Box 1). Nevertheless, especially our participants from the dynamic network treatment reached high levels of cooperation. In addition, participants seemed to follow reciprocal strategies, as they answered that their decisions to cooperate or to defect were conditional on previous decisions of their partners (see Fig. S5).

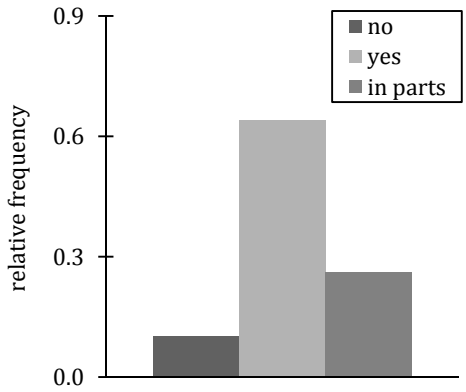

Figure S4 Answers of participants of the static and dynamic network treatments to "I have tried to earn as much money as possible" (n = 200).

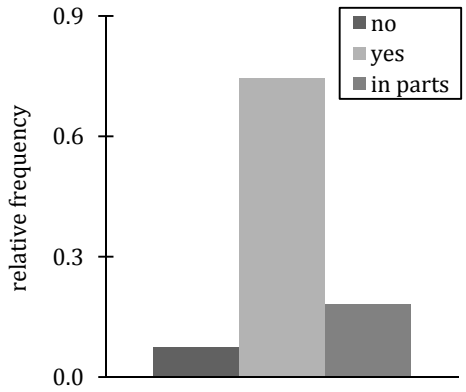

Figure S5 Answers of participants of the static and dynamic network treatments to "Which strategy (blue, orange) I played, depended a lot on the previous decisions of the respective partner" (n = 200).

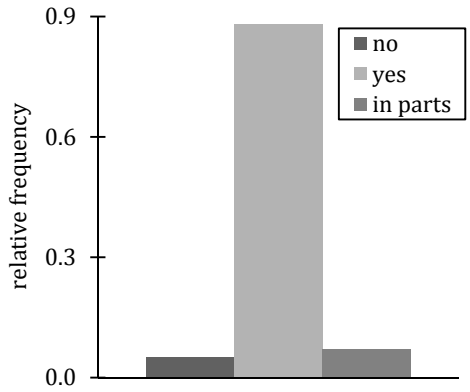

Figure S6 Answers of participants of the dynamic network treatment to "Whether I continued to play with a partner depended a lot on his previous decisions" (n = 100).

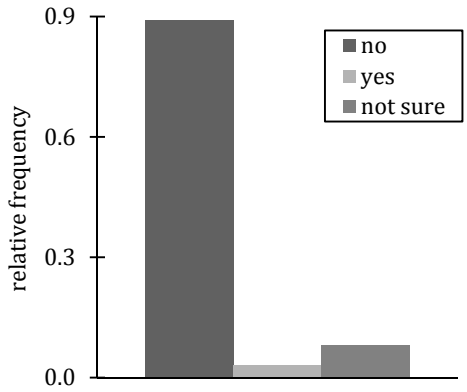

Figure S7 Answers of participants of the dynamic network treatment to "Have you played with Oberon at least once?". The alias "Oberon" was never given to any participant, thus the answer should be "no" (n = 100).

A great majority (88 %) in the dynamic network treatment answered to have used the previous decisions of a partner to determine whether they would keep or break this link (see Fig. S6). Furthermore, we asked participants in the dynamic network treatment whether they could remember to have played with a particular partner. We presented them different aliases; one of them was a moon name never used as a participant's alias during the experiment (i.e. Oberon). Thus, if participants made use of aliases to consider whether they have met this particular partner before and then which decisions to make in the prisoner's dilemma, then they should be able to recall that they have not been paired with the non-existing player. In line with this assumption, only 11 participants out of 100 considered to have played with the non-existing player (see Fig. S7).

Appendix S3 Dynamic networks: Clustering and agent-based simulations

Clustering Score

In our assessment of network clustering we considered a clustering score per individual in terms of whether "your friends are each others friends" (triangular motifs in the network; from here on "Friends of Friends" or FoF). We calculated FoF as:

$$FoF_{ir} = \sum\sum L_{ijr} L_{ikr} L_{jkr}$$

where individual i can be linked to individuals j, and k in round r. The link status is represented by L. $L_{ijr} = 1$ when individuals i and j have been linked for more than one round in round r, otherwise $L_{ijr} = 0$. Per round FoF can therefore be maximally 3 (all neighbours are each other's neighbours). We then cumulatively sum FoF scores over all rounds. Thus the maximal cluster score is 87, where individuals are maximally clustered on every round. Because we only count clustering when all the links defining them are older than one round, the stability of clusters in time becomes important for the clustering score. As long as a cluster is stable we increment the score, but as soon as one link is broken any new cluster has to be in existence for at least one round for the score to increase again. Thus the breaking up of clusters will decrease the clustering score. Individuals with high cluster scores will therefore be those interacting within "cliques" that are stable over time.

Clustering under random link breaking: agent-based simulations

We ran simulations with a simple model implemented in the C programming language. Each simulation emulated a single experimental session of 10 individuals each with a maximum of

3 links over 30 rounds (time steps) of prisoner's dilemma games. Simulations were initialized with the same initial network configuration as the experiment (see Fig. 1). Links were broken randomly, and individuals were re-linked as in the experiment: any individuals with fewer than 3 links was re-linked randomly to another individuals with less than 3 links until there were no more individuals that could be linked.

To compare the simulation results with experimental conditions we simulated the model using the average link breaking rates of 30 rounds over time obtained from each experimental group of the dynamic network treatment. These breaking rates decreased over time, which could in itself affect the stability of network clusters (more stable with lower breaking rates). We took this into account by including this decrease in the simulations directly. Thus, decisions with whom to break links were random and actual break rates were pre-fixed. We then simulated 1000000 "experiments". To calculate an overall "experiment" average (identical to the experimental data), we calculated one average FoF score from 10 random selected simulated groups giving us 100000 "experiment" data points for the average FoF score under random link breaking. From this we obtained the distribution of clustering scores expected for random link breaking (7.35 ± 0.38), and thus could determine whether the average clustering score obtained from our experimental groups was significantly greater than would be expected under random link breaking. In this way we reveal the impact of active link breaking on individual neighbour selection and network structure.

CHAPTER III VENDETTAS OF COSTLY PUNISHMENT

I DARE YOU TO PUNISH ME – VENDETTAS IN A GAME OF COOPERATION

with Dirk Semmann[1], Ralf Sommerfeld[2], Jürgen Krambeck[2], and Manfred Milinski[2]

[1] Courant Research Center Evolution of Social Behavior, University of Göttingen, Germany

[2] Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Plön, Germany

In preparation

Abstract

Everybody has heard of neighbors, who have been fighting over some minor topic for years. The fight goes back and forth, giving the neighbors a hard time. This kind of reciprocal punishment is known as vendettas and is a cross-cultural phenomenon. In general, punishment is seen as a mechanism for maintaining cooperative behavior. However, this kind of punishment excludes vendettas a priori. Vendettas pose a special kind of evolutionary problem, since they incur high costs on individuals, i.e. costs of punishing plus costs of being punished, without any benefits. Theoretically speaking vendettas do not evolve under natural selection. However, we find that under experimental conditions human participants retaliated frequently and contradictory to theory even engaged in cost-intense punishment vendettas, especially when punishment was unjustified or rather ambiguous. Punishment was targeted at defectors in the beginning, but soon provocations led to mushrooming of counter-punishments. Remarkably, participants were, nevertheless, able to enhance cooperation levels in a public goods game. Some participants even seemed to anticipate the outbreak of costly vendettas and delayed their punishment to the last possible moment. Overall, results indicate that current evolutionary models fail to consider an important aspect of interactions, which conjecturally include concerns of equity and reputation.

Keywords

Cooperation, evolutionary game theory, punishment, public goods game, revenge, vendetta

Introduction

Many species, especially humans, frequently cooperate and provide help to each other (for recent reviews, see e.g. Pennisi 2009; Melis & Semmann 2010). Cooperative behavior prevails despite theoretical problems explaining its evolution. That is why cooperate if one could enjoy the benefits provided by others and refrain from costly cooperative behavior oneself? This is the so-called free-rider problem (Dawes 1980). One hotly debated mechanism to prevent free-riding is punishment – a widely spread behavior among humans and animals (for reviews, see Clutton-Brock & Parker 1995; Sigmund 2007; Jensen 2010; Milinski & Rockenback in press). However, punishment can escalate into vendettas where "I punish you, because you punished me; but you already punished me, because I punished you before", and so on. How can punishment then be beneficial for cooperation?

Punishment is understood as a behavior that has costs for the opponent and somewhat lower costs to the punishing individual itself. As punishment is costly, there is no incentive to do so. This situation is analogous to the free-rider problem of cooperation, whereby non-punishers represent second-order free-riders (Boyd & Richerson 1992). The second-order free-rider problem has been investigated intensively and under certain conditions punishment is evolutionary stable (Henrich & Boyd 2001; Boyd *et al.* 2003; Brandt *et al.* 2003; Gintis *et al.* 2003; Fowler 2005; Hauert *et al.* 2007). Moreover, an extensive amount of experimental research shows that humans employ costly punishment and that thereby cooperation is enhanced (e.g. Yamagishi 1986; Ostrom *et al.* 1992; Fehr & Gächter 2002; Rockenbach & Milinski 2006; Egas & Riedl 2008; Gächter *et al.* 2008; Herrmann *et al.* 2008; but see Wu *et al.* 2009). Even symbolic gestures of punishment (Masclet *et al.* 2003) and the mere threat of punishment (Fehr & Gächter 2002) raise cooperation levels. However, earnings are usually negatively affected, because the costs of punishment cannot be compensated by higher cooperative benefits (Ostrom *et al.* 1992; Fehr & Gächter 2002; Egas & Riedl 2008). On the other hand, if interactions last very long, negative effects of punishment costs can be overcome at the group level (Gächter *et al.* 2008).

Previous research in the area of costly punishment has mainly concentrated on situations where punishment cannot be retaliated (e.g. Henrich & Boyd 2001; Fehr & Gächter 2002). Under most natural conditions this is not true, i.e. usually punishment can be avenged by victims. One only needs to look at the epic dramas that have been described in history and

literature. For instance, Shakespeare (1954) wrote about it in Romeo and Juliet where the Montague and Capulet families were deeply involved in a vendetta. Vendettas are a cross-cultural phenomenon (Ericksen & Horton 1992). There are blood vendettas between Turkish farmers lasting as long as 60 years (İçli 1994). Vendettas occurred in the Mediterranean area in the nineteenth-century (Gould 2000) and they proliferate in science (Hellman 1998, 2006). Sometimes these vendettas escalate and then one reads headlines like "A 20-year feud between two neighbors [...] revved up this week, ending in bloodshed" (The Local 2010). These yearlong vendettas often begin with a punishment of one party, which is perceived as unjustified by the victim (Stillwell *et al.* 2008), and turn into a more serious conflict.

Recently, there has been growing interest in the effect of retaliation on cooperative games with punishment (Denant-Boemont *et al.* 2007; Dreber *et al.* 2008; Nikiforakis 2008; Nicklisch & Wolff 2009). They show that humans avenge punishment regardless of its negative effect on payoffs. However, in most cases cooperation cannot be sustained by revengeful punishment. Up to now the possibility that punishment can escalate into vendettas has been disregarded by restricting punishment to a single retaliation stage (Denant-Boemont *et al.* 2007; Janssen & Bushman 2008; Nikiforakis 2008). In other works, the focus lays on other topics neglecting the analysis of possible escalations (Denant-Boemont *et al.* 2007; Dreber *et al.* 2008; Nicklisch & Wolff 2009). Only Nikiforakis and Engelmann have explicitly studied vendettas, but the interpretation of their results remains unclear due to discrepancies between two paper versions (Nikiforakis & Engelmann 2008, 2011). Their participant samples comes from different countries (i.e. London, UK and Melbourne, Australia; which include differences in the experimental procedures, e.g. group size differences), but vendettas only occur in Melbourne (which is not explicitly stated in the 2011 version). This, however, seems crucial, as the occurrence (or non-occurrence) of vendettas goes along with different results and conclusions about the effect of escalating punishment on cooperation levels and earnings for the London-sample (2008) and the combined samples (2011). Hence, while reporting no country differences for first-stage punishment, country differences for vendettas are not discussed even though vendettas are their main focus. Hence, the issue of whether humans engage in costly punishment, which can escalate into vendettas, and how cooperation is affected remains unanswered.

Despite the vengefulness observed in humans, theoretical research shows that vendettas of punishment are not an evolutionary stable behavior. In repeated interactions the best respond to an opponent's defection is defection and not punishment. On top of that, defection is preferred as a response to an opponent's punishment (Rand *et al.* 2009b). In other words, one expects little punishment and no vendettas. Furthermore, though a single

stage of retaliation is not enough to stabilize cooperative behavior (Janssen & Bushman 2008), it can be beneficial for the evolution of cooperation based on a conformism bias to have more than three opportunities to punish and punish back (Wolff 2009). However, in either case individuals should abstain from counter-punishing and let the mere threat operate. In line with this, concealing the punisher's identity, and thus making retaliation harder, has positive effects on cooperation (Janssen & Bushman 2008). Therefore from a theoretical point of view, one does not expect to find vendettas of punishment in cooperative games.

In this study we allow for vendettas by combining the public goods game (PG; Hardin 1968; Ledyard 1995) with multiple rounds of costly punishment. In this setting, we can investigate the occurrence of vendettas, as it is more realistic to assume that victims can punish their punisher in the same way immediately or later. Rational choice theory, however, assumes that people should take this behavior into account. They should reason that if they punish others and when there is the possibility to be punished back, they not only will have to pay to punish, but also the fine imposed on them due to being punished by their victim. This leads to exaggerated costs of punishment that should be avoided by the rational individual. Following this logic, if there is no punishment, then there is no incentive to invest in the PG anymore. Nevertheless, studies of costly punishment where vendettas are impossible show that people do indeed engage in punishment, which then stabilizes PG contributions (e.g. Fehr & Gächter 2002). Therefore, albeit the high costs of possible vendettas, we expect participants to engage in punishment. Additionally, we also anticipate the occurrence of vendettas, as they are observed in the real world. Subsequently, it will be highly interesting to see how cooperative behavior and overall payoffs in the PG will be affected.

Methods

First-semester biology students from the Universities of Kiel, Hamburg and Münster, Germany, as well as Vienna, Austria, joined the experiment voluntarily. A total of 96 participants were randomly assigned into 6 sessions of 16 participants each. In each experimental session, participants were randomly seated in front of an individual computer with partitions between participants. In order not to disclose their real identity, but still allow for individual recognition within the game, participants received an alias at the beginning of the game. Participants were told that they have to make decisions during the experiment whether to invest their money or not in different situations. A short introduction ensured that the participants understood how to handle the computer, that they are

completely anonymous throughout and after the game concerning their behavior within the experiment, that they should not talk to one another or draw attention to them during the experiment and that they will receive all their earnings anonymously in cash. After the experiment, each participant could collect her earnings out of an envelope entitled with her alias from behind partitions (as described in Semmann *et al.* 2005). Thus, the participant herself was the only person who knew her identity in the experiment. The experimental sessions lasted about 60 minutes and participants earned on average 13.41€ ± 6.92 (mean ± SD).

Our experimental design follows the one of Fehr and Gächter (2002) where the participants were arranged into subgroups of four individuals each, and first played a PG round followed by the possibility to punish other members of the subgroup. The starting money was set to 20€ for each participant. In the PG situation, participants had to decide whether or not to contribute 1.00€ to a PG. They were informed that the sum of all contributions will be multiplied by 1.6 and distributed equally among all subgroup members irrespective of their contribution (0.00 or 1.00€; this is in contrast to the continuous contributions in Fehr & Gächter 2002). With a group size of four, this results in a marginal PG payoff of 0.4. In the following punishment round, participants were informed about the PG investments of all subgroup members and could then assign a punishment from 0 to 10 units (each unit represents 0.10€) for each subgroup member separately. Following again Fehr and Gächter, each point of punishment assigned resulted in a threefold fine to be paid by the punished subgroup member. If for instance a player invested 0.30€ (= 3 units) to punish somebody, the account of the punished member was reduced by 0.90€ (= 9 units).

The difference to the previous study is that instead of just one, a sequence of five punishment rounds was played after the initial PG round. In these successive rounds participants were provided with complete information about previous punishment investments of all subgroup members. Thus, in addition to the others' PG decisions, they knew exactly who punished whom with how much money for each previous round. In each experimental session 16 participants played the mentioned sequence of rounds (PG followed by five punishment rounds) three times (= three periods). Between each period, the participants were reshuffled into new subgroups of four individuals in a way that excluded any kind of reputation building and direct reciprocity between periods. No participant, being aware of this condition, did ever meet a previous subgroup member in later periods again.

For statistical analyses SPSS 18.0.2 and R 2.12.1 were used. A 5%-level of significance is used and probabilities are reported as two tailed. Furthermore, analyses were done on the session level, if not stated otherwise. Exceptions are the generalized linear mixed models where session effects are considered in terms of random factors.

Results

Cooperation in the public goods game

As the research design is adopted from Fehr and Gächter (2002), we also applied their statistical analysis where this is feasible. Our results show that the level of cooperation in the PG round increased (comparing period 1 vs. period 3; Wilcoxon signed-rank test: $Z = 2.21$, $n = 6$, $p < 0.05$; see Fig. 1).

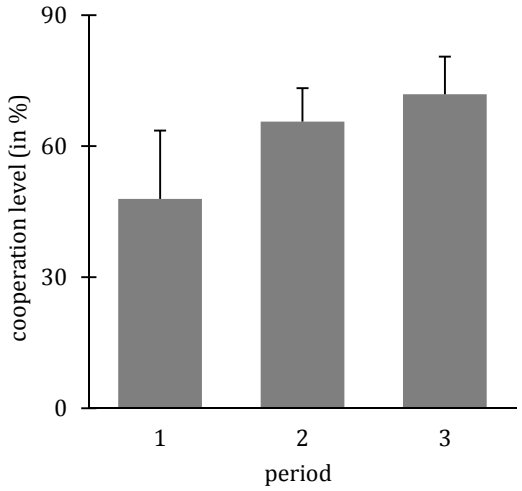

Figure 1 Average cooperation in the public goods rounds (+ SD). Cooperation levels significantly increased over time (comparing period 1 vs. period 3; Wilcoxon signed-rank test: $Z = 2.21$, $n = 6$, $p < 0.05$).

Punishment after the public goods round

Punishment was frequent. In overall 15 rounds of punishment and with the possibility to punish up to three subgroup members, 85.4% of the participants punished at least once; 52.1% at least five times; and 21.9% at least 10 times. Within period 1 investment in punishment did not change over the five rounds of punishment (see Fig. 2; Friedman test: $\chi^2 = 2.12$, $df = 4$, $n = 6$, $p = 0.71$). However, we found significant changes in periods 2 and 3 (Friedman test: period 2, $\chi^2 = 11.42$, $df = 4$, $n = 6$, $p < 0.05$; period 3, $\chi^2 = 14.08$, $df = 4$, $n = 6$, $p < 0.01$). In period 1, participants did not yet know the total number of rounds played in each period, afterwards they could guess. In periods 2 and 3, we observed an increase in punishment investment in the very last round. To analyze this last round effect, we compared punishment in the last and the second-last round. The respective differences were significantly different in period 2 (Wilcoxon signed-rank test: $Z = 2.20$, $n = 6$, $p < 0.05$) and we found a trend in period 3 ($Z = 1.58$, $n = 6$, $p = 0.12$). Further analysis revealed that the high punishment investment in round 5 was due to few participants (in each period: 10 out of 96), who invested high amounts to punish (period 2: 0.85€ ± 0.23; period 3: 0.91€ ± 0.19).

These participants revenged their punishment of round 4 (period 2: 30%; period 3: 26%), but also delayed their revenge of being punished in rounds 1 to 3 (period 2: 40%; period 3: 47%).

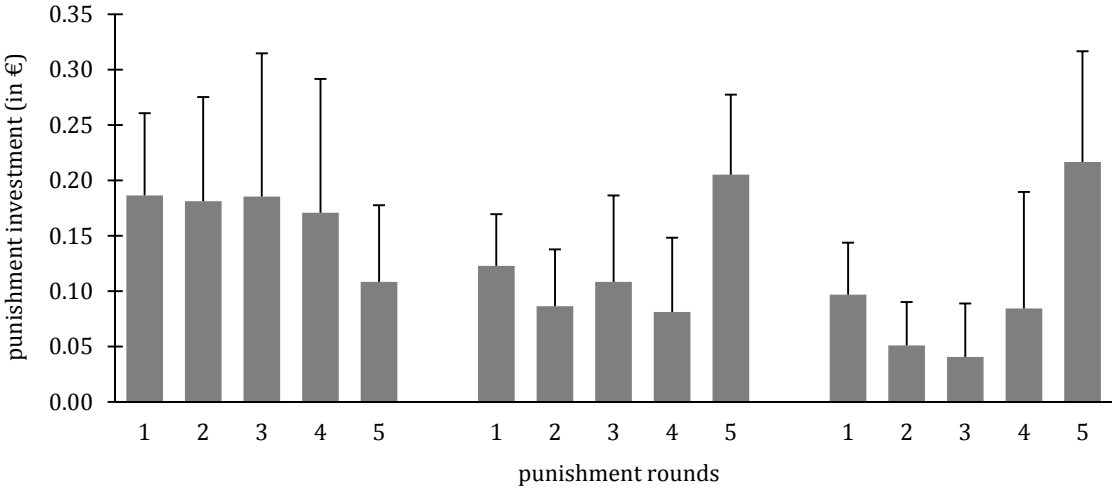

Figure 2 Average punishment investment (+ SD) per player. In each of the three periods participants played one round of public good followed by five rounds of punishment.

Multiple rounds of punishment allow participants to punish back after receiving a fine. Indeed on average up to 80% retaliated their punishment (period 1: 0.80% ± 0.09; period 2: 0.58% ± 0.20; period 3: 0.42% ± 0.17). Within acts of punishment there was a significant relationship between punishment investment by the punisher and counter-punishment investment by the opponent. We used generalized linear mix effect models (GLMMs), in which we included *punisher identity* and *counter-punisher identity* nested within *sessions* as random factors, to model the received *counter-punishment* (0 to 10 units) in the current round as a function of *original punishment* (1 to 10 units) in the previous round. GLMMs were fitted by Laplace approximation assuming Poisson error distribution. We found that the higher the original fine the higher the counter-punishment (round 1 – round 2: intercept = -0.85, SE = 0.25, $p < 0.001$; β = 0.15, SE = 0.06, $p < 0.05$; round 2 – round 3: intercept = -0.61, SE = 0.24, $p < 0.05$; β = 0.18, SE = 0.06, $p < 0.01$; round 3 – round 4: intercept = -0.75, SE = 0.21, $p < 0.001$; β = 0.23, SE = 0.04, $p < 0.001$; round 4 – round 5: intercept = -1.03, SE = 0.30, $p < 0.001$; β = 0.17, SE = 0.04, $p < 0.001$).

To analyze the motives of participants to punish we used GLMMs to model *punishment* (0 to 10 units) as a function of *participant's and opponent's PG decisions*, *subgroup members' PG decisions*, and *provocation* (i.e. in punishment rounds 2 to 5 the punishment investment by the opponent in the previous round). We controlled for differences in *periods* and in *participants*, who are nested within *experimental sessions*, and included these as random

factors. We looked at the given models for each punishment round separately; hereby allowing motives for punishment to differ between rounds. GLMMs were fitted by Laplace approximation assuming Poisson error distribution. The variance inflation factors are all less than 1.25, which indicates that multicollinearity is not a problem in the models' estimations (Greene 2008). In punishment round 1 the participant's and her opponent's behavior in the PG predicted the punishment investment of the participant, i.e. if both contributed then punishment became less likely (see Tab. 1). In subsequent rounds of punishment the importance of the PG behavior varies. However, the behavior of the two other subgroup members is now important, as the more of them contributed, the more likely punishment of the remaining subgroup member became. In addition, the previous amount of punishment by the opponent significantly increased investments by the participant to punish the opponent in the current round.

Table 1 Results of the generalized linear mixed models to model punishment investment.

	round 1	***round 2***	***round 3***	***round 4***	***round 5***
intercept	-3.36 *** (0.33)	-3.45 *** (0.43)	-3.64 *** (0.49)	-3.52 *** (0.38)	-2.97 *** (0.35)
P contributed and O did not contribute into the PG [1]	2.55 *** (0.20)	0.71 ** (0.25)	1.07 *** (0.24)	-0.20 (0.24)	-0.02 (0.17)
P did not contribute and O contributed into the PG [1]	0.72 ** (0.27)	0.82 *** (0.22)	0.83 *** (0.23)	0.39 (0.23)	-0.47 *** (0.14)
P and O did not contribute into the PG [1]	1.33 *** (0.26)	0.72 ** (0.26)	0.84 *** (0.24)	-0.61 * (0.28)	-0.17 (0.18)
other two subgroup members' behavior in PG [2]	0.17 (0.10)	0.22 * (0.11)	0.28 ** (0.11)	0.47 *** (0.12)	0.30 *** (0.09)
provocation	n/a	0.48 *** (0.04)	0.42 *** (0.03)	0.43 *** (0.03)	0.20 *** (0.02)

Note: Provided are the estimates, the standard errors in brackets and the p-values as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The *period*, the *participant's identity* and the *experimental session* were added as random factor in all models ($n = 864$, in each round 96 participants could punish up to 3 subgroup members in 3 periods). For punishment in round 1 no previous provocation (in terms of punishment investment by the opponent in the previous round) is possible.

[1] The contribution of both, the participant (P) and her opponent (O), into the public good (PG) served as reference group of the categorical fixed factor *participant's and opponent's PG decisions*.

[2] The behavior of the remaining two subgroup members was coded as 0, 1, or both contributed into the PG.

In line with the results from the GLMMs for punishment round 1, punishment was directed at non-contributing (i.e. defecting) participants. In particular, contributors, who punished defectors, spend the most money on punishment (see Fig. 3; punishment significantly differs between outcomes of PG behavior of punisher and opponent: Friedman test, $\chi^2 = 12.2$, $df = 3$, $n = 6$, $p < 0.01$). In all subsequent rounds the punishment investment

did not differ according to the PG behavior of the punisher and the opponent (see *Supplementary Information*, Fig. S1). This is in line with the GLMMs, as they showed that now the behavior of other subgroup members and provocations gained importance.

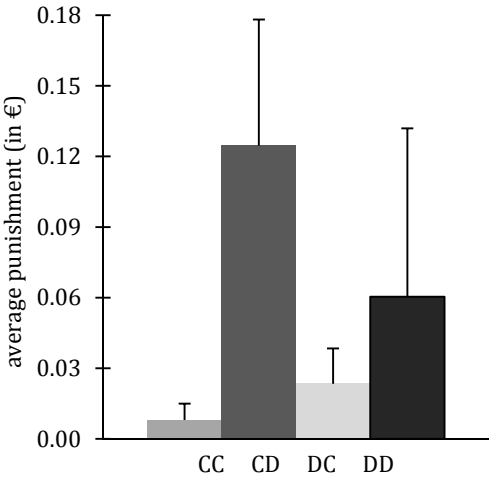

Figure 3 Average punishment investment (+ SD) in the first round of punishment (pooled over all periods). Participants could either contribute into the public good, C, or defect, D. Hence, in CD a contributor punished a defector (CC, DC, DD, respectively; Friedman test: $\chi^2 = 12.2$, $df = 3$, $n = 6$, $p < 0.01$).

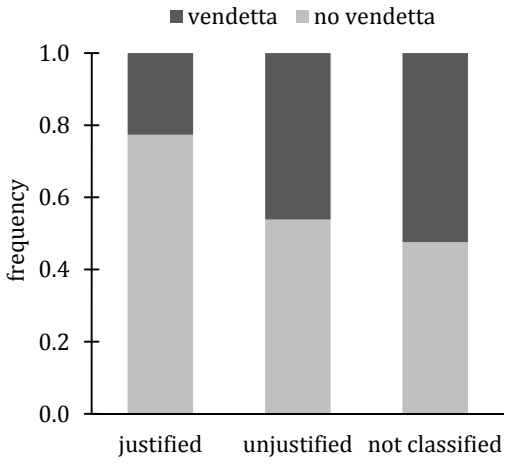

Figure 4 Frequencies where a participant punished a subgroup member in punishment round 1 and either a vendetta or no vendetta occurred (pooled over all periods). Punishment was classified as justified if a contributing participant punished a non-contributor ($n = 106$, individual level); it was termed unjustified if a non-contributing participant punished a contributor ($n = 26$); all other cases were rather ambiguous and not further classified ($n = 42$).

Vendettas of costly punishment

A minimum of three sequential punishments was defined as a vendetta, i.e. player A started by punishing player B, who retaliated this punishment, and was again punished by player A in the next round. In the experiment, we observed 71 vendettas in total. On average participants were involved in 1.48 ± 0.88 vendettas and vendettas lasted on average 3.89 ± 0.39 rounds. In addition, a clear pattern arises when looking at punishments in round 1 and whether a vendetta developed or not on the level of individual interactions (see Fig. 4). Justified punishment of a non-contributor by a contributor was most frequent, but did not lead to vendettas in most cases (77%). All other punishments, i.e. unjustified punishment of a contributor by a non-contributing participant and ambiguous punishment (a contributor punished a contributor; or non-contributor punished a non-contributor), triggered a vendetta in about 50% of the time. Those players engaging in vendettas pay large

costs, since it includes their punishment investment and counter-punishment fines. Comparing average payoffs of participants that were involved in vendettas (10.44€ ± 4.25) and participants that were never involved in a vendetta (i.e. neither started one nor did counter-punish that resulted in a vendetta; 17.05€ ± 1.29) showed, that the latter earned significantly more money (sign test: $n = 6$, $p < 0.05$). This is also true for players retaliating punishment (11.87€ ± 2.84) versus players refraining completely from retaliating (17.15€ ± 1.34; sign test: $n = 6$, $p < 0.05$).

Discussion

Just as vendettas occur under natural conditions (Ericksen & Horton 1992; İçli 1994; Gould 2000), so did the participants in our experiment of a public goods game with five rounds of punishment frequently engage in vendettas. This happened even though vendettas are cost-intense, as one has to pay costs for punishing and costs for being punished; multiplied by several instances. Nevertheless, we find that participants frequently retaliated punishment and that each participant was involved in on average 1.5 vendettas, i.e. at least three sequential rounds of punishment, during the experiment. Despite the costliness of vendettas (i.e. they significantly reduced earnings compared to players, who abstained completely from vendettas) we observed an-eye-for-an-eye counter-punishment. This supports the view that counter-punishment possibly escalating into vendettas is due to an attempt to restore equity between participants (Adams 1965; Fehr & Schmidt 1999; Dawes *et al.* 2007; Stillwell *et al.* 2008; Brosnan *et al.* 2010). In the example of fighting neighbors, both see themselves as victims and both go on to restore (subjective) justice. The durations of vendettas were rather long. In fact, participants' vendettas lasted on average about four out of five rounds. Vendettas normally started with an unjustified punishment (i.e. a non-contributor punished a contributor), or when the meaning of the punishment was rather ambiguous (i.e. a contributor punished a contributor; or non-contributor punished a non-contributor). When the punished individual had defected and was "properly" punished by a cooperative participant then vendettas seldomly started (i.e. only in 23% of all cases). The initial social interaction of the PG was relevant for the first punishment, i.e. defectors attracted the highest punishment. In later rounds players primarily reacted to provocations (previous punishment). In addition, participants relied on the behavior of other subgroup members as a social reference point: the more those cooperated the more likely the remaining subgroup member deserved punishment.

Despite the frequency of costly punishment, retaliations, and even vendettas cooperativity increased over time. This occurred even though direct reciprocity and

reputation building between PG rounds was excluded. Results contradict earlier findings of revengeful punishment where cooperation is not sustained (Denant-Boemont *et al.* 2007; Nikiforakis 2008). The increase of cooperation is presumably due to the effect of the first punishment round where high amounts of punishment were targeted at defecting participants. Punishment of non-contributors as a direct response to their defection (though without the possibility of escalating punishments) is also observed in previous studies (Fehr & Gächter 2002; Egas & Riedl 2008). Nevertheless, in experiments earnings are usually negatively affected (e.g. Fehr & Gächter 2002; Egas & Riedl 2008), which is especially true for participants, who engaged in retaliation and vendettas. This makes punishment as a mechanism to solve the free-rider problems in PG situations unlikely (Dreber *et al.* 2008). Furthermore, just as we so do experimental studies report of unjustified punishment, which in general has been termed anti-social punishment (Wu *et al.* 2009; Denant-Boemont *et al.* 2007; Nikiforakis 2008; Dreber *et al.* 2008). However, the evolution of cooperation is not supported in the presence of anti-social punishment (Gächter *et al.* 2010; Rand *et al.* 2010). In our study, anti-social punishment acts frequently led to vendettas, making the original unjustified or anti-social punishment very costly. Given that punishment can escalate, this could serve as means to reduce anti-social punishment to a minimum over time.

Remarkably, by quickly adjusting to the given experimental set-up some participants were able to avoid costly vendettas (as soon as they could guess the number of rounds in later periods). As an indication for avoiding retaliation and the possible anticipation of outbreaks of costly vendettas, is the behavior of some participants, who delayed their punishment to the very last round. In addition, these participants invested high amounts to punish, indicating a final revenge for being punished in previous rounds where they patiently refrained from immediate counter-punishment to avoid the danger of paying counter-punishment fines.

Our results are in accordance with earlier findings that humans are willing to punish and retaliate (e.g. Denant-Boemont *et al.* 2007; Egas & Riedl 2008; Nikiforakis 2008; Jensen 2010). We extended this line of research by showing that acts of punishment can escalate into vendettas. However, the behavior of our participants is in contradiction to theoretical postulations that vendettas should not occur under natural selection (Janssen & Bushman 2008; Wolff 2009), as defection is the proper response evolving after provoking punishment (Rand *et al.* 2009b). Nevertheless, a tendency to avenge can also be found in animals (Clutton-Brock & Parker 1995; Jensen *et al.* 2007). For instance, Japanese macaques sometimes use indirect revenge against an aggressor's kin (Aureli *et al.* 1992). These counter-aggressive acts seem to have regulatory effects, as they happen in the presence of the aggressor, who however is unable to intervene, and thus these acts can serve as means to

reduce the likelihood of further attacks of the aggressor against the revenging individual. Vendettas in human societies are also attributed a functional quality (Elster 1990; Gould 2000). For one, vendettas are thought to provide rules for escalating conflicts and thereby they might reduce the likelihood of full escalation. Additionally, social norms prescribe which kind of behavior is to be avenged. Here, we also found that vendettas occur only under certain circumstances: after unjustified or ambiguous punishment, but rarely after justified punishment. Such counter-punishments could relate to social norms like "showing strength" or "avoiding to losing face". Furthermore, "natural" vendettas occur more frequently in regions where the institutional law is rather weak or absent (Elster 1990). Considering real-world observations, we find it worthwhile to investigate multiple rounds of punishment in an experimental setting where punishment can be peer-based, but in addition institutionalized. Due to assured institutionalized punishment, peer-punishment might become less important, resulting in a reduced likelihood of vendettas. That cooperation is promoted by institutional punishment has been shown theoretically (Sigmund *et al.* 2010), but whether this inhibits vendettas on the peer-level remains to future research.

The aim of our study was to examine whether vendettas of punishment occur, and to test how the existence of vendettas then would affect cooperation. Here, punishers are not protected, in the sense that they had to take on the consequences of their punishing behavior. As punishment is costly and negative for individuals (Dreber *et al.* 2008), it is better to abstain from it. This is especially true in our setting where punishment could escalate into cost-intense vendettas. Nevertheless, our participants engaged quite frequently in vendettas. In conclusion, these results indicate that evolutionary models so far neglected important aspects of real-life interactions, like equity and reputational concerns, as animals and humans frequently retaliate and as vendettas occur across human societies.

Acknowledgements

We are grateful to M.A. Nowak, H. Brendelberger, T.B.H. Reusch, J. Marotzke, F.G. Barth, and M. Bähler for their support. We thank the students of the Universities of Kiel, Hamburg and Münster, Germany, as well as Vienna, Austria for their participation in this study. D.S. and K.F. are founded by the German Initiative of Excellence of the German Science Foundation (DFG).

APPENDIX TO CHAPTER III

Average punishment investment

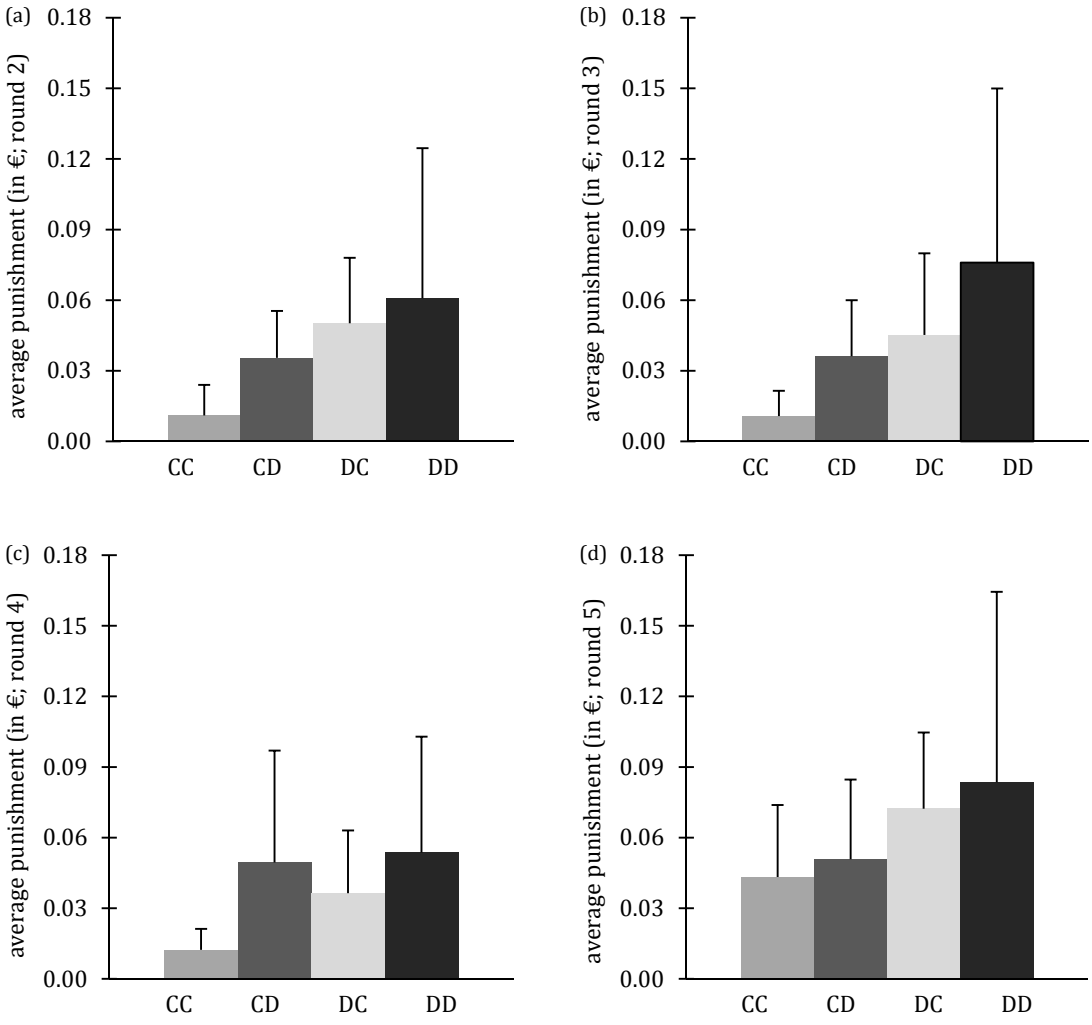

Figure S1 Average punishment investment (+ SD) in the (a) second, (b) third, (c) fourth and (d) fifth round of punishment (pooled over all periods). Participants could either contribute into the public good, C, or defect, D. Hence, in CD a contributor punished a defector (CC, DC, DD, respectively; Friedman test: (a) $\chi^2 = 5.0$, $df = 3$, $n = 6$, $p = 0.17$; (b) $\chi^2 = 2.95$, $df = 3$, $n = 6$, $p = 0.40$; (c) $\chi^2 = 2.29$, $df = 3$, $n = 6$, $p = 0.52$; (d) $\chi^2 = 2.6$, $df = 3$, $n = 6$, $p = 0.46$).

General Discussion

Is cooperation abundant?

The major aim of this thesis is to determine conditions under which cooperative behavior is established and maintained. To do so, predictions from evolutionary models were derived and human behavior was investigated in experimental settings. In the introduction I outlined examples of cooperation in humans and animals (see also Hammerstein 2003; Kappeler & van Schaik 2006). The present thesis contributes further empirical evidence to the list of cooperative endeavors. In the experiments, the majority of participants cooperated, i.e. the levels of cooperation in the iterated prisoner's dilemma and the public goods game were hardly ever below 50%. Nevertheless, the question whether "cooperation is abundant" can only in parts be answered with yes (cf. Chapter 1). Before outlining the conditions of cooperative behavior in regard to this thesis and presenting an outlook on future research, I will summarize the results of cooperation in humans.

In **Chapter 1**, I investigated for the first time, to my knowledge, whether different numbers of social partners affect cooperativity. Participants interacted either in one iterated prisoner's dilemma (cf. Box 1 p. 20) or in three iterated prisoner's dilemma at a time. The precise endpoints of the interactions were unknown to participants. Results revealed that when interacting with multiple partners cooperation levels were on average lower. Participants in the multiple-games setting treated their partners quite differently, despite the fact that identical monetary incentives were used in each dyadic interaction. This contradicts the assumption of game independence of traditional evolutionary game theory (Maynard Smith 1982; Nowak 2006a) and experimental results in the context of groups (Falk *et al.* 2010). Throughout three dyadic interactions participants were only able to establish one cooperative relationship. Here, no difference in cooperativity compared to the single-game setting could be found, additionally cooperation levels increased over time. Having one trustworthy and high-income relationship, the "temptation" to defect seemed to be enhanced, or participants are more strictly reciprocating defection. This resulted in one rather defective relationship where cooperation levels remained as low as 30% to 40% (the cooperation level of the third relationship was between the other two). Overall, I demonstrated that interactions in the iterated prisoner's dilemma with an unknown ending are defined by reciprocal behavior based on reactive strategies close to generous tit-for-tat, supporting the theoretical concept of direct reciprocity. However, this is in contrast to the

expected win-stay lose-shift like strategies, which is superior in evolutionary terms and was observed in previous studies (Wedekind & Milinski 1996; Milinski & Wedekind 1998). In sum, the classic mechanism of cooperation via direct reciprocity (cf. *General Introduction*) does not always lead to a cooperative outcome when simultaneously interacting in multiple pairwise games. Therefore, a new type of models is required which account for behavioral differences within the relationships of a single individual to better understand human cooperation in multiple dyadic, real-life interactions.

The study of **Chapter 2** provided first insights into how cooperation is influenced by the dynamic structure of social networks in experiments with humans. Relative to reciprocity in static relationships, the ability to change partners via an active-link-breaking mechanism (Pacheco *et al.* 2006a, 2006b, 2008), and hence the generation of a dynamic network, enhanced cooperation. This result is especially important since the theoretically predicted cooperation enhancing effects of social, static structure could not be found in so far (Grujić *et al.* 2010; Traulsen *et al.* 2010). Within dynamic networks cooperative links were maintained while links with defectors were broken. This biased link breaking resulted in an assortment process between participants on the network. In fact, assortment occurred in the form of "cooperative cliques". This network assortment is remarkable, because it occurred on top of behavioral assortment through direct reciprocity and was beyond the perception of participants. The formation of cooperative cliques occurred at a higher-level and represents a self-organized pattern. Hence, biased link breaking generated an interaction between behavior and network structure (i.e. cooperative cliques): behavioral decisions affect the network structure (i.e. clique formation), which forms the social ecology of participants (i.e. cooperators interacting with cooperators) and this feeds back again on behavioral decision and so forth. In general, the results highlight the importance of co-evolutionary processes. Here, behavior (at the individual level) and structure (at the network level) both evolve and exert selective pressures on the other, and thus have to be considered as correlated evolutionary processes.

The central issue of **Chapter 3** was whether punishment acts, which are not protected from retaliations, can still promote cooperation. In accordance with previous experimental findings, participants frequently punished and retaliated (e.g. Denant-Boemont *et al.* 2007; Nikiforakis 2008). Here, I extend this line of research by showing that acts of punishment, in contradiction to evolutionary theory (Janssen & Bushman 2008; Rand *et al.* 2009b; Wolff 2009), frequently escalated into cost-intense vendettas. In fact, counter-punishing participants were willing to match up their punishment fine, making the proverb an-eye-for-an-eye very real and supporting theories of equity (e.g. Adams 1965; Stillwell *et al.* 2008). The outbreak of vendettas was normally triggered by unjustified or ambiguous punishment.

In contrast, justified punishers were largely protected from on-going revenges, indicating underlying rules or social norms for punishment behavior. Remarkably, by quickly adjusting to the given experimental set-up some participants were able to avoid retaliation and occurrences of vendettas by delaying their punishment to the very last moment. The social interaction in the public goods game was most relevant for the first punishment, i.e. defectors attracted the highest punishment fines. This effect of coherent punishment seemed strong enough to raise cooperation levels over time. Overall, these results indicate that evolutionary models so far neglected important aspects of real-life interactions, like equity and reputational concerns, as vendettas occur across human societies.

Conditions that help to solve the puzzle

The results of the present thesis (in line with former experimental evidence) show that for cooperative behavior to prevail additional incentives, like conditional behavior or punishment, are needed. However, some mechanisms of reciprocity are not as strong and universally applicable as formerly believed. In long-lasting relationships cheating can be kept under control by using conditional strategies (Wedekind & Milinski 1996; Milinski & Wedekind 1998), though this is not always successful when one is simultaneously involved in more than one social relationship (Chapter 1). Another cheater-control mechanism is punishment. Punishment is very effective in group endeavors (e.g. Fehr & Gächter 2002), but also debated as a maladaptation with regard to cooperation (Dreber *et al.* 2008). I could demonstrate that punishment remains effective even in light of costly conflict escalation (Chapter 3); nevertheless, this brings into question the evolutionary origin of punishment in this scenario, because theory does not observe the emergence of vendettas (Rand *et al.* 2009b). In sum, in Chapters 1 and 3 I could extend our understanding of cooperation when it faces the threat of exploitation in pairwise interactions and exploitation in group endeavors.

Apart from further deepening the knowledge of these two known conditions for cooperation, I could also contribute empirical evidence for a novel condition. In models, dynamic networks have been shown to enhance cooperation (Perc & Szolnoki 2010), but empirical support was missing so far. Here, I demonstrated that the potential of rejecting partners via active link breaking leads to assortment of cooperative individuals, and that the generated network structure greatly enhanced cooperation (Chapter 2). This study demonstrated how two mechanisms for the evolution of cooperation – direct reciprocity and

network reciprocity[4] (cf. also *General Introduction*) – can be integrated and it shows that interactions between mechanisms can arise. Whereas in the static-network condition direct reciprocity could operate and a cooperative outcome was obtained, the comparison to the dynamic-network condition allowed network reciprocity to have a cooperation enhancing effect on top of direct reciprocity. The relationship between different mechanisms in experiments on human cooperation has not been investigated. An exception is the study by Rockenbach and Milinski (2006) where the combination of direct and indirect reciprocity enhances cooperation greatly.

Outlook – evolutionary solutions to cooperation

This paragraph will bring together the results of Chapters 1 to 3 and based on that I will outline possible future research questions on the evolution of cooperation.

A link between results of this thesis can be found if punishment is seen in a broader light. One way of punishing is by imposing a costly fine on someone (see Chapter 3). Another way to carry out "punishment" is by refusing to continue a relationship and rather "try one's luck" with a new partner (see Chapter 2). Here, the decision for a new partner is self-serving but can reduce the defector's fitness as a by-product. Depending on the goal an individual wants to accomplish with the help of a second party, and the number of alternative partners punishment via partner switching can either cause high or low costs. If the goal does not require special skills (i.e. many individuals possess the skill) and thus various alternative partners are available, then punishment costs are low and it might be worthwhile to stop a relationship which yields no good return and start one with another individual. Nevertheless, the time frame when partner switching is still low in costs might be narrow. Within dynamic networks (Chapter 2) half of the participants demanded a new partner at the beginning of the iterated prisoner's dilemma. Thus, many alternative partners where in the pool of participants that needed to be assigned a new partner at random. However, this changed rather quickly and participants looking for new partners towards the end of the game were confronted with the fact that they were quite likely to receive their old partner again simply because of lacking alternatives. In general, partner switching is a widespread control mechanism in animals. In interspecies mutualism, for instance, the clients (i.e. various reef fish species) of the cleaner wrasse, *Labroides dimidiatus*, switch to a new cleaner after defection by their current one (Bshary & Schäffer 2002). This forces the cleaner to be

[4] Note that network reciprocity was originally defined for static networks or spatial structures, like lattices (Nowak 2006b). Here the idea that structure impacts the evolution of behaviour is applied in the context of dynamic networks.

more cooperative (Bshary & Grutter 2005). More precisely, in cleaner-client interactions so-called residents (reef fish that live within defined territories and only have access to one cleaner fish) make use of punishment by chasing defective cleaners; whereas so-called visitors (pelagic fish that travel in larger areas and can use the services of various cleaners) make use of partner switching if exploited by a cleaner (Bshary 2010). In conclusion, it seems worthwhile to study the combined effect of costly punishment and partner switching as conditions to achieve cooperation in humans[5]. Individuals can adjust their "punishing behavior" according to the costs of punishment and the current costs of partner switching. I conjecture that this combination leads to low-cost partner switching when new relationships are still being established, whereas this pattern changes as soon as more and more long-term relationships are formed. Defectors would now be "educated" by costly punishment.

The combination of the outcomes of the first two chapters shows that within flexible relationships, in which partners can be exchanged (if necessary) cooperation prevails even though individuals seem to be tempted to defect with some interaction partners. This highlights the importance of studying cooperation in networks or structured populations, in particular in dynamic networks. In the last years, theoretical research on the evolution of cooperation has concentrated on structured populations (Nowak & May 1992; Szabó & Fáth 2007; Lion *et al.* 2011) and dynamic networks as well as heterogeneous (i.e. in terms of social connection numbers) networks have become popular topics for analyses (e.g. Santos *et al.* 2006c; Assenza *et al.* 2008; Pacheco *et al.* 2008). However, empirical evidence is lagging behind: so far only the impacts of static network structure (Cassar 2007; Kirchkamp & Nagel 2007; Traulsen *et al.* 2010; Grujić *et al.* 2010) and of dynamic-homogeneous network structure (Chapter 2) have been investigated. Hence, I in particular encourage future experimental research to include dynamic-heterogeneous networks. Within heterogeneous networks so-call hubs exist, i.e. individuals who have a high degree of connectedness in comparison to the rest of the population. These hubs clear the way for cooperation to prevail (Santos & Pacheco 2005, 2006; Santos *et al.* 2006c). In fact, in static-heterogeneous networks cooperators occupy hubs and defectors can only survive in nodes with a low connectedness, which greatly reduces their exploitive opportunities (Santos *et al.* 2006c). The reduced cooperation when having multiple relationships (Chapter 1) can be tackled by dynamic-heterogeneous networks[6]. In conjunction with the processes in static-heterogeneous

[5] Straight forward parallels to the idea of cost-dependent partner switching can be found in biological market theory (Noë *et al.* 1991; Noë & Hammerstein 1995; Bshary & Noë 2003) where supply and demand affect levels of cooperation.

[6] In this case not every participants would receive the same number of partners, but within a social network the number of partners would vary greatly between participants. However, having many interactions in real life causes higher costs of maintaining all relationships, for instances one has to invest in meetings, phone calls, sending postcards when on holidays, and so on. Thus, additional constraints or costs arise that are usually

networks (Santos & Pacheco 2006) and results of biased link breaking to defectors in dynamic networks (Chapter 2) the advantage of dynamic-heterogeneous networks lies in the possibility of rapid defector exclusion which is additionally supported by the nature of the network structure. Here, selection can act on behavior and structure, thus generating a co-evolutionary process. In general, no matter what the behavioral tendency of an individual is, it is best to have many links to cooperators. The arising problem is that defectors, who have many links to cooperators, fare very well and should dominate the population. However, cooperators can control cheating either by starting to defect as well, or by breaking the link to defectors and seeking new partners. In consequence, defectors (i) end up in defector-defector relationships, which reduce their payoff compared to defector-cooperator relationships, or (ii) they are abandoned by their partners. If cooperators are rare, defectors most likely receive defectors as new partners, or if partners are rare in general defectors might even be left with no new partner at all, both reduces their payoff. As a result, cooperators are provided with the chance to positively assort themselves and hence recover. Being a cooperator has the consequence of attracting many cooperative partners and to remain in cooperative long-term relationships (cf. Chapter 2). This allows cooperators to become a hub. Having reached this stage, two scenarios are possible in an experimental set-up. On the one hand, once cooperators became hubs who are linked to cooperative partners, and thus they have a very high payoff, it becomes particularly hard for defectors to invade due to the heterogeneous nature of the network, and cooperation prevails (this has been demonstrated theoretically for static-heterogeneous networks, Santos & Pacheco 2006). On the other hand, for cooperative individuals on hubs the "temptation" to defect increases and they lapse at least in parts into defective behavior (cf. Chapter 1). If a cooperative hub individual now becomes a defector, then however these defective hubs would soon collapse (see above). As a consequence many individuals (possibly conditional cooperators) are seeking new partners paving the way for cooperation to rise again. As a result, a cycle of cooperative hubs and defective hubs emerges (J. Pacheco, personal communication). As profound theoretical and empirical evidence is lacking, however, for the time being it remains elusive whether the mechanisms within dynamic-heterogeneous networks can promote cooperation in the long-run.

In conclusion, in the present thesis I could demonstrate that many pairwise interactions reduce cooperative behavior of humans (Chapter 1). However, when social partners can be

---

neglected in models (e.g. Nowak & May 1992; Ohtsuki *et al.* 2006; Santos *et al.* 2006c; Pacheco *et al.* 2008; Fu *et al.* 2009; but for an experiment, see Ahn *et al.* 2009) and will have to be studied in the future.

rejected – generating a dynamic network and resulting in assortment of individuals – cooperation can prevail (Chapter 2). In addition, within group endeavors punishment, though cost-intense escalations of conflict can arise, is nevertheless effective in promoting cooperation (Chapter 3). One next and logical step is to further combine these new pieces of knowledge as outlined above. It stands to reason that reduced cooperation can easily be overcome by combining partner switching and punishment, or by dynamic-heterogeneous networks (all is nonetheless prevalent in nature), but this possibly results in a much more effective cheater control. To do so only a joint way of both theoretical and empirical work can broaden our knowledge on the evolution of cooperation in particular and prosocial behavior in general.

SUMMARY

The general goal of this thesis was to study conditions under which prosocial behavior is established and maintained. I approached these questions by applying predictions and assumptions of evolutionary game theory to human cooperation within experimental settings. Cooperative behavior is an evolutionary puzzle, because over time natural selection should favor cheating behavior. Nevertheless, cooperation is ubiquitous in nature. Cooperative behavior is costly and provides benefits to other individuals, whereas cheating - or defection - does not incur costs but offers benefits allocated by cooperative behavior of others. Thus, cooperation cannot be an evolutionary stable strategy without special mechanisms that can reduce the chance of exploitation. In this thesis, I examined (i) the effects of multiple interactions on reciprocal behavior, and (ii) the impact of social structure on dyadic relationships using the prisoner's dilemma paradigm. Additionally, (iii) I investigated the effect of punishment, which can trigger conflict escalation, in a public goods game.

In Chapter 1, I investigated whether and in which way different numbers of social interaction partners affect cooperative behavior in the iterated prisoner's dilemma (IPD). In line with the concept of direct reciprocity, I found that participants in the traditional IPD with an unknown ending frequently cooperated. Overall, behavior most closely followed reactive strategies close to generous tit-for-tat. However, when participants interacted with multiple partners in three independent IPDs at a time average cooperativity decreased significantly. Further analyses showed that these participants had only one cooperative relationship similar to the relationship in the setting of a single IPD, but these participants could not establish cooperation in a second relationship (the cooperation level of the third relationship was located between these two). These results contradict the traditional assumption of evolutionary game theory of game independence as there seems to be an enhanced temptation to defect in some relationships, but not all, when interacting with three instead of only one social partner. Hence, theory needs to start modeling explicitly the impacts of different numbers of partners to account for behavioral differences within the relationships of a single individual. Studying cooperation within heterogeneous networks, where the number of partners varies, is a first attempt.

Further, in Chapter 2 I addressed the impact of social structure on cooperation. Relationships were characterized by an underlying network structure. Most previous models have neglected relationship networks, whereas within the recent decade, theoretical

research has started to include such structure. However, empirical evidence is lagging behind the development of theoretical insights. Here, I overcome such shortcomings by an experimental set-up where participants interacted in multiple but independent IPDs either within a static or dynamic network. In the latter, participants were given the option to break their social links after each prisoner's dilemma round. In accordance with theoretical predictions, cooperation levels were higher in dynamic networks compared to static networks. Additionally, participants in dynamic networks changed their social environment by biased link breaking to defectors. Hence, an assortment on the network took place and cooperative clusters emerged. This assortment is remarkable, because it occurred on top of behavioral assortment through direct reciprocity and beyond the perception of participants, and represents a self-organized pattern. In sum, these results highlight the importance of dynamic social networks, show that higher-order structures emerge above the individual level, and that these eventually feed back on selection processes.

In Chapter 3, I examined the impact of costly punishments on cooperative behavior where punishment acts can potentially escalate between participants. Groups of four played a public goods game which was followed by five rounds of punishment. Thus, whereas punishers are usually protected from retaliation this was not the case in this study. In the experiment, I found that sequences of costly punishment between participants, so-called vendettas, frequently occurred especially when punishment was unjustified or rather ambiguous. This finding contradicts theory which shows that vendettas do not evolve, as they are too cost-intense and reciprocated defection is the superior alternative strategy. Nevertheless, cooperation levels increased over time. This is presumably due to the fact that early punishment was mainly directed at defecting group members, which seemed to have increased their motivation to contribute into the public good. Moreover, some participants seemed to anticipate the outbreak of cost-intense vendettas and delayed their punishment to the last possible moment. These results indicate that evolutionary models so far neglected an important aspect of real-life interactions, as animals and humans frequently retaliate and as vendettas occur across human societies. So far one can only speculate that equity and reputational concerns are such central aspects.

In conclusion, the present thesis successfully identified conditions under which cooperation between unrelated individuals can be established or when cooperation levels remain low. On the one hand, this thesis provides empirical support for recent models of dynamic networks, but on the other hand limitations of other models could be pointed out. In addition, the thesis contributed further knowledge to the understanding of costly punishment and direct reciprocity in humans.

ZUSAMMENFASSUNG

Das Ziel der vorliegenden Arbeit war es, Bedingungen, unter denen prosoziales Verhalten entsteht, zu untersuchen. Hierzu wurden Vorhersagen und Annahmen der evolutionären Spieltheorie auf menschliches Kooperationsverhalten angewendet. Kooperatives Verhalten wird als evolutionäres Rätsel betrachtet, da natürliche Selektion betrügerisches Verhalten im Laufe der Zeit eigentlich begünstigen sollte. Nichtsdestotrotz ist Kooperation überall in der Natur vorzufinden. Per Definition ist kooperatives Verhalten kostenverursachend für den Handelnden und bietet gleichzeitig Vorteile oder Gewinne für andere Personen. Betrug oder Defektion dagegen verursachen keine Kosten, aber die bereitgestellten Vorteile anderer können dennoch genutzt werden. Infolgedessen ist Kooperation ohne die Unterstützung von Mechanismen, die die Gefahr der Ausbeutung reduzieren, keine evolutionär stabile Strategie. In der vorliegenden Arbeit wurden nun folgende Aspekte untersucht: (i) reziprokes Verhalten in multiplen paarweisen Interaktionen und (ii) die Auswirkung von sozialen Strukturen auf dyadische Beziehungen im Gefangenendilemma; und (iii) die Verwendung von Bestrafung unter Berücksichtigung der Möglichkeit von Konflikteskalation im öffentlichen-Güter-Spiel.

In Kapitel 1 wurde untersucht, ob und in welcher Weise sich unterschiedliche Anzahlen von Interaktionspartnern auf kooperatives Verhalten im wiederholten Gefangenendilemma (kurz IPD) auswirken. Gemäß den Annahmen der direkten Reziprozität zeigte sich, dass die Versuchspersonen im traditionellen IPD mit unbekanntem Endpunkt mehrheitlich kooperierten. Insgesamt entsprach das Verhalten reaktiven Strategien ähnlich zu „großzügigem Tit-For-Tat“. Wenn die Versuchspersonen mit mehreren Partnern in drei IPDs gleichzeitig interagierten, sankt die durchschnittliche Kooperativität allerdings signifikant ab. Weiterführende Analysen zeigten, dass diese Versuchspersonen nur eine kooperative Beziehung ähnlich der Beziehung aus dem ein-Partner IPD etablieren konnten, dass aber keine Kooperation in einer zweiten Beziehung aufgebaut werden konnten (das Kooperationslevel der dritten Beziehung lag zwischen diesen beiden). Diese Resultate widersprechen der traditionellen Annahme der evolutionären Spieltheorie, die eine Unabhängigkeit von Spielen annimmt, da eine erhöhte „Versuchung“ in einigen Beziehungen zu bestehen scheint, wenn man mit drei anstelle von nur einem Sozialpartner interagiert. All dies deutet daraufhin, dass Modelle explizit den Effekt von unterschiedlichen Anzahlen von Partnern mitaufnehmen sollten, um so dem differenzierenden Verhalten eines Individuums

gerecht zu werden. Ein Anfang stellt hier die Erforschung von Kooperation in heterogenen Netzwerken dar.

Die Auswirkung von sozialen Strukturen auf Kooperation wurde in Kapitel 2 betrachtet. Beziehungen können durch eine zugrundeliegende Netzwerkstruktur charakterisiert werden. Bisher wurde diese Gegebenheit in theoretischen Überlegungen zumeist ignoriert und erst kürzlich fanden Netzwerkstrukturen Berücksichtigung in Modellen. Empirische Erkenntnisse zu diesen Modellen gab es bisher kaum, so dass sich dieses Kapitel genau dieser Lücke widmete. Hier interagierten die Versuchspersonen in mehreren, unabhängigen IPDs entweder innerhalb eines statischen oder eines dynamischen Netzwerkes. In Letzterem hatten die Versuchspersonen die Möglichkeit ihre sozialen Verbindungen nach jeder Gefangenendilemma-Runde zu verändern. In Übereinstimmung mit theoretischen Modellen war die Kooperation in den dynamischen Netzwerken höher als in den statischen. Darüber hinaus veränderten die Versuchspersonen der dynamischen Netzwerke ihr soziales Umfeld durch ein bevorzugtes Beenden von Beziehungen zu Defektoren. Hierdurch fand eine Sortierung innerhalb des Netzwerkes statt und es bildeten sich kooperative Cliquen. Diese Selbstorganisation ist bemerkenswert, weil sie zusätzlich zum Effekt der direkten Reziprozität auftrat und weil die Versuchspersonen die Cliquenbildung auf Netzwerkebene nicht wahrnehmen konnten. Zusammenfassend zeigen diese Resultate die hohe Bedeutung von dynamischen sozialen Netzwerken auf und belegen, dass Strukturen höherer Ordnung neben dem Verhalten auf Individuumsebene entstehen können, welche dann wiederum in Wechselwirkung zum Selektionsdruck stehen können.

In Kapitel 3 wurde der Einfluss von kostenverursachender Bestrafung, die potentiell zwischen Versuchspersonen eskalieren kann, auf kooperatives Verhalten untersucht. Vierergruppen spielten das öffentliche-Güter-Spiel mit fünf aufeinanderfolgenden Bestrafungsrunden. In der Regel sind Bestrafer aufgrund des Experimentalaufbaus vor Vergeltung geschützt, hier war dies jedoch nicht der Fall. Tatsächlich entwickelten sich Sequenzen von kostenverursachender Bestrafung zwischen Versuchspersonen, sogenannte Vendetten. Sie traten besonders häufig auf, wenn die Bestrafung als ungerecht oder als beliebig eingestuft wurde. Diese Resultate stehen im Widerspruch zu theoretischen Modellen, in denen Vendetten nicht evolvieren, da sie zu kostenintensiv sind und Defektion die bessere Verhaltensalternative darstellt. Nichtsdestotrotz stieg die Kooperation im Laufe der Zeit an. Dies ist vermutlich darauf zurückzuführen, dass die Bestrafung der ersten Runde noch auf defektierende Gruppenmitglieder gerichtet war, welches ihre Motivation in das öffentliche Gut einzuzahlen letztlich erhöhte. Ferner schienen einige Versuchspersonen sogar den möglichen Ausbruch von kostenintensiven Vendetten zu antizipierten und verzögerten ihre Bestrafung bis zum letztmöglichen Zeitpunkt. Diese Resultate zeigen auf,

dass Modelle einen wichtigen Aspekt bisher vermissen lassen, da sowohl Tiere als auch Menschen häufig Vergeltung üben und sich gerade in menschlichen Gesellschaften Vendetten finden lassen. Spekulativ ist anzunehmen, dass Equity und Reputation gerade solche Aspekte dar stellen.

Zusammenfassend konnten mit dieser Arbeit Bedingungen identifizieren werden, unter denen Kooperation zwischen nicht-verwandten Personen entsteht und unter denen Kooperation niedrig ausfällt. Auf der einen Seite konnten neuere Modelle zu dynamischen sozialen Netzwerken empirisch untermauert werden - auf der anderen Seite wurden Schwachpunkte in anderen Modellen ausgewiesen. Zusätzlich trug diese Arbeit weitere Erkenntnisse zum Verständnis der kostenverursachenden Bestrafung und der direkten Reziprozität beim Menschen bei.

REFERENCES

Abbot, P., et al. (2011). Inclusive fitness theory and eusociality. *Nature*, 471, E1-E4.

Adams, J.S. (1965). Inequity in social exchange. In: *Advances in experimental social psychology* (ed. Berkowitz, L.). Academic Press: New York, pp. 267-299.

Ahn, T.K., Esarey, J. & Scholz, J.T. (2009). Reputation and cooperation in voluntary exchanges: Comparing local and central institutions. *The Journal of Politics*, 71, 398-413.

Amaral, L.A.N., Scala, A., Barthélémy, M. & Stanley, H.E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences USA*, 97, 11149-11152.

Andreoni, J. & Miller, J.H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal*, 103, 570-585.

Aoyagi, M. & Fréchette, G. (2009). Collusion as public monitoring becomes noisy: Experimental evidence. *Journal of Economic Theory*, 144, 1135-1165.

Assenza, S., Gómez-Gardeñes, J. & Latora, V. (2008). Enhancement of cooperation in highly clustered scale-free networks. *Physical Review E*, 78, 017101.

Aureli, F., Cozzolino, R., Cordischi, C. & Scucchi, S. (1992). Kin-oriented redirection among japanese macaques: An expression of a revenge system? *Animal Behaviour*, 44, 283-291.

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books: New York.

Axelrod, R. & Hamilton, W.D. (1981). The evolution of cooperation. *Science*, 211, 1390-1396.

Barclay, P. (2004). Trustworthiness and competitive altruism can also solve the 'tragedy of the commons'. *Evolution and Human Behavior*, 25, 209-220.

Barrett, L., Henzi, S.P., Weingrill, T., Lycett, J.E. & Hill, R.A. (1999). Market forces predict grooming reciprocity in female baboons. *Proceedings of the Royal Society B: Biological Sciences*, 266, 665-670.

Bednar, J., Chen, Y., Liu, T.X. & Page, S. (2010). Behavioral spillovers and cognitive load in multiple games: An experimental study. *Working Paper*, Department of Political Science. University of Michigan: Ann Arbor.

Boerlijst, M.C. & Hogeweg, P. (1991). Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. *Physica D*, 48, 17-28.

Boyd, R., Gintis, H., Bowles, S. & Richerson, P.J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences USA*, 100, 3531-3535.

Boyd, R. & Richerson, P.J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13, 171-195.

Brandt, H., Hauert, C. & Sigmund, K. (2003). Punishment and reputation in spatial public goods games. *Proceedings of the Royal Society B: Biological Sciences*, 1099-1104.

Brauchli, K., Killingback, T. & Doebeli, M. (1999). Evolution of cooperation in spatially structured populations. *Journal of Theoretical Biology*, 200, 405-417.

Brosnan, S.F., Houser, D., Leimgruber, K., Xiao, E., Chen, T. & de Waal, F.B.M. (2010). Competing demands of prosociality and equity in monkeys. *Evolution and Human Behavior*, 31, 279-288.

Brown, J.L. (1983). Cooperation – A biologist's dilemma. *Advances in the Study of Behavior*, 13, 1-37.

Bshary, R. (2010). Cooperation between unrelated individuals - A game theoretic approach. In: *Animal behavior: Evolution and mechanims* (ed. Kappeler, P.M.). Springer: Berlin, pp. 213-240.

Bshary, R. & Bergmüller, R. (2008). Distinguishing four fundamental approaches to the evolution of helping. *Journal of Evolutionary Biology*, 21, 405-420.

Bshary, R. & Grutter, A.S. (2005). Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biology Letters*, 1, 396-399.

Bshary, R. & Noë, R. (2003). Biological markets - The ubiquitous influence of partner choice on the dynamics of cleaner fish - client reef fish interactions. In: *Genetic and cultural evolution of cooperation* (ed. Hammerstein, P.). MIT Press: Cambridge, pp. 167-184.

Bshary, R. & Schäffer, D. (2002). Choosy reef fish select cleaner fish that provide high-quality service. *Animal Behaviour*, 63, 557-564.

Cason, T.N., Savikhin, A. & Sheremeta, R.M. (2010). Behavioral spillovers in coordination games. *Working Paper*, Department of Economics, Purdue University: West Lafayette.

Cassar, A. (2007). Coordination and cooperation in local, random and small world networks: Experimental evidence. *Games and Economic Behavior*, 58, 209-230.

Chen, Z., Gao, J., Cai, Y. & Xu, X. (2011). Evolutionary prisoner's dilemma game in flocks. *Physica A*, 390, 50-56.

Cheney, D.L., Moscovice, L.R., Heesen, M., Mundry, R. & Seyfarth, R.M. (2010). Contingent cooperation between wild female baboons. *Proceedings of the National Academy of Sciences USA*, 107, 9562-9566.

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. *Animal Behaviour*, 34, 1562-1566.

Cooper, R. & Ross, T.W. (1996). Cooperation without reputation: Experimental evidence from prisoner's dilemma games. *Games and Economic Behavior*, 12, 187-218.

Croft, D.P., et al. (2006). Social structure and co-operative interactions in a wild population of guppies (*poecilia reticulata*). *Behavioral Ecology and Sociobiology*, 59, 644-650.

Dal Bó, P. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *The American Economic Review*, 95, 1591-1604.

Dawes, C.T., Fowler, J.H., Johnson, T., McElreath, R. & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, 446, 794-796.

Dawes, R.M. (1980). Social dilemmas. *Annual Review of Psychology*, 31, 169-193.

de Waal, F.B.M. & Brosnan, S.F. (2006). Simple and complex reciprocity in primates. In: *Cooperation in primates and humans: Mechanisms and evolution* (eds. Kappeler, P.M. & van Schaik, C.P.). Springer Verlag: Berlin, pp. 85-106.

Denant-Boemont, L., Masclet, D. & Noussair, C.N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*, 33, 145-167.

Do, A.-L., Rudolf, L. & Gross, T. (2010). Patterns of cooperation: Fairness and coordination in networks of interacting agents. *New Journal of Physics*, 12, 063023.

Doebeli, M. & Hauert, C. (2005). Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecology Letters*, 8, 748-766.

Dorogotsev, S.N. & Mendes, J.F.F. (2003). *Evolution of networks: From biological nets to the internet and WWW*. Oxford University Press: Oxford.

Dreber, A., Rand, D.G., Fudenberg, D. & Nowak, M.A. (2008). Winners don't punish. *Nature*, 452, 348-351.

Duffy, J. & Ochs, J. (2009). Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior*, 66, 785-812.

Egas, M. & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 275, 871-878.

Elster, J. (1990). Norms of revenge. *Ethics*, 100, 862-885.

Ericksen, K.P. & Horton, H. (1992). "Blood feuds": Cross-cultural variations in kin group vengeance. *Cross-Cultural Research*, 26, 57-85.

Falk, A., Fischbacher, U. & Gächter, S. (2010). Living in two neighborhoods - Social interaction effects in the laboratory. *Economic Inquiry*.

Fehl, K., van der Post, D.J. & Semmann, D. (2011). Co-evolution of behaviour and social network structure promotes human cooperation. *Ecology Letters*, 14, 546-551.

Fehr, E. & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140.

Fehr, E. & Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114, 817-868.

Fewell, J.H. (2003). Social insect networks. *Science*, 301, 1867-1870.

Fletcher, J.A. & Doebeli, M. (2009). A simple and general explanation for the evolution of altruism. *Proceedings of the Royal Society B: Biological Sciences*, 276, 13-19.

Fowler, J.H. (2005). Altruistic punishment and the origin of cooperation. *Proceedings of the National Academy of Sciences USA*, 102, 7047-7049.

Fu, F., Chen, X., Liu, L. & Wang, L. (2007). Promotion of cooperation induced by the interplay between structure and game dynamics. *Physica A*, 383, 651-659.

Fu, F., Hauert, C., Nowak, M.A. & Wang, L. (2008). Reputation-based partner choice promotes cooperation in social networks. *Physical Review E*, 78, 026117.

Fu, F., Wu, T. & Wang, L. (2009). Partner switching stabilizes cooperation in coevolutionary prisoner's dilemma. *Physical Review E*, 79, 036101.

Gächter, S., Herrmann, B. & Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2651-2661.

Gächter, S., Renner, E. & Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322, 1510.

Gintis, H., Bowles, S., Boyd, R. & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24, 153-172.

Gould, R.V. (2000). Revenge as sanction and solidarity display: An analysis of vendettas in nineteenth-century Corsica. *American Sociological Review*, 65, 682-704.

Greene, W.H. (2008). *Econometric analysis*. (6 edn). Pearson Prentice Hall: Upper Saddle River.

Greiner, B. (2004). The online recruitment system ORSEE 2.0 - A guide for the organization of experiments in economics. *Working Paper Series in Economics*, 10. University of Cologne: Cologne, Germany.

Gross, T. & Blasius, B. (2008). Adaptive coevolutionary networks: A review. *Journal of the Royal Society Interface*, 5, 259-271.

Grujić, J., Fosco, C., Araujo, L., Cuesta, J.A. & Sánchez, A. (2010). Social experiments in the mesoscale: Humans playing a spatial prisoner's dilemma. *PLoS ONE*, 5, e13749.

Hamilton, W.D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology*, 7, 1-52.

Hammerstein, P. (2003). *Genetic and cultural evolution of cooperation*. MIT Press: Cambridge.

Hanaki, N., Peterhansl, A., Dodds, P.S. & Watts, D.J. (2007). Cooperation in evolving social networks. *Management Science*, 53, 1036-1050.

Hardin, G. (1968). The tragedy of the commons. *Science*, 162, 1243-1248.

Harinck, F., Van Dijk, E., Van Beest, I. & Mersmann, P. (2007). When gains loom larger than losses: Reversed loss aversion for small amounts of money. *Psychological Science*, 18, 1099-1105.

Hauert, C. & Doebeli, M. (2004). Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature*, 428, 643-646.

Hauert, C., Holmes, M. & Doebeli, M. (2006). Evolutionary games and population dynamics: Maintenance of cooperation in public goods games. *Proceedings of the Royal Society B: Biological Sciences*, 273, 2565-2570.

Hauert, C., Traulsen, A., Brandt, H., Nowak, M.A. & Sigmund, K. (2007). Via freedom to coercion: The emergence of costly punishment. *Science*, 316, 1905-1907.

Hauk, E. (2003). Multiple prisoner's dilemma games with(out) an outside option: An experimental study. *Theory and Decision*, 54, 207-229.

Hellman, H. (1998). *Great feuds in science: Ten of the liveliest disputes ever*. Wiley: New York.

Hellman, H. (2006). *Great feuds in mathematics: Ten of the liveliest disputes ever*. Wiley: Hoboken.

Henrich, J. & Boyd, R. (2001). Why people punish defectors. Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, 208, 79-89.

Herre, E.A. & Wcislo, W.T. (2011). In defence of inclusive fitness theory. *Nature*, 471, E8-E9.

Herrmann, B., Thöni, C. & Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319, 1362-1367.

İçli, T.G. (1994). Blood feud in turkey: A sociological analysis. *British Journal of Criminology*, 34, 69-74.

Ifti, M., Killingback, T. & Doebeli, M. (2004). Effects of neighbourhood size and connectivity on the spatial continuous prisoner's dilemma. *Journal of Theoretical Biology*, 231, 97-106.

Janssen, M.A. & Bushman, C. (2008). Evolution of cooperation and altruistic punishment when retaliation is possible. *Journal of Theoretical Biology*, 254, 541-545.

Jensen, K. (2010). Punishment and spite, the dark side of cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2635-2650.

Jensen, K., Call, J. & Tomasello, M. (2007). Chimpanzees are vengeful but not spiteful. *Proceedings of the National Academy of Sciences USA*, 104, 13046-13050.

Jun, T. & Sethi, R. (2009). Reciprocity in evolving social networks. *Journal of Evolutionary Economics*, 19, 379-396.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-292.

Kappeler, P.M. & van Schaik, C.P. (2006). *Cooperation in primates and humans: Mechanisms and evolution*. Springer: Berlin.

Kirchkamp, O. & Nagel, R. (2007). Naive learning and cooperation in network experiments. *Games and Economic Behavior*, 58, 269-292.

Konno, T. (2011). A condition for cooperation in a game on complex networks. *Journal of Theoretical Biology*, 269, 224-233.

Kossinets, G. & Watts, D.J. (2006). Empirical analysis of an evolving social network. *Science*, 311, 88-90.

Kurzban, R., DeScioli, P. & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28, 75-84.

Ledyard, J.O. (1995). Public goods: A survey of experimental research. In: *Handbook of experimental economics* (eds. Kagel, J.H. & Roth, A.E.). Princeton Universtiy Press: Princeton, pp. 111-194.

Lehmann, L. & Keller, L. (2006). The evolution of cooperation and altruism – A general framework and a classification of models. *Journal of Evolutionary Biology*, 19, 1365-1376.

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

Lion, S., Jansen, V.A.A. & Day, T. (2011). Evolution in structured populations: Beyond the kin versus group debate. *Trends in Ecology and Evolution*, 26, 193-201.

Lion, S. & van Baalen, M. (2008). Self-structuring in spatial evolutionary ecology. *Ecology Letters*, 11, 277-295.

Masclet, D., Noussair, C.N., Tucker, S. & Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93, 366-380.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press: Cambridge.

Maynard Smith, J. & Price, G.R. (1973). The logic of animal conflict. *Nature*, 246, 15-18.

Melis, A.P. & Semmann, D. (2010). How is human cooperation different? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2663-2674.

Milinski, M. (2006). Reputation, personal identity and cooperation in a social dilemma. In: *Cooperation in primates and humans: Mechanisms and evolution* (eds. Kappeler, P.M. & van Schaik, C.P.). Springer: Berlin, pp. 263-274.

Milinski, M., Pfluger, D., Külling, D. & Kettler, R. (1990). Do sticklebacks cooperate repeatedly in reciprocal pairs? *Behavioral Ecology and Sociobiology*, 27, 17-21.

Milinski, M. & Rockenback, B. (in press). On the interaction of the stick and the carrot in social dilemmas. *Journal of Theoretical Biology*.

Milinski, M., Semmann, D. & Krambeck, H.-J. (2002). Reputation helps to solve the 'tragedy of the commons'. *Nature*, 415, 424-426.

Milinski, M. & Wedekind, C. (1998). Working memory constrains human cooperation in the prisoner's dilemma. *Proceedings of the National Academy of Sciences USA*, 95, 13755-13758.

Nicklisch, A. & Wolff, I. (2009). Cooperation norms in multiple-stage punishment. *Working Paper*, Max Planck Institute for Research on Collective Goods: Bonn, Germany.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92, 91-112.

Nikiforakis, N. & Engelmann, D. (2008). Feuds in the laboratory? A social dilemma experiment. *Working Paper*, Research Paper Series 1058, Department of Economics, The University of Melbourne: Melbourne.

Nikiforakis, N. & Engelmann, D. (2011). Altruistic punishment and the threat of feuds. *Journal of Economic Behavior and Organization*, 78, 319-332.

Noë, R. & Hammerstein, P. (1995). Biological markets. *Trends in Ecology and Evolution*, 10, 336-339.

Noë, R., van Schaik, C.P. & van Hooff, J.A. (1991). The market effect: An explanation for pay-off asymmetries among collaborating animals. *Ethology*, 87, 97-118.

Nowak, M.A. (2006a). *Evolutionary dynamics: Exploring the equations of life*. Harvard University Press: Cambridge.

Nowak, M.A. (2006b). Five rules for the evolution of cooperation. *Science*, 314, 1560-1563.

Nowak, M.A. & May, R.M. (1992). Evolutionary games and spatial chaos. *Nature*, 359, 826-829.

Nowak, M.A. & Sigmund, K. (1990). The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematicae*, 20, 247-265.

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

Nowak, M.A. & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364, 56-58.

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

Nowak, M.A. & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437, 1291-1298.

Nowak, M.A., Tarnita, C.E. & Wilson, E.O. (2010). The evolution of eusociality. *Nature*, 466, 1057-1062.

Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M.A. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441, 502-505.

Ohtsuki, H. & Iwasa, Y. (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, 239, 435-444.

Ohtsuki, H. & Nowak, M.A. (2007). Direct reciprocity on graphs. *Journal of Theoretical Biology*, 247, 462-470.

Ostrom, E., Walker, J. & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86, 404-417.

Pacheco, J.M., Traulsen, A. & Nowak, M.A. (2006a). Active linking in evolutionary games. *Journal of Theoretical Biology*, 243, 437-443.

Pacheco, J.M., Traulsen, A. & Nowak, M.A. (2006b). Coevolution of strategy and structure in complex networks with dynamical linking. *Physical Review Letters*, 97, 258103.

Pacheco, J.M., Traulsen, A., Ohtsuki, H. & Nowak, M.A. (2008). Repeated games and direct reciprocity under active linking. *Journal of Theoretical Biology*, 250, 723-731.

Pennisi, E. (2009). On the origin of cooperation. *Science*, 325, 1196-1199.

Perc, M. & Szolnoki, A. (2010). Coevolutionary games - A mini review. *BioSystems*, 99, 109-125.

Rand, D.G., Armao, J.J., Nakamaru, M. & Ohtsuki, H. (2010). Anti-social punishment can prevent the co-evolution of punishment and cooperation. *Journal of Theoretical Biology*, 265, 624-632.

Rand, D.G., Dreber, A., Ellingsen, T., Fundenberg, D. & Nowak, M.A. (2009a). Positive interactions promote public cooperation. *Science*, 325, 1272-1275.

Rand, D.G., Ohtsuki, H. & Nowak, M.A. (2009b). Direct reciprocity with costly punishment: Generous tit-for-tat prevails. *Journal of Theoretical Biology*, 256, 45-57.

Rapoport, A. & Chammah, A.M. (1965). *Prisoner's dilemma: A study in conflict and cooperation*. University of Michigan Press: Ann Arbor.

Reeve, H.K. & Hölldobler, B. (2007). The emergence of a superorganism through intergroup competition. *Proceedings of the National Academy of Sciences USA*, 104, 9736-9740.

Richerson, P.J., Boyd, R. & Henrich, J. (2003). Cultural evolution of human cooperation. In: *Genetic and cultural evolution of cooperation* (ed. Hammerstein, P.). MIT Press: Cambridge, pp. 357-388.

Rockenbach, B. & Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature*, 444, 718-723.

Santorelli, L.A., et al. (2008). Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae. *Nature*, 451, 1107-1101.

Santos, F.C. & Pacheco, J.M. (2005). Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters*, 95, 098104.

Santos, F.C. & Pacheco, J.M. (2006). A new route to the evolution of cooperation. *Journal of Evolutionary Biology*, 19, 726-733.

Santos, F.C., Pacheco, J.M. & Lenaerts, T. (2006a). Cooperation prevails when individuals adjust their social ties. *PLoS Computational Biology*, 2, 1284-1291.

Santos, F.C., Pacheco, J.M. & Lenaerts, T. (2006b). Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences USA*, 103, 3490-3494.

Santos, F.C., Rodrigues, J.F. & Pacheco, J.M. (2006c). Graph topology plays a determinant role in the evolution of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 273, 51-55.

Savikhin, A. & Sheremeta, R.M. (2010). Simultaneous decision-making in competitive and cooperative environments. *Working Paper*, The University of Chicago: Chicago.

Schino, G. & Aureli, F. (2010). The relative roles of kinship and reciprocity in explaining primate altruism. *Ecology Letters*, 13, 45-50.

Sefton, M., Shupp, R. & Walker, J.M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45, 671-690.

Selten, R. & Stoecker, R. (1986). End behavior in sequences of finite prisoner's dilemma supergames: A learning theory approach. *Journal of Economic Behavior and Organization*, 7, 47-70.

Semmann, D., Krambeck, H.-J. & Milinski, M. (2004). Strategic investment in reputation. *Behavioral Ecology and Sociobiology*, 56, 248-252.

Semmann, D., Krambeck, H.-J. & Milinski, M. (2005). Reputation is valuable within and outside one's own social group. *Behavioral Ecology and Sociobiology*, 57, 611-616.

Shakespeare, W. (1954). *The tragedy of Romeo and Juliet*. Yale University Press: New Haven.

Sigmund, K. (2007). Punish or perish? Retaliation and collaboration among humans. *Trends in Ecology and Evolution*, 22, 593-600.

Sigmund, K., De Silva, H., Traulsen, A. & Hauert, C. (2010). Social learning promotes institutions for governing the commons. *Nature*, 466, 861-863.

Sober, E. & Wilson, D.S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Harvard University Press: Cambridge.

Stillwell, A.M., Baumeister, R.F. & Del Priore, R.E. (2008). We're all victims here: Toward a psychology of revenge. *Basic and Applied Social Psychology*, 30, 253-263.

Szabó, G. & Fáth, G. (2007). Evolutionary games on graphs. *Physics Reports*, 446, 97-216.

Szolnoki, A., Perc, M. & Danku, Z. (2008). Making new connections towards cooperation in the prisoner's dilemma game. *Europhysics Letters*, 84, 50007.

The Local (2010, June 10). *Neighbourhood feud ends in bloody weed-whacker attack*. Retrieved Mai 24, 2011, from http://www.thelocal.de/society/20100610-27771.html.

Traulsen, A. & Nowak, M.A. (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences USA*, 103, 10952-10955.

Traulsen, A., Semmann, D., Sommerfeld, R.D., Krambeck, H.-J. & Milinski, M. (2010). Human strategy updating in evolutionary games. *Proceedings of the National Academy of Sciences USA*, 107, 2962-2966.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35-57.

Turner, P.E. & Chao, L. (1999). Prisoner's dilemma in an RNA virus. *Nature*, 398, 441-443.

van Schaik, C.P. & Kappeler, P.M. (2006). Cooperation in primates and humans: Closing the gap. In: *Cooperation in primates and humans: Mechanisms and evolution* (eds. Kappeler, P.M. & van Schaik, C.P.). Springer: Berlin, pp. 3-21.

Voelkl, B. & Kasper, C. (2009). Social structure of primate interaction networks facilitates the emergence of cooperation. *Biology Letters*, 5, 462-464.

Wedekind, C. & Milinski, M. (1996). Human cooperation in the simultaneous and the alternating prisoner's dilemma: Pavlov versus generous tit-for-tat. *Proceedings of the National Academy of Sciences USA*, 93, 2686-2689.

West, S.A., et al. (2006). Cooperation and the scale of competition in humans. *Current Biology*, 16, 1103-1106.

West, S.A., Griffin, A.S. & Gardner, A. (2007a). Evolutionary explanations for cooperation. *Current Biology*, 17, R661-R672.

West, S.A., Griffin, A.S. & Gardner, A. (2007b). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, 20, 415-432.

Wilson, D.S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences USA*, 72, 143-146.

Wilson, D.S. (1983). The group selection controversy: History and current status. *Annual Review of Ecology and Systematics*, 14, 159-187.

Wilson, E.O. & Hölldobler, B. (2005). Eusociality: Origin and consequences. *Proceedings of the National Academy of Sciences USA*, 102, 13367-13371.

Wolff, I. (2009). Counterpunishment revisited: An evolutionary approach. *Working Paper*. Center for Empirical Research in Economics and Behavioral Sciences, University of Erfurt: Erfurt, Germany.

Wu, B., Zhou, D., Fu, F., Luo, Q., Wang, L. & Traulsen, A. (2010). Evolution of cooperation on stochastic dynamical networks. *PLoS ONE*, 5, e11187.

Wu, J.-J., et al. (2009). Costly punishment does not always increase cooperation. *Proceedings of the National Academy of Sciences USA*, 106, 17448-17451.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110-116.

Yamauchi, A., Tanimoto, J. & Hagishima, A. (2011). An analysis of network reciprocity in prisoner's dilemma games using full factorial designs of experiment. *BioSystems*, 103, 85-92.

ACKNOWLEDGEMENTS

First, I would like to thank Dirk Semmann without whom this thesis would not have been possible. He hired me even though I was “only” a psychologist and tried to turn me into a biologist with great effort, patiently explaining those biological concepts which were unfamiliar to me. He was always up to date on what I was doing. I could always stop by in his office to discuss new ideas and experimental set-ups.

Special thanks go to Daniel van der Post, who was always ready to have a coffee break in the kitchen, to discuss our various projects and ideas, and to give me detailed feedback on my drafts. I sincerely thank Mathias Franz for spreading his statistical knowledge. I would like to thank Margarete Boos for keeping me on the psychological track and her enthusiasm for interdisciplinary research. Outside the department, I would like to thank Arne Traulsen as a reliable source of theoretical knowledge. Further, I wish to thank Manfred Milinski, Ralf Sommerfeld, and Jürgen Krambeck.

I would like to thank Johannes Pritz, Frederic Nowak and Bernhard Brauner for helping with technical issues such as keeping the lab running and programming the software for the experiments. Thanks to Uta Schröder and Christine Wittge for helping out in the lab and hunting down new participants. A big thank you also goes to the students for their participation in the experiments.

Of course I also want to thank everyone in Courant Research Center Evolution of Social Behavior for stimulating discussions and interesting exchanges of ideas. Financial support for this work was provided by the German Initiative of Excellence of the German Science Foundation (DFG).

CURRICULUM VITAE

Katrin Fehl

Born November, 28th 1981 in Bad Soden-Salmünter, Germany

Present Position

Since 12/2008	PhD-student, Courant Research Center Evolution of Social Behavior, University of Göttingen
	Thesis: "Experimental and theoretical investigations of the emergence and sustenance of prosocial behavior in groups"

Education and Academic Degrees

2008	Diploma in Psychology University of Marburg, Germany
	Diploma-Thesis: „Interventionen bei unfairem Verhalten durch unbeteiligte Dritte: Reaktionen von Beobachtern auf Normbruch im Ultimatumspiel"
2002 – 2008	Studies of Psychology University of Marburg, Germany and Aarhus University, Denmark
2001 – 2002	Au Pair, Kansas City, USA
2001	Abitur (A-level), Berufliches Gymnasium, Schlüchtern, Germany

List of Publications

Fehl, K., van der Post, D.J. & Semmann, D. (2011). Co-evolution of behaviour and social network structure promotes human cooperation. *Ecology Letters*, 14, 546-551.

Overgaard, M., Fehl, K., Mouridsen, K., Bergholt, B. & Cleeremans, A. (2008). Seeing without seeing? Degraded conscious vision in a blindsight patient. *PLoS ONE*, 3, 1-4.

Fink, B., Brewer, G., Fehl, K. & Neave, N. (2007). Instrumentality and lifetime number of sexual partners. *Personality and Individual Differences*, 43, 747-756.

ERKLÄRUNG ÜBER EIGENE LEISTUNGEN

Ich versichere, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Die Stellen, die anderen Werken wörtlich oder sinngemäß entnommen sind, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass diese Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat.

Die Publikationen, wie sie in den Kapiteln 1 bis 3 repliziert sind, wurden von mir selbst verfasst. Dirk Semmann leitete alle Arbeiten als Dissertationsbetreuer an. Daniel van der Post programmierte die Agent-based Simulationen zu Kapitel 2. Ralf Sommerfeld und Manfred Milinski entwickelten die Idee zur eskalierenden Bestrafung aus Kapitel 3 und führten die Datenaufnahme durch. Alle Koautoren wirkten bei der Finalisierung der Manuskripte mit.

Göttingen, Juni 2011