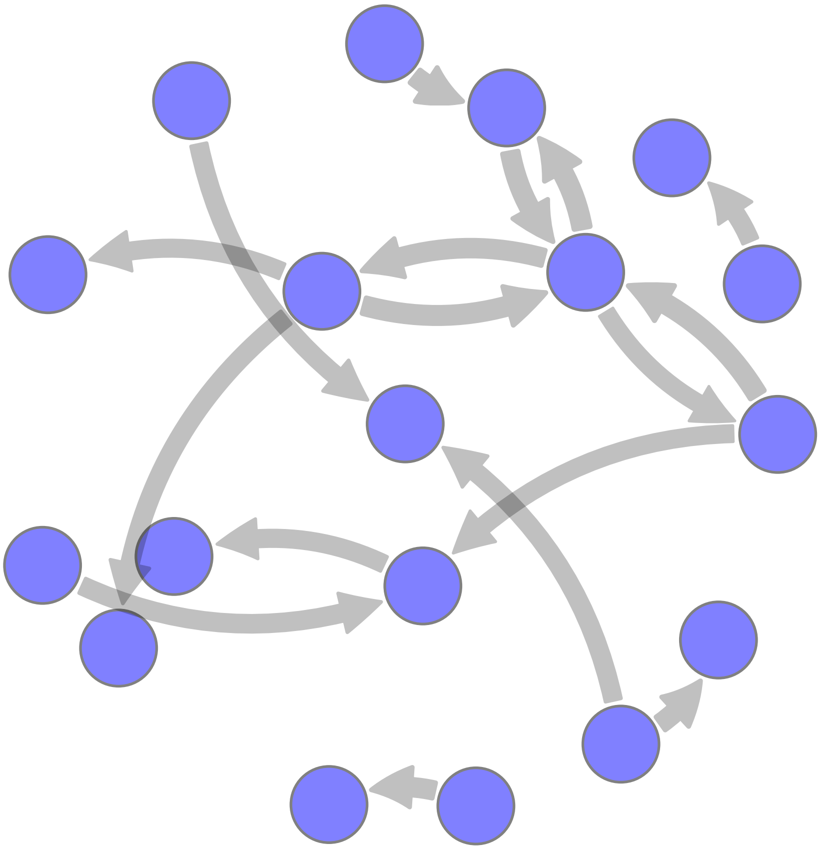
Partner control mechanisms in repeated social interactions

Matthias Wubs - PhD Thesis - University of Neuchâtel - 2017



PhD Thesis submitted to:
Faculty of Science, Institute of Biology, University of Neuchâtel, Switzerland

Prof. Redouan Bshary, University of Neuchâtel, supervisor
Prof. Laurent Lehmann, University of Lausanne, supervisor
Prof. Klaus Zuberbühler, University of Neuchâtel, internal examiner
Prof. Peter Hammerstein, Humboldt University of Berlin, external examiner

Defended on March 28, 2017

UNIVERSITÉ DE
NEUCHÂTEL

Faculté des Sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel – Suisse
Tél : + 41 (0)32 718 21 00
E-mail : secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Monsieur Matthias WUBS

Titre:

"Partner control mechanisms in repeated social interactions"

sur le rapport des membres du jury composé comme suit:

- Prof. Redouan Bshary, directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Laurent Lehmann, co-directeur de thèse, Université de Lausanne, Suisse
- Prof. Klaus Zuberbühler, Université de Neuchâtel, Suisse
- Prof. Peter Hammerstein, Humboldt-Universität, Berlin Allemagne

Neuchâtel, le 11 avril 2017

Le Doyen, Prof. R. Bshary

General abstract

Individuals that interact socially regularly have to make decisions whether to help another individual (provide some payoff benefit, possible at a personal payoff cost) or not. Here, a pair of individuals is best off if both individuals help each other, but a single individual may gain a relative payoff advantage by not helping (defecting) while the partner does choose to help, thus creating a dilemma. Because social interactions are often repeated, individuals can condition the actions they take on the actions taken by their partner in previous rounds of interaction. The so-called partner control mechanisms positive reciprocity (where acts of helping are conditioned on receiving help from the partner), punishment, and partner switching have all been shown to stabilize cooperation in populations where the individuals engage in repeated pairwise interactions. What remains unclear, however, is under which conditions each partner control mechanism will be dominant in a population if the the partner control mechanisms coevolve.

Additionally, the expressed behaviour in repeated social interactions may depend on the state of the environment in which the interactions take place. Social grooming in primates is likely to depend on the food competition that the individuals experience. If food is clustered and monopolizable, high ranked individuals may defend food sources, where low ranked individuals then need to groom high ranked ones in order to be tolerated on the food source, resulting in grooming being directed up the hierarchy. However, the conditions that cause grooming to be traded for grooming or grooming to be traded for tolerance have yet to be quantified.

In this thesis, I developed several agent-based models in order to investigate both the coevolution of various partner control mechanisms and the grooming patterns in primates.

In chapter one, I demonstrated that in a well-mixed population the likelihood of partner switching being the dominant partner control mechanism increases with increasing number of rounds of interaction. Furthermore, if interactions are localized to small groups of unrelated individuals, then punishment is more likely to be favoured by selection compared to the well-mixed case, while the conditions where positive reciprocity is dominant are less clearly defined.

In chapter two, I investigated how limited migration of offspring, overlapping generations, and complexity costs affect the competitiveness of each partner control mechanism. It is shown that the relative competitiveness of partner switching is increased due to generational overlap, while punishment is most strongly negatively affected by complexity costs. Additionally, while the conditions where these partner control mechanisms are dominant in the population increases if the population is kin structured, the conditions where unconditional helping is dominant over conditional helping strategies are rather stringent.

In chapter three, I developed a reinforcement learning model that simulated grooming and feeding interactions in primates. The model generated patterns of grooming reciprocity in the absence of food competition, while grooming was found to be directed up the hierarchy if individuals compete for food. It is shown that grooming up the hierarchy may not necessarily increase with increasing food competition, and that an increase in aggressiveness causes grooming to become more reciprocal.

In summary, helping behaviours occur in a diversity of repeated social interactions in natural populations. By exploring a large range of conditions, the models developed in this thesis provide insights regarding which partner control mechanism is likely to be dominant in a population. In addition, the model from chapter three shows how grooming patterns may depend on a variety of relevant parameters.

Keywords: social behaviour; repeated interactions; partner control mechanisms; simulation models; cooperation; helping

Résumé général

Les individus qui interagissent socialement doivent souvent décider entre aider leur prochain, en leur procurant un bénéfice a un possible cout, ou pas. Tandis qu'une paire d'individus s'en sort mieux si les deux s'entraident, un individu peut tirer un bénéfice relatif s'il se décide à ne pas aider tandis que l'autre l'aide. Cette situation crée un dilemme social. Parce que les interactions sociales sont souvent répétées, les individus peuvent conditionner leurs propres actions sur les actions de leur partenaire lors d'interactions passées. Il existe trois mécanismes de contrôle de partenaire qui stabilisent la coopération lors d'interactions répétées entre paires d'individus: la réciprocité positive (la décision d'aider est conditionnée sur l'aide du partenaire lors de l'interaction précédente), la punition, et le changement de partenaire. Mais les conditions sous lesquelles un mécanisme domine sur les autres lorsque les trois mécanismes coévoluent dans une même population sont mal comprises.

Un autre point qu'il est important de considérer pour comprendre les interactions sociales est que le comportement exprimé peut aussi dépendre sur l'environnement. Il est probable que le toilettage social chez les primates, par exemple, dépend de la compétition sur les ressources de nourriture. Si la nourriture est concentrée dans l'espace et facilement monopolisée, les individus en haut de la hiérarchie sociale peuvent défendre ces ressources, auquel cas les individus en bas de la hiérarchie doivent les toiletter afin d'être tolérer sur leur territoire et accéder à leur ressource. Afin de comprendre ce comportement, il est nécessaire de comprendre les conditions sous lesquelles le toilettage est échangé contre du toilettage, ou contre la tolérance sur un territoire.

Dans cette thèse, je développe et utilise des modèles agents basés pour explorer la coévolution des mécanismes de contrôle de partenaire et l'évolution de patrons de toilettage chez les primates.

Dans le premier chapitre, je démontre que dans une population panmictique, plus les interactions sont répétées, plus il est probable que le changement de partenaire soit le mécanisme de contrôle de partenaire dominant. Si les interactions sont restreintes au sein de petits groupes de non-apparentés, alors la punition est le mécanisme le plus probablement favorisé par la sélection. Les conditions pour que la réciprocité positive soit dominante sont moins clairement définies.

Dans le deuxième chapitre, j'étudie comment la dispersion limitée chez la progéniture, le recouvrement de génération, et le cout de la complexité influencent la compétitivité de chaque mécanisme de contrôle. Il est démontré que la compétitivité du changement de partenaire est accrue par le recouvrement de génération tandis que celle de la punition est très diminuée par le cout de complexité. De plus, alors que les conditions pour que ces mécanismes de contrôle de partenaire soient dominants s'élargissent avec la structure d'apparentement, les conditions pour que l'aide inconditionnelle domine sur l'aide conditionnelle se restreignent.

Dans le troisième chapitre, je développe un model d'apprentissage par renforcement pour simuler les interactions de toilettage et l'accès a la nourriture chez les primates. On observe des patrons de toilettage réciproque lorsqu'il n'y pas de compétition pour les ressources, et des patrons de toilettage dirigé vers le haut de l'échelle hiérarchique lorsque la compétition pour les ressources augmente. Il est montré que l'effort dans le toilettage n'augmente pas forcément avec la compétition pour les ressources, et qu'une augmentation d'agressivité amène le toilettage à être plus réciproque.

En bref, les comportements sociaux d'entraide dans les milieux naturels sont observés dans une multitude d'interactions sociales répétées. En explorant une gamme variée de conditions, les modèles développés dans cette thèse permettent de mieux comprendre quels mécanismes de contrôle de partenaire sont favorisés par la sélection naturelle. De plus, le modèle du troisième chapitre montre que les patrons de toilettage dépendent de plusieurs paramètres naturels importants.

Mots clés: comportement social; interactions répétées; mécanismes de contrôle de partenaire; modèles de simulation; coopération; aider

Table of Contents

General Introduction **1**
- Evolution of social strategies . . . 2
- Ecological influence on behaviour . . . 3
- Main question . . . 4
- Approach . . . 4
- Thesis outline . . . 4
- References . . . 6

Chapter one - Coevolution between positive reciprocity, punishment, and partner switching in repeated interactions **11**
- Introduction . . . 12
- The model . . . 13
- Results . . . 16
- Discussion . . . 17
- References . . . 21
- Table and figures . . . 24
- Supplementary material . . . 26

Chapter two - Coevolution of positive reciprocity, punishment, partner switching, and unconditional helping in a kin structured population **35**
- Introduction . . . 36
- The model . . . 38
- Results . . . 40
- Discussion . . . 44
- References . . . 47
- Table and figures . . . 50
- Supplementary material . . . 55

Chapter three - A model on grooming up the hierarchy in primates **63**
- Introduction . . . 64
- The model . . . 65
- Results . . . 71
- Discussion . . . 75
- References . . . 79
- Table and figures . . . 83
- Supplementary material . . . 90

General Discussion **97**
- Coevolution of partner control mechanisms . . . 97
- Grooming up the hierarchy . . . 98
- Unification of both approaches . . . 99
- Empirically informed models . . . 100
- Concluding remarks . . . 101
- References . . . 102

Acknowledgements **105**

General introduction

Organisms (of whatever species) frequently engage in social interactions (Dugatkin, 1997; Sachs et al., 2004). There are countless examples where the expressed behaviours of an organism affect the fitness of those in its spatial neighbourhood. These behaviours vary from bacteria releasing siderophores (molecules to absorb iron) that other bacteria can profit from (Griffin et al., 2004), to plants providing nectar to various insect species (Kearns et al., 1998), to fish aggregating in large schools (Castro et al., 2002), to cooperative breeding in birds and mammals (Arnold and Owens, 1998; Jennions, 1994). Regularly, the interacting individuals have to make decisions whether to help the other (providing some benefit), possible at a personal cost (i.e., an investment), or not to help the other. Provided that the benefit of helping outweighs the cost, then a pair of individuals would be best off by helping each other. However, one individual can gain a relative advantage by not reciprocating the help it receives (defecting), a situation that is elegantly captured in the Prisoner's Dilemma (Luce and Raiffa, 1957; Rapoport and Chammah, 1965). Despite the apparent dilemma, examples of cooperation (i.e., an interaction between individuals that benefits all individuals involved) are widespread in nature. While the existence of helping behaviours and cooperation has often been presented as a puzzle, a large number of solutions have been suggested (Lehmann and Keller, 2006; West et al., 2007; Leimar and Hammerstein, 2010; Bshary and Bronstein, 2011). A number of solutions comprise conditional decision making in repeated interactions, and in this thesis I investigate various such solutions.

Hamilton showed that a social trait (such as the propensity to help) can evolve if the fitness cost to the individual (c) is outweighed by the fitness benefit it confers to the recipient (b) multiplied by the coefficient of relatedness (r), yielding Hamilton's rule: $rb - c > 0$ (Hamilton, 1964a,b). However, interactions between unrelated individuals (e.g., in mutualisms, Bronstein, 2001, 2009), where $r = 0$, require alternative solutions, where the individuals thus has to gain a direct fitness benefit. Here, the solution comes from the fact that many social interactions are repeated. For example, an insect may visit the same plant more than once, and birds may raise several broods together. Then, cooperation can be maintained in a population if the individuals involved use social strategies that ensure that helpful individuals are benefited more through the interactions than defecting ones (West et al., 2007; Bshary and Bronstein, 2011). Studies investigating the optimal strategies in a variety of games are numerous (e.g., Axelrod and Hamilton, 1981; Boyd and Lorberbaum, 1987; Boyd and Richerson, 1988, 1992; Enquist and Leimar, 1993; Clutton-Brock and Parker, 1995; Leimar, 1997b; Joyce et al., 2006; Izquierdo et al., 2010, 2014). However, due to the rich diversity of repeated social interactions in nature, many open questions remain. For example, while many studies characterize conditions where a specific helping strategy is favoured by selection over defecting strategies, relatively few studies pit different helping strategies against one another in order to find which type is likely to be selected under those conditions. Additionally, once a strategy has evolved, the expressed behaviour may still largely depend on the environment under which it is expressed. In order to fully understand the diversity of cooperation and helping behaviours in nature, models that capture the essential aspects of the interactions are required.

Evolution of social strategies

Evolution is the change in allele frequency over generations (Schoener, 2011). The alleles, in concert with the environment, determine the phenotype of the individual (its body size, colour, etc). Individuals that are better adapted to the current environment will leave more offspring than others, and the heritable traits of those individuals will be represented more in the next generation. As evolution shapes any trait, so too does it shape social strategies (i.e., the specification of the actions an individual will take in social interactions, Fudenberg and Maskin, 1986). Since the actions an individual takes will influence its fitness, those individuals that take the "best" actions will produce more offspring, who in turn behave similar to their parents.

Evolutionary game theory is the study of the evolution of strategies in a population of interacting individuals (Maynard Smith, 1982; Hammerstein and Selten, 1994; Weibull, 1995; Hofbauer and Sigmund, 1998). It emerged from the field of game theory where the focus is on optimal strategic behaviour between rational agents (Osborne and Rubinstein, 1994; Myerson, 1997). If interactions are repeated, then an individual can respond to the behaviour of another. In other words, the actions an individual takes can be conditioned on the actions of others. If one just considers helping and defecting behaviours, an infinite set of strategies can be hypothesized. Actions can be taken probabilistically, conditional on the partner only, conditional on the outcome of previous interactions (taking into account both the partner and oneself), and the memory to remember the history of interactions can vary from absent to complete. In addition, other types of actions can be included in the model (for example, stop the interaction with the current partner). There thus exists a wide variety of potential social strategies. Understanding which ones are likely to evolve, and how this depends on the conditions under which the individuals interact is a key topic in evolutionary biology.

Partner control mechanisms

In repeated social interactions, an individual may have different options on how to respond to a defector, where a conditional response that reduces the partner's payoff relative to that of cooperating with it is defined as a partner control mechanism. The individual can condition its decision to help a partner based on whether the partner has helped the individual in previous interactions, which is called positive reciprocity. The most well-known strategy using positive reciprocity is tit-for-tat, which helps (cooperates in game theoretic terms) on the first interaction, while in following interactions it copies the actions of the partner in the previous interaction. In competition with a range of different strategies, tit-for-tat was found to do best on average (Axelrod and Hamilton, 1981). Consequently, a number of studies then identified conditions under which tit-for-tat and other positive reciprocal strategies could maintain cooperation in a population of interacting individuals, and how this may depend on group size, partner quality, cognitive constraints, and more (Boyd and Richerson, 1988; Leimar, 1997a; Fishman et al., 2001; Sherratt and Roberts, 2001; André, 2014).

While in game theoretic models individuals are assumed to be stuck with one another, in nature the interaction structure is more likely to represent a biological market where individuals may choose their preferred interaction partner (Noë and Hammerstein, 1994; Hammerstein and Noë, 2016). A partner control mechanism that allows an individual to interact with a variety of individuals is partner switching, where the individual can choose to stop the interaction with the defector and search for a new partner. Although partner switching is not an active partner choice mechanism, it still allows individuals to avoid prolonged interactions with specific individuals. Here, factors such as the cost of switching, the number of interactions, and others may determine whether a switching strategy evolves (Aktipis, 2004; Izquierdo et al., 2010, 2014; Enquist and Leimar, 1993; Hamilton and Taborsky, 2005; Joyce et al., 2006).

An individual may also respond to a defector by punishing it, i.e., pay a personal payoff cost that reduces the payoff of the defector (Clutton-Brock and Parker, 1995). Punishing strategies have been investigated widely in order to identify conditions where it may stabilize cooperation in a population. Among the examined factors are the reputation of the individual for being a punisher or not, the structure of the population, group size, the effect of retaliation, and more (Boyd and Richerson, 1992; Boyd et al., 2003; Janssen and Bushman, 2008; Santos et al., 2011; Shutters, 2011; Powers et al., 2012; Wolff, 2012).

Despite this wealth of models, there are relatively few that investigate when one type of response will yield more payoff for an individual (and thus a higher fitness) compared to others. However, in order to understand which conditions favour the evolution of which partner control mechanism, models are needed where the coevolution of partner control mechanisms is investigated.

Ecological influence on behaviour

Additionally, one can study how the ecology influences the behaviour of an individual (Maher and Lott, 2000). Substantial variation in behavioural patterns can be found between populations of the same species (Dall et al., 2004). Although some of this variation is due to genetic effects (Fuller, 1960), many species live in highly variable environments and constantly need to adjust their behaviour to the current state of the environment. Here, one can think of an individual that feeds in different patches. The energy intake can vary between patches, but will also decay over time (as food depletes). In order to optimize intake, individuals have to use a learning rule to determine the duration of time spend on a patch (McNamara and Houston, 1985). Also the social environment can vary between populations or within populations over time (Borgeaud et al., 2016). For an individual to navigate its actions when interacting repeatedly with both the social and non-social environment it is thus likely to use learning rule to optimize its behaviour (Harley, 1981; Börgers and Sarin, 1997; Camerer and Hua Ho, 1999). Therefore, an interesting area of research is to investigate how the state of the environment affects patterns of behaviour.

Primate grooming

The exchange of grooming in primates is an interesting example where the environment is likely to influence the grooming patterns. Grooming is a helpful act where individuals groom each other in order to remove ectoparasites and strengthen the social bond (Tanaka and Takefushi, 1993; Dunbar, 1991). Grooming is often found to be traded reciprocally (trading grooming for grooming), where two individuals in a dyad groom each other equally much over longer periods of time (Rowell et al., 1991; Barrett et al., 1999; Leinfelder et al., 2001; Pazol and Cords, 2005). Alternatively, grooming is regularly found to be directed up the hierarchy, where high ranked individuals receive more grooming than low ranked individuals (Seyfarth, 1980; Ventura et al., 2006; Wei et al., 2012; Roubová et al., 2015). It is argued that this is the result of food competition (van Schaik, 1989; Sterck et al., 1997; Kappeler and van Schaik, 2002). If food sources are clustered and monopolizable, then high ranked individuals are expected to defend food sources and exclude lower ranked ones from feeding there. In order to gain access to a food patch, a low ranked individual needs to groom the high ranked one in order to be tolerated on the patch (Ventura et al., 2006; Carne et al., 2011; Tiddi et al., 2011; Xia et al., 2012).

Although there is ample empirical evidence supporting both types of grooming (reciprocal grooming and grooming up the hierarchy), the conditions where either is expected to occur are based on verbal reasoning alone. Up until now, no quantitative model has shown that food competition can result in grooming being traded for tolerance, and thus there are no predictions as to how the intensity of food competition modulates the level of grooming up the hierarchy. Therefore, understanding how grooming patterns are affected by the ecology of the population is a key topic in primatology (Seyfarth, 1980; Schino, 2001).

Main question

In this work I study repeated social interactions using two different approaches. First, I study the coevolution of various conditional strategies. The question I address is: "Under which conditions will positive reciprocity, partner switching, or punishment be dominant in a population if these partner control mechanisms coevolve?". These three mechanisms comprise the main type of responses to stabilize cooperation in a two player game, and thus understanding the conditions where each is relatively favoured over the alternative ones is of critical importance. The second question is specifically addressed to primates: "What are the conditions that cause grooming to be directed up the hierarchy?". Grooming can be traded reciprocally (for grooming) or for rank-related benefits (e.g., tolerance on food patches). It is argued that food competition favours the latter type of trading, but no model has demonstrated this effect until now, and thus no quantitative predictions regarding how food competition modulates the grooming up the hierarchy exist.

Approach

The technique I use to study these questions is agent-based modelling. Agent-based modelling is a relatively recent scientific method, which has increased greatly in popularity in recent years (Niazi and Hussain, 2011). An agent-based model consists of a set of interacting agents (often individuals), where the agents make decisions (take actions) according to a specified set of rules, where the rules of interaction (e.g., model parameters that determine payoff) are specified by the experimenter (Bonabeau, 2002; Bankes, 2002; An et al., 2009). In this thesis I develop two types of agent-based models: an evolutionary one and a learning one. In the evolutionary model, agents collect payoff through their interaction with one another, where agents that accumulate most payoff generate most new agents (offspring) in a following time step (generation). As such, agents with the best decision rules (strategies) will proliferate over time, and by tracking which strategies are represented most in the population, one can make inferences regarding the adaptiveness of these strategies in nature. In the learning model, the agents change their behaviour by updating the probabilities of taking actions as a result of payoffs received. Here, the emergent property of the model is not which rule performs best, but the actions that are taken as a result of the state of the environment (the model parameters).

Thesis outline

Coevolution of partner control mechanisms

In the first chapter I study the coevolution of various conditional strategies in pairwise repeated interactions. For this, I develop an agent-based model where the interacting individuals have a variety of conditional responses to both cooperators and defectors. The strategy set includes strategies where individuals respond to defection with defection (positive reciprocity), strategies that punish defectors (punishment), and strategies that stop interacting with defectors and switch to a new partner (partner switching). Additionally, combinations of these responses are allowed (e.g., punish defectors and defect in response to a cooperator), while also allowing the same conditional responses to cooperators. The set of strategies is thus entirely unbiased. By simulating the evolution of the population where the individuals interact in pairs, I vary the conditions and by letting the population evolve I find which strategies (and what conditional responses) are best adapted to those conditions.

Chapter two is a continuation of the first chapter. With the same basic set-up, I add three factors that arguably add more realism to the model. These factors are: 1) selective survival of individuals over successive time steps, resulting in overlap in generations; 2) limited migration of offspring between demes, resulting in buildup of relatedness within demes; 3) a cognitive cost for strategies, that correlates with its complexity. In most species there is an obvious overlap of generations (parents and offspring coexisting at some point in time). Furthermore, offspring are usually not equally likely to interact with any other individual in the population, but often remain in the vicinity of their parents and thus will be interacting with closely related individuals (Chapuisat et al., 1997; Knight et al., 1999). Finally, research has shown that there are cognitive costs associated with strategies (Mery and Kawecki, 2003, 2004; Kotrschal et al., 2013). In the second chapter I therefore investigate how these factors in isolation or combined affect the competitiveness of the different partner control mechanisms: positive reciprocity, punishment, and partner switching.

Grooming up the hierarchy

In the third chapter I investigate the grooming behaviour in primates, and how grooming patterns are affected by the ecology of the population. For this, I develop an agent-based model where the agents (primates) choose whom to groom, where to feed, and whom to tolerate on food patches. The individuals take actions according to dynamic motivations for the various actions. Since primates live in highly variable environments (both food patches and the social environment can vary greatly in a primate's life), the individuals in the model use learning rules that determine how motivations are updated. Through learning the individuals update the motivations for the actions each interaction time step, where the updating depends on the payoff it received in the previous step. By varying the model parameters (such as the number of available food patches, the aggressiveness, and more) I investigate how the state of the environment determines the grooming patterns that emerge.

References

Aktipis, C. A. Know when to walk away: contingent movement and the evolution of cooperation. *Journal of Theoretical Biology*, 231(2):249–260, 2004.

An, G., Mi, Q., Dutta-Moscato, J., and Vodovotz, Y. Agent-based models in translational systems biology. *WIREs Syst Biol Med*, 1(2):159–171, 2009.

André, J. B. Mechanistic constraints and the unlikely evolution of reciprocal cooperation. *J. Evol. Biol.*, 27(4):784–795, 2014.

Arnold, K. E. and Owens, I. P. F. Cooperative breeding in birds: a comparative test of the life history hypothesis. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1398): 739–745, 1998.

Axelrod, R. and Hamilton, W. D. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

Bankes, S. C. Agent-based modeling: A revolution? *Proceedings of the National Academy of Sciences*, 99(suppl 3):7199–7200, 2002.

Barrett, L., Henzi, S. P., Weingrill, T., Lycett, J. E., and Hill, R. A. Market forces predict grooming reciprocity in female baboons. *Proceedings of the Royal Society of London B: Biological Sciences*, 266(1420):665–670, 1999.

Bonabeau, E. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7280–7287, 2002.

Borgeaud, C., Sosa, S., Bshary, R., Sueur, C., and van de Waal, E. Intergroup Variation of Social Relationships in Wild Vervet Monkeys: A Dynamic Network Approach. *Frontiers in Psychology*, 7, 2016.

Börgers, T. and Sarin, R. Learning Through Reinforcement and Replicator Dynamics. *Journal of Economic Theory*, 77(1):1–14, 1997.

Boyd, R. and Lorberbaum, J. P. No pure strategy is evolutionarily stable in the repeated Prisoner's Dilemma game. *Nature*, 327(6117):58–59, 1987.

Boyd, R. and Richerson, P. J. The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132(3):337–356, 1988.

Boyd, R. and Richerson, P. J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13(3):171–195, 1992.

Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, 100(6):3531–3535, 2003.

Bronstein, J. L. The exploitation of mutualisms. *Ecology Letters*, 4(3):277–287, 2001.

Bronstein, J. L. The evolution of facilitation and mutualism. *Journal of Ecology*, 97(6):1160–1170, 2009.

Bshary, R. and Bronstein, J. L. A General Scheme to Predict Partner Control Mechanisms in Pairwise Cooperative Interactions Between Unrelated Individuals. *Ethology*, 117(4):271–283, 2011.

Camerer, C. and Hua Ho, T. Experience-weighted Attraction Learning in Normal Form Games. *Econometrica*, 67(4):827–874, 1999.

Carne, C., Wiper, S., and Semple, S. Reciprocation and interchange of grooming, agonistic support, feeding tolerance, and aggression in semi-free-ranging Barbary macaques. *American Journal of Primatology*, 73(11):1127–1133, 2011.

Castro, J., Santiago, J., and Santana-Ortega, A. A general theory on fish aggregation to floating objects: An alternative to the meeting point hypothesis. 11(3):255–277, 2002.

Chapuisat, M., Goudet, J., and Keller, L. Microsatellites Reveal High Population Viscosity and Limited Dispersal in the Ant Formica paralugubris. *Evolution*, 51(2):475–482, 1997.

Clutton-Brock, T. H. and Parker, G. A. Punishment in animal societies. *Nature*, 373(6511): 209–216, 1995.

Dall, S. R. X., Houston, A. I., and McNamara, J. M. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, 7(8):734–739, 2004.

Dugatkin, L. A. The Evolution of Cooperation. *BioScience*, 47(6):355–362, 1997.

Dunbar, R. I. M. Functional Significance of Social Grooming in Primates. *Folia Primatologica*, 57 (3):121–131, 1991.

Enquist, M. and Leimar, O. The evolution of cooperation in mobile organisms. *Animal Behaviour*, 45(4):747–757, 1993.

Fishman, M. A., Lotem, A., and Stone, L. Heterogeneity Stabilizes Reciprocal Altruism Interactions. *Journal of Theoretical Biology*, 209(1):87–95, 2001.

Fudenberg, D. and Maskin, E. The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54(3):533–554, 1986.

Fuller, J. L. Behavior Genetics. *Annual Review of Psychology*, 11(1):41–70, 1960.

Griffin, A. S., West, S. A., and Buckling, A. Cooperation and competition in pathogenic bacteria. *Nature*, 430(7003):1024–1027, 2004.

Hamilton, I. M. and Taborsky, M. Contingent movement and cooperation evolve under generalized reciprocity. *Proceedings of the Royal Society B: Biological Sciences*, 272(1578):2259–2267, 2005.

Hamilton, W. D. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, 7 (1):1–16, 1964a.

Hamilton, W. D. The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, 7 (1):17–52, 1964b.

Hammerstein, P. and Noë, R. Biological trade and markets. *Phil. Trans. R. Soc. B*, 371(1687): 20150101+, 2016.

Hammerstein, P. and Selten, R. *Chapter 28 Game theory and evolutionary biology*, volume 2, pages 929–993. Elsevier, 1994.

Harley, C. B. Learning the evolutionarily stable strategy. *Journal of Theoretical Biology*, 89(4): 611–633, 1981.

Hofbauer, J. and Sigmund, K. *Evolutionary Games and Population Dynamics.* Cambridge University Press, 1 edition, 1998.

Izquierdo, L. R., Izquierdo, S. S., and Vega-Redondo, F. Leave and let leave: A sufficient condition to explain the evolutionary emergence of cooperation. *Journal of Economic Dynamics and Control*, 46:91–113, 2014.

Izquierdo, S. S., Izquierdo, L. R., and Vega-Redondo, F. The option to leave: Conditional dissociation in the evolution of cooperation. *Journal of Theoretical Biology*, 267(1):76–84, 2010.

Janssen, M. A. and Bushman, C. Evolution of cooperation and altruistic punishment when retaliation is possible. *Journal of Theoretical Biology*, 254(3):541–545, 2008.

Jennions, M. Cooperative breeding in mammals. *Trends in Ecology & Evolution*, 9(3):89–93, 1994.

Joyce, D., Kennison, J., Densmore, O., Guerin, S., Barr, S., Charles, E., and Thompson, N. S. My Way or the Highway: a More Naturalistic Model of Altruism Tested in an Iterative Prisoners' Dilemma. *Journal of Artificial Societies & Social Simulation*, 9(2), 2006.

Kappeler, P. M. and van Schaik, C. P. Evolution of Primate Social Systems. *International Journal of Primatology*, V23(4):707–740, 2002.

Kearns, C. A., Inouye, D. W., and Waser, N. M. Endangered Mutualisms: The Conservation of Plant-Pollinator Interactions. *Annual Reviews in Ecology and Systematics*, 29:83–112, 1998.

Knight, M. E., Oppen, M. J. H. V., Smith, H. L., Rico, C., Hewitt, G. M., and Turner, G. F. Evidence for male-biased dispersal in Lake Malawi cichlids from microsatellites. *Molecular Ecology*, 8(9):1521–1527, 1999.

Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., Immler, S., Maklakov, A. A., and Kolm, N. Artificial Selection on Relative Brain Size in the Guppy Reveals Costs and Benefits of Evolving a Larger Brain. *Curr Biol*, 23(2):168–171, 2013.

Lehmann, L. and Keller, L. The evolution of cooperation and altruism a general framework and a classification of models. *Journal of Evolutionary Biology*, 19(5):1365–1376, 2006.

Leimar, O. Reciprocity and communication of partner quality. *Proceedings of the Royal Society of London B: Biological Sciences*, 264(1385):1209–1215, 1997a.

Leimar, O. Repeated Games: A State Space Approach. *Journal of Theoretical Biology*, 184(4): 471–498, 1997b.

Leimar, O. and Hammerstein, P. Cooperation for direct fitness benefits. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553):2619–2626, 2010.

Leinfelder, I., de Vries, H., Deleu, R., and Nelissen, M. Rank and grooming reciprocity among females in a mixed-sex group of captive hamadryas baboons. *American Journal of Primatology*, 55(1):25–42, 2001.

Luce, R. D. and Raiffa, H. *Games and Decisions.* Whiley, New York, NY, USA, 1957.

Maher, C. R. and Lott, D. F. A Review of Ecological Determinants of Territoriality within Vertebrate Species. *The American Midland Naturalist*, 143(1):1–29, 2000.

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

McNamara, J. M. and Houston, A. I. Optimal foraging and learning. *Journal of Theoretical Biology*, 117(2):231–249, 1985.

Mery, F. and Kawecki, T. J. An operating cost of learning in Drosophila melanogaster. *Animal Behaviour*, 68(3):589–598, 2004.

Mery, F. and Kawecki, T. J. A fitness cost of learning ability in Drosophila melanogaster. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1532):2465–2469, 2003.

Myerson, R. B. *Game Theory: Analysis of Conflict*. Harvard University Press, 1st paperback edition edition, 1997.

Niazi, M. and Hussain, A. Agent-based computing from multi-agent systems to agent-based models: a visual survey. *Scientometrics*, pages 1–21, 2011.

Noë, R. and Hammerstein, P. Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. 35(1):1–11, 1994.

Osborne, M. J. and Rubinstein, A. *A Course in Game Theory*. The MIT Press, first edition, twelf edition, 1994.

Pazol, K. and Cords, M. Seasonal variation in feeding behavior, competition and female social relationships in a forest dwelling guenon, the blue monkey (Cercopithecus mitis stuhlmanni), in the Kakamega Forest, Kenya. *Behavioral Ecology and Sociobiology*, 58(6):566–577, 2005.

Powers, S. T., Taylor, D. J., and Bryson, J. J. Punishment can promote defection in group-structured populations. *Journal of Theoretical Biology*, 311:107–116, 2012.

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

Roubová, V., Konečná, M., Šmilauer, P., and Wallner, B. Whom to Groom and for What? Patterns of Grooming in Female Barbary Macaques (Macaca sylvanus). *PLoS ONE*, 10(2):e0117298+, 2015.

Rowell, T. E., Wilson, C., and Cords, M. Reciprocity and partner preference in grooming of female blue monkeys. 12(4):319–336, 1991.

Sachs, J. L., Mueller, U. G., Wilcox, T. P., and Bull, J. J. The Evolution of Cooperation. *The Quarterly Review of Biology*, 79(2):135–160, 2004.

Santos, M. D., Rankin, D. J., and Wedekind, C. The evolution of punishment through reputation. *Proceedings of the Royal Society B: Biological Sciences*, 278(1704):371–377, 2011.

Schino, G. Grooming, competition and social rank among female primates: a meta-analysis. *Animal Behaviour*, 62(2):265–271, 2001.

Schoener, T. W. The Newest Synthesis: Understanding the Interplay of Evolutionary and Ecological Dynamics. *Science*, 331(6016):426–429, 2011.

Seyfarth, R. M. The distribution of grooming and related behaviours among adult female vervet monkeys. *Animal Behaviour*, 28(3):798–813, 1980.

Sherratt, T. N. and Roberts, G. The importance of phenotypic defectors in stabilizing reciprocal altruism. *Behavioral Ecology*, 12(3):313–317, 2001.

Shutters, S. T. Punishment Leads to Cooperative Behavior in Structured Societies. *Evolutionary Computation*, 20(2):301–319, 2011.

Sterck, E. H. M., Watts, D. P., and van Schaik, C. P. The evolution of female social relationships in nonhuman primates. *Behavioral Ecology and Sociobiology*, 41(5):291–309, 1997.

Tanaka, I. and Takefushi, H. Elimination of External Parasites(Lice) Is the Primary Function of Grooming in Free-ranging Japanese Macaques. *Anthropological Science*, 101(2):187–193, 1993.

Tiddi, B., Aureli, F., Polizzi di Sorrentino, E., Janson, C. H., and Schino, G. Grooming for tolerance? Two mechanisms of exchange in wild tufted capuchin monkeys. *Behavioral Ecology*, 22(3):663–669, 2011.

van Schaik, C. P. The ecology of social relationships amongst female primates. In Standen, V. and Foley, R. A., editors, *Comparative socioecology. The behavioural ecology of humans and other mammals.*, pages 195–218. Blackwell, Oxford, 1989.

Ventura, R., Majolo, B., Koyama, N. F., Hardie, S., and Schino, G. Reciprocation and interchange in wild Japanese macaques: grooming, cofeeding, and agonistic support. *American Journal of Primatology*, 68(12):1138–1149, 2006.

Wei, W., Qi, X.-G., Guo, S.-T., Zhao, D.-P., Zhang, P., Huang, K., and Li, B.-G. Market Powers Predict Reciprocal Grooming in Golden Snub-Nosed Monkeys (Rhinopithecus roxellana). *PLoS ONE*, 7(5):e36802+, 2012.

Weibull, J. W. *Evolutionary game theory*, volume 31. MIT Press, Cambridge, MA, 1995.

West, S. A., Griffin, A. S., and Gardner, A. Evolutionary Explanations for Cooperation. *Current Biology*, 17(16):R661–R672, 2007.

Wolff, I. Retaliation and the role for punishment in the evolution of cooperation. *Journal of Theoretical Biology*, 315:128–138, 2012.

Xia, D., Li, J., Garber, P. A., Sun, L., Zhu, Y., and Sun, B. Grooming Reciprocity in Female Tibetan Macaques Macaca Thibetana. *American Journal of Primatology*, 74(6):569–579, 2012.

Chapter one

Coevolution between positive reciprocity, punishment, and partner switching in repeated interactions

Status: Published in Proceedings of the Royal Society B
Wubs M, Bshary R, Lehmann L. 2016 Coevolution between positive reciprocity, punishment, and partner switching in repeated interactions. *Proc. R. Soc. B* **283**: 20160488.

Contributions: All authors contributed to the conceptual design and the writing of the manuscript. MW wrote the code and performed the analysis.

Abstract

Cooperation based on mutual investments can occur between unrelated individuals when they are engaged in repeated interactions. Individuals then need to use a conditional strategy to deter their interaction partners from defecting. Responding to defection such that the future payoff of a defector is reduced relative to cooperating with it is called a partner control mechanism. Three main partner control mechanisms are (i) to switch from cooperation to defection when being defected ("positive reciprocity"), (ii) to actively reduce the payoff of a defecting partner ("punishment"), or (iii) to stop interacting and switch partner ("partner switching"). However, such mechanisms to stabilize cooperation are often studied in isolation from each other. In order to better understand the conditions under which each partner control mechanism tends to be favoured by selection, we here analyse by way of individual-based simulations the coevolution between positive reciprocity, punishment, and partner switching. We show that random interactions in an unstructured population and a high number of rounds increase the likelihood that selection favours partner switching. In contrast, interactions localized in small groups (without genetic structure) increase the likelihood that selection favours punishment and/or positive reciprocity. This study thus highlights the importance of comparing different control mechanisms for cooperation under different conditions.

Introduction

Interactions where all participants gain a direct net fitness benefit, namely cooperation, are widespread in natural populations (Dugatkin, 1997). Many cases of cooperation involve investments; that is, the reduction of current personal payoff by some amount in order to increase the partner's payoff. This observation raises the question how individuals can ensure that their investments yield future benefits; that is, how they can avoid being defected by their partner over repeated bouts of interactions. When individuals engage in repeated interactions over their lifespan, the evolution of cooperation is often modelled as an iterated Prisoner's Dilemma game where individuals have to choose whether to cooperate or defect at each interaction stage. The payoffs are such that mutual cooperation yields a higher payoff than mutual defection, while to defect yields a higher payoff than to cooperate in each single round, irrespective of the partner's action, hence the dilemma. In order to deter a partner from defecting and stabilizing cooperation in a repeated game, an individual can use a conditional strategy that reduces a defecting partner's payoff relative to that of cooperating with it. We define the broad type of such a conditional response as a partner control mechanism (Bshary and Bronstein, 2011).

Different types of partner control mechanisms have been proposed to stabilize cooperation in the repeated Prisoner's Dilemma game. Perhaps the most well known is positive reciprocity, where cooperative acts are reciprocated by cooperation in future interactions, whereas defection is not, thus making defection unfavourable in the long run. An often-studied strategy using positive reciprocity as a partner control mechanism is tit-for-tat (TFT), which starts by cooperating and then in subsequent rounds implements the previous action of the partner (Axelrod and Hamilton, 1981; Kreps et al., 1982; Rubinstein, 1986). Although positive reciprocity is often favoured by selection in evolutionary models (Axelrod and Hamilton, 1981; Boyd and Richerson, 1988; Leimar, 1997; André and Day, 2007), its relevance outside humans has been questioned (Hammerstein, 2003, but see Raihani and Bshary, 2011; Taborsky et al., 2016).

Another partner control mechanism is punishment, which comes at an immediate payoff cost to the actor, but also reduces the payoff of a defector relative to cooperating (Clutton-Brock and Parker, 1995; Nakamaru and Iwasa, 2006; Powers et al., 2012). Although punishment thus comes at a cost to the punisher, this can be overcome if punishment results in the partner being more cooperative in the long run. Punishment can be favoured by selection in evolutionary models of repeated interactions (Clutton-Brock and Parker, 1995), and examples of punishment as a partner control mechanism can be found in natural populations (reviewed in Raihani et al., 2012).

Still another partner control mechanism is partner switching (Enquist and Leimar, 1993; Joyce et al., 2006; McNamara et al., 2008; Izquierdo et al., 2010, 2014). By partner switching an individual can avoid being exploited by a defector by simply stopping the interaction. Although switching entails an opportunity cost because it necessitates finding a new partner, it has been shown to be favoured by selection in the iterated Prisoner's Dilemma game (McNamara et al., 2008), and several examples of partner switching have been suggested in nature (Cresswell, 1999; Bshary and Schäffer, 2002; Schwagmeyer, 2014).

For individuals interacting in an iterated Prisoner's Dilemma game positive reciprocity, punishment, and partner switching are predicted as main partner control mechanisms capable of stabilizing cooperation (Bshary and Bronstein, 2011). However, the evolution of these three main types of partner control mechanisms for cooperation is generally investigated in isolation from each other. It thus remains unclear under which conditions selection will favour one mechanism over another. More recently, however, different partner control mechanisms have been investigated together (Joyce et al., 2006; Izquierdo et al., 2010). In a landmark study, Izquierdo et al. (Izquierdo et al., 2010) have shown that selection favours partner switching over TFT. However, this study has assumed that switching does not incur any costs, it excluded the strategic option to punish partners, and restricted the analysis to a population with random interactions only, which are all factors that may change which mechanism is favoured by selection. In order to predict which partner control mechanisms are likely to be observed in natural populations, it is important to consider the coevolution of positive reciprocity, partner switching, and punishment, and understand the conditions under which one partner control mechanism is favoured over the others by selection.

Here, we present an evolutionary model where we let positive reciprocity, punishment, and partner switching coevolve when interactions are random in the population and when they occur in groups in a panmictic population (i.e., no genetic structure within groups, Haystack model of population structure Maynard Smith, 1964). The aim of this study is to identify the partner control mechanisms favoured under different conditions, and we therefore chose the Prisoner's Dilemma as a payoff matrix for the pairwise interactions, where defection always yields a higher single round payoff, and thus selection for responding to defection is strong. We explore the role of the proximate costs and benefits of cooperation, punishment, and switching on these dynamics, as well as the role of interactions localized to groups and the duration of punishment. Our results show that, when interactions occur at random between all population members, the likelihood that partner switching is favoured by selection increases if the number of interactions in an individual's lifespan increases. However, when interactions are localized to groups, we find that punishment generally dominates in sizable groups, unless punishment efficiency is reduced. In the latter case, we do find conditions where positive reciprocity outcompetes alternative partner control mechanisms, but we were unable to identify a particular factor that would consistently favour it.

The model

Population and lifecycle

We consider a haploid population of constant size with a total number of $N = d \times n$ adult individuals, which are subdivided into d groups of equal size n. The lifecycle is marked by the following events. First, group members interact socially with each other and accumulate payoffs. Next, each individual produces a large number of offspring proportionally to accumulated payoff, and dies. Finally, offspring disperse randomly (with probability $1/d$ to a given group, including the natal one) and compete randomly with exactly n individuals reaching adulthood in each group. Hence, the population is panmictic (no genetic structure will be obtained).

Social interactions

In the social interaction phase of the lifecycle, individuals play a repeated game for T rounds, whose stage game consists of a pairwise extensive-form game (see Fudenberg and Tirole, 1996 for a description of different types of games). The per-round extensive-form game consists of five sequential moves where the individuals of a pair choose actions simultaneously during each move (Fig. 1), and where pair rematching may occur during each round, as follows.

Move 0: random pairing. Each unpaired individual (all individuals in the first round) gets randomly paired with another unpaired individual. Individuals cannot influence this process, i.e., there is no partner choice.

Move 1: the Prisoner's Dilemma. Each individual in a pair can either cooperate (action C) or defect (action D). To cooperate means paying a payoff cost C_{h} to contribute a payoff benefit B_{h} to the partner, whereas defection has no effect on payoff.

Move 2: leaving. Each individual can either leave its partner (action L) or stay (action S) and a pairbond is broken if at least one individual leaves. A payoff cost of C_{l} is paid by both individuals of a broken pair and only unbroken pairs are engaged in the forthcoming move 3 and 4, otherwise, individuals are added to a pool of individuals that will be paired in move 0 of the next round.

Move 3: punishment. Each individual in a pair can either opt to punish its partner (action P) or not punish (action N). Playing action P incurs a payoff cost C_p to self and reduces by D_p the payoff of the partner. Only punished individuals enter the next move.

Move 4: response to punishment. A punished individual has three possible (re)actions available. (i) It receives the punishment but "ignores" it and does not change any future action if the pairbond is maintained (action I). (ii) The individual leaves its partner, namely it expresses action L as in move 2 with the same payoff consequences. (iii) The individual alters its behaviour (action A), which means that, if it played action D (C) in move 1, it will cooperate (defect) in the next z rounds in move 1. An individual that has switched to defection (cooperation) owing to punishment and is punished again, will again change its behaviour in move 1 for z rounds.

In addition to a fixed cost C_l of partner switching, we also consider an alternative cost function for individuals that leave (or were left) in either move 2 or 4, where the cost depends on the number of unpaired individuals at the end of a round. For this, we consider the function

$$C_\mathrm{l}(i) = \frac{a}{1+i^k}, \tag{1}$$

which decreases as the number i of unpaired individuals in the population increases, where $a > 0$ determines the maximum cost, and $k > 0$ the shape. Thus, we assume that if a larger number of individuals is searching for a partner, then the cost of finding a partner is reduced.

Strategies

We assume that individuals use pure strategies, which deterministically specify the actions to be taken at moves 1-4 of the stage game, possibly conditionally on past actions. The strategy of an individual for the entire game is specified by a vector $s = (x_1, x_2, x_3, x_4)$, where x_k represents the move-wise strategy the individual uses when faced with a choice at move $k \in \{1, 2, 3, 4\}$.

In the supplementary material, table S1, we list all move-wise strategies, which are obtained as follows. We assume that the strategy for move 1 specifies an action taken when the individual first interacts with its partner, and an action taken in subsequent rounds is conditioned on what the partner did in the previous round in move 1. This move-wise strategy can thus be written as $a_1 a_\mathrm{C} a_\mathrm{D}$, where $a_1, a_\mathrm{C}, a_\mathrm{D} \in \{\mathrm{C}, \mathrm{D}\}$. Here, a_1 is the action taken the first time the two individuals in a pair interact, a_C is the action taken if the partner cooperated in the previous round, and a_D is the action taken if the partner defected in the previous round. We thus have a total of 8 (2^3) move-wise strategies for move 1: {CCC, CCD, CDC, CDD, DCC, DCD, DDC, DDD}.

For move 2, the decision to leave or stay is assumed to be conditional on the action taken by the partner in move 1 of the current round. Hence, the move-wise strategy can be written as $a_\mathrm{C} a_\mathrm{D}$, where $a_\mathrm{C} \in \{\mathrm{L}, \mathrm{S}\}$ ($a_\mathrm{D} \in \{\mathrm{L}, \mathrm{S}\}$) gives the action taken when the partner cooperated (defected), whereby $x_2 \in \{\mathrm{LL}, \mathrm{LS}, \mathrm{SL}, \mathrm{SS}\}$.

Likewise, for move 3, the decision to punish or not to punish the partner is assumed to be conditional on the action taken by the partner in move 1, so that the move-wise strategy is $a_\mathrm{C} a_\mathrm{D}$, where $a_\mathrm{C} \in \{\mathrm{P}, \mathrm{N}\}$ ($a_\mathrm{D} \in \{\mathrm{P}, \mathrm{N}\}$) is the action taken when the partner cooperated (defected), whereby $x_3 \in \{\mathrm{PP}, \mathrm{PN}, \mathrm{NP}, \mathrm{NN}\}$. Importantly though, we assume that if an individual punishes its partner in this move and the pair is not broken in the next move, then the individual expresses in move 1 of the next round the same action it expressed in this round. This is assumed to avoid individuals responding to the action of the partner both by punishing and by (possibly) changing their own action in move 1 of the following round, and thus take two conditional actions as a response to one action of its partner. Because we want to compare strategies that differ in their response to defecting individuals, we did not allow individuals that punish in the current round to take a conditional action in move 1 of the following round. Finally, the response to punishment in move 4 is simply given by $x_4 \in \{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$.

Removing phenotypically indistinguishable strategies

As there are eight different alternatives for x_1, 4 for x_2 and x_3, 3 for x_4 (see supplementary material, table S1), there is a total of 384 strategies. However, given the setup of our model, many strategies in the strategy space are phenotypically indistinguishable. By phenotypically indistinguishable strategies, we mean those strategies that at no point in the game would act differently from one another, and so will be neutral in an evolutionary model. Therefore, to decrease the complexity of the model, we removed strategies from the strategy space as follows. Per set of phenotypically indistinguishable strategies, only one strategy was used. For example, consider the set of strategies with the same move-wise strategy for move 1 (e.g., $x_1 = \text{CCC}$) and that always leaves the partner in move 2 ($x_2 = \text{LL}$). Strategies from this set never reach move 3 and 4, and thus will always behave similarly, despite having different move-wise strategies for these moves. The 92 strategies that remain after removing phenotypically indistinguishable strategies are shown in the supplementary material (table S3).

Pooling strategies into classes

Although there are many strategies in the model, we are mainly interested in cooperative strategies that differ in their response to defection, i.e., cooperative strategies using different partner control mechanisms. A cooperative strategy is defined as a strategy that, when paired with another cooperative strategy, will always cooperate in move 1 of the game, without punishing or leaving the partner. Within the set of cooperative strategies, we can distinguish between classes of strategies that differ in their partner control mechanism: no response (no control), conditional play in the Prisoner's Dilemma (move 1), leaving (move 2), or punishment (move 3). Each of these four classes consists of three strategies that differ only in their response to punishment (move 4). Because we are interested in comparing partner control mechanisms, when comparing frequencies of strategies, we will do so according to class, i.e., in our analysis, we will always pool the frequencies of the strategies belonging to the same class.

Here, we will give a verbal description of each of the six classes of strategies that we consider (supplementary material, table S2). Each strategy of the *positive reciprocity* class (denoted $\mathcal{R}_c$) cooperates on the first interaction. It cooperates in subsequent rounds if the partner cooperated in the previous round and defects if the partner defected in the previous round, without leaving or punishing the partner. Each strategy of the *partner switching* class (denoted $\mathcal{S}_c$) cooperates on the first round, cooperates if the partner cooperates, does not punish, but leaves as soon as the partner defects. Each strategy of the *punishment* class (denoted $\mathcal{P}_c$) cooperates on the first round, cooperates in subsequent rounds, does not leave, but punishes a partner that defects. Each strategy of the *always cooperate* (denoted $\mathcal{C}_c$) and *always defect* class (denoted $\mathcal{D}_c$) always cooperates (defects), and does not express any conditional play in move 1-3. The remaining $92 - 5 \times 3 = 77$ strategies will be pooled in "*rest*".

Analyses

In order to analyse the model, we used individual-based simulations to track the frequencies of the six classes of strategies ($\mathcal{R}_c$, $\mathcal{S}_c$, $\mathcal{P}_c$, $\mathcal{C}_c$, $\mathcal{D}_c$, and "*rest*") in the population over generations. Strategies are assumed to be inherited from parent to offspring with probability $1 - \mu$. With probability μ, the offspring mutates to another strategy taken at random among all remaining strategies. To form the next generation of offspring, we use multinomial sampling over the aggregate payoff of each strategy type of the parental generation with a baseline payoff guaranteeing there can be no negative payoff (Wright-Fisher process, Ewens, 2004).

For all reported results (Figs. 2 and 3), we ran the simulations for 10^6 generations and computed the time average frequency of the six classes of strategies starting with uniformly sampled initial frequencies. We also evaluated the total frequency of cooperation in the population, which we define as the average frequency over the whole population and length of the repeated game of the pairs of individuals in the population where both individuals in the pair cooperated in the Prisoner's Dilemma game.

Results

We first present results assuming that the population consists of a single group ($d = 1, n = 10000$), so that the pairing process (move 0, Fig. 1) is random at the population level. We will refer to this as the well-mixed case. Then, we introduce group structure ($d = 250$, varying n), where the pairing process occurs at the group level but with otherwise similar parameters to show how this factor alters the relative effectiveness of each partner control mechanism.

Well-mixed population

Our results are based on the following baseline parameter values: $B_{\mathrm{h}} = 2, C_{\mathrm{h}} = 1, D_{\mathrm{p}} = 2, C_{\mathrm{p}} = 1, C_{\mathrm{l}} = 1, \mu = 0.01$, whereas we let T vary between 1 and 30 (table 1) and set $z = T$, so that the behavioural change after punishment lasts indefinitely. We find that the average frequency of cooperation in the population is strongly dependent on the number of rounds (T) per generation (Fig. 2a, black line). When the game is one shot ($T = 1$), conditional strategies are unable to affect payoff or behaviour in future rounds, and thus cooperation is selected against (less than 1%), which is consistent with the standard result that defection is favoured in such cases (Axelrod and Hamilton, 1981). As the number of rounds is increased, the frequency of cooperation quickly increases, with more than 90% of mutual cooperation for $T \geq 6$.

Additionally, we find that the number of rounds has a strong influence on which partner control mechanism is favoured by selection. Our main results are as follows.

For $4 \leq T \leq 6$, we find that the positive reciprocity class ($\mathcal{R}_c$) is dominant (Fig. 2a). Here, the number of rounds is very low, and thus the costs of punishment or partner switching in the first rounds cannot be negated in later rounds of mutual cooperation. Switching to defection to minimize payoff losses is more beneficial for the lifetime payoff and thus the $\mathcal{R}_c$ class is selected for.

For intermediate T ($7 \leq T \leq 9$), we find that the punishment class ($\mathcal{P}_c$) dominates (Fig. 2a). Although $\mathcal{R}_c$ and $\mathcal{P}_c$ strategies gain equal payoffs when paired with each other, their respective payoff gain will differ considerably when paired with a defector. An $\mathcal{R}_c$ strategy switches to defection when paired with a defector resulting in both players gaining the baseline payoff. A $\mathcal{P}_c$ strategy, however, continues to cooperate while punishing defection. If the recipient of punishment switches to play cooperate, then through several rounds of mutual cooperation, a $\mathcal{P}_c$ strategy is likely to obtain more payoff than an $\mathcal{R}_c$ strategy. This difference in payoff between $\mathcal{R}_c$ and $\mathcal{P}_c$ when matched with defectors may thus explain why for a higher number of rounds of interaction selection will favour the $\mathcal{P}_c$ class over the $\mathcal{R}_c$ class. However, not all strategies respond to punishment by altering behaviour, and thus $\mathcal{P}_c$ strategies cannot force all individuals to cooperate. Some partnerships can therefore be very costly for these individuals as they pay double costs (cooperating and punishing).

Finally, for large T ($T \geq 10$), we find that the switching class ($\mathcal{S}_c$) dominates the population (Fig. 2a). Strategies in the $\mathcal{S}_c$ class do not face the problem of prolonged costly partnerships as they will always leave uncooperative individuals. Two $\mathcal{S}_c$ strategies will therefore always manage to find each other in a well-mixed population, given enough rounds. When the number of rounds increases, $\mathcal{S}_c$ strategies will have more rounds to reap the benefits of mutual cooperation once a cooperative partner has been found, and thus the $\mathcal{S}_c$ class outcompetes both the $\mathcal{R}_c$ and the $\mathcal{P}_c$ class for $T \geq 10$. If the cost of switching is increased to $C_{\mathrm{l}} = 5$, however, then the number of rounds needed for the $\mathcal{S}_c$ class to dominate is increased to $T \geq 70$ (Fig. 2b). In all simulations where $d = 1$ (single group), we find that switching is generally favoured when T is large enough. The finding that a high number of rounds favours partner switching is robust even when the cost of switching increases exponentially with fewer number of unpaired individuals (using eq. 1, $a = 100, k = 0.9$, Fig. 2c; see the supplementary material, Fig. S2 for other parameter values).

Group structured population

We now introduce group structure (without genetic structure as dispersal is random to any group) into the population, setting the number of groups (d) to 250 while varying group size (n). Otherwise, we use the same set of parameter values as in the baseline case for the well-mixed population ($B_\mathrm{h} = 2, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, \mu = 0.01$, Fig. 2a) with $T = 30$ and $z = T$. Our main aim is to determine the conditions where the $\mathcal{S}_c$ class dominates in frequency.

Interestingly, switching only dominates in very large groups ($n \geq 300$, Fig. 3a). Instead, we find that the $\mathcal{P}_c$ class is dominant for any group size lower than 300. The $\mathcal{P}_c$ class coexists in these simulations with a strategy that always defects, punishes other defectors, and alters behaviour if punished. While the $\mathcal{P}_c$ individuals can force such individuals to cooperate, other strategies will either be exploited or punished.

To determine the robustness of the result that the $\mathcal{P}_c$ class tends to dominate in a group structured population, we relaxed the assumption of punishment altering behaviour for the lifetime of the individual (in move 4). Such a strong effect of punishment is unlikely to occur in nature, and punished individuals may attempt to defect again after several interactions. We find that the evolutionary success of punishment is strongly dependent on this parameter. If $z = 5$, then the $\mathcal{P}_c$ class is still dominant in groups up to a size of 52 (Fig. 3b). In larger groups however, it is first the $\mathcal{R}_c$ class that dominates, whereas for $n \geq 76$, the $\mathcal{S}_c$ class is dominant. Strikingly, if the cost of switching partner is absent as well ($C_\mathrm{l} = 0$), the $\mathcal{S}_c$ class is still outcompeted by the $\mathcal{P}_c$ class in small groups ($n \leq 28$, Fig. 3c). This may stem from the fact that if individuals interact in small groups, a partner switcher may be rematched with the individual it left on the previous round and may end up repeatedly interacting with the same defector (despite switching every round). The $\mathcal{P}_c$ class therefore still dominates in small groups, because its payoff is mostly dependent on how a defecting individual responds to punishment, but not on the composition of the group it is in. This effect largely persists in a structured population if T is small, unless the cost of punishment is doubled, in which case the $\mathcal{R}_c$ class takes over (supplementary material, Fig. S4).

Sensitivity analysis

To test the robustness of the various results presented here, we have performed additional analyses testing a larger part of the parameter space adding up to at least 15000 different parameter combinations for which we have run simulations. The results of these analyses are presented in the supplementary material.

Discussion

Cooperative individuals can use partner control mechanisms; that is, broad types of conditional strategies to reduce the lifetime payoff of defectors relative to cooperators. Three partner control mechanisms (positive reciprocity, punishment, and partner switching) have all been shown to be able to stabilize cooperation in panmictic populations in separate models (Axelrod and Hamilton, 1981; García and Traulsen, 2012; Izquierdo et al., 2014). However, few studies have investigated under which conditions selection would favour one partner control mechanism over another. Here, we have addressed this issue by investigating the coevolution of these three control mechanisms in a panmictic population in which the interaction structure is either well-mixed (i.e., all individuals are potential partners) or group structured with interactions occurring only locally among a small number of individuals (with no genetic structure within groups). In most simulations, we find a polymorphism where the different classes of strategies coexist. However, it is clear that under most conditions a specific class of strategies tends to be favoured by selection over alternatives and thus dominates in this polymorphism.

Our key result for the well-mixed case is that the likelihood of partner switching being favoured by selection over positive reciprocity, punishment, and defection increases if the number of rounds of interaction is larger (Fig. 2 and supplementary material, Figs. S1-S3). For fewer number of rounds punishment and positive reciprocity tend to be favoured, but which of the two classes dominates depends on changes in various parameters, and thus no general conclusion can be reached here. When interactions are localized to the group level, punishment is relatively more favoured in small and moderately sized groups for otherwise similar parameter values as in the well-mixed interactions case, and this is for both small and large number of rounds (Figs. 3a and S4a). Positive reciprocity dominates under certain conditions in a group structured population when punishment efficiency is reduced; for example, for a high number of rounds, intermediate group size, and a low duration of the effect of punishment (Fig. 3b), or for a low number of rounds and high cost of punishment (supplementary material, Fig. S4c). We did not, however, identify a specific factor that would consistently induce positive reciprocity to dominate the other control mechanisms. In the following, we will first discuss each control mechanism separately and then evaluate how our results connect to empirical research.

Switching

In our analysis, partner switching emerges as the dominant partner control mechanism when many potential partners exist and many interactions take place during an individuals lifespan, unless the cost of switching is high and the number of rounds of interaction is insufficiently large to compensate for these costs. These results make intuitive sense if one considers how the three control mechanisms respond to unconditional defectors: punishers and positive reciprocators may spend their entire life with a defecting partner, whereas partner switchers leave and will invariably end up with another cooperative individual and hence reap the benefits of cooperation as long as enough rounds are played. Izquierdo et al. (2010) have already shown that partner switching is a powerful partner control mechanism stabilizing cooperation; if it is cost-free, then it dominates over positive reciprocity. Our results extend their insights by showing that switching can be favoured by selection over not only positive reciprocity, but also punishment in a well-mixed population, with the caveat that a sufficient number of rounds of interaction must take place.

Switching (when linked to cooperation) is a cognitively simple strategy that, via the exploration of partner behaviour, rejects defectors and tends to assort with cooperators. It can thus be regarded as a primitive form of partner choice. Although more active mechanisms of partner choice exist, such as using information about past behaviour of individuals or other signals of cooperative behaviour (Ashlock et al., 1996; Janssen, 2008), partner switching allows individuals to respond to variation in the population in the same way. This generally tends to stabilize cooperation because, if individuals can exert some level of choice in the presence of variation of the expression of cooperation, the system of interacting individuals functions as a biological market where cooperators end up assorted with themselves (Johnstone and Bshary, 2008; McNamara and Leimar, 2010).

A critical result of our model, however, is that the size of the interaction group has a clear impact on the likelihood of a partner switcher to find the right partner, and thus the evolutionary success of partner switching. Relaxing the assumption of well-mixed interaction opportunities (McNamara et al., 2008; Izquierdo et al., 2010), we find that the prevalence of partner switching diminishes the smaller the number of potential interaction partners gets. This conclusion holds even if partner switching is free of opportunity costs (Fig. 3c). The reason for this result is that the smaller the group the more likely it becomes that switchers can only be rematched with their defecting partner as nobody else is available. In other words, the market for interaction partners becomes increasingly restricted with decreasing numbers of potential interaction partners.

Punishment

Via punishment an individual can actively attempt to change the behaviour of its partner, by paying a small payoff cost to reduce the payoff of its defecting partner, thereby making cooperation more attractive. Punishment is more favoured when the population is group structured (compared with unstructured), up to relatively large group sizes, especially if punishment results in the defecting recipient changing its behaviour to cooperation indefinitely ($z = T$, Fig. 3a). Importantly, a punisher can induce cooperative behaviour in a conditionally defecting partner but switchers cannot, which gives punishment an advantage when the number of potential partners and hence the number of unmatched cooperators is limited. For the same reason, punishment outcompetes positive reciprocity for various parameter value combinations, because within the limits of the strategy space explored in this paper, the behaviour of the partner and focal individual can be more easily aligned through punishment than through positive reciprocity. Therefore, we find in group structured populations that selection generally favours punishment over positive reciprocity and partner switching in sizable groups (Fig. 3). If one of the parameters influencing punishment efficiency is changed (i.e., high cost of punishment, low payoff reduction for the recipient of punishment, or short behavioural change after being punished), then we find that alternative classes of strategies dominate (supplementary material, Fig. S5).

Positive reciprocity

The conditions where positive reciprocity is favoured over punishment and partner switching are less easily characterized. Although in group structured populations we find that punishment dominates often in sizable groups (Fig. 3), when punishment efficiency is decreased, there are various conditions where positive reciprocity dominates instead (Fig. 3b, and supplementary material, Figs. S4 and S5). However, depending on the number of rounds of interaction, cost of partner switching, and other parameters we also find conditions where the always defect class or the switching class dominates in the population (supplementary material, Fig. S5). In sum, there is not a specific factor that would consistently increase the likelihood of positive reciprocity dominating the population.

Our analyses suggest that $\mathcal{R}_c$ strategies may often be outcompeted by other control mechanisms, because $\mathcal{R}_c$ individuals paired with defectors are unable to reach the cooperative outcome (both individuals play C in move 1). That is, there exists no strategy in our strategy set that would exploit unconditional cooperators, but that can also "identify" the $\mathcal{R}_c$ strategy and cooperate with it. Such strategies would require several rounds of interaction (and thus a large memory) to identify that the partner is playing TFT. Punishment, on the other hand, is a much more direct signal (a single punishing act) to which defectors can respond. Thus, if strategy complexity is limited to one round of memory, then the $\mathcal{S}_c$ and $\mathcal{P}_c$ class can still reach the cooperative outcome when paired with a defector, but the $\mathcal{R}_c$ class cannot. Therefore, even though the $\mathcal{R}_c$ class avoids being exploited by defectors by switching to play defect as well, it gains less payoff than other classes of control mechanisms and is thus frequently outcompeted. This does not necessarily mean that positive reciprocity can never be favoured. As the results show, we have found conditions where positive reciprocity dominates (Figs. 2a and 3b, and supplementary material). More importantly, however, our results show that the deterministic play and a single round of memory of our $\mathcal{R}_c$ class (as in the TFT strategy) causes it to often be outcompeted by classes of strategies that do manage to reach a cooperative outcome with their partners. Therefore, for positive reciprocity to evolve, it is likely necessary that strategies evolve that take into account a larger history of the interaction or play less deterministically.

Connection to the empirical literature

It is still a largely unanswered question of how frequently each of the three partner control mechanisms investigated here occurs in natural populations. According to current evidence, there are very few examples for punishment (Raihani et al., 2012), while there are various examples for positive reciprocity (Taborsky et al., 2016). Regarding partner switching, we are aware only of clear interspecific examples where partner switching in response to defection occurs. For example, in an interspecific interaction between client and cleaner fish, it has been observed that client reef fish with access to several cleaning stations use a partner switching strategy in response to a defecting client even though they could alternatively use punishment - as clients without choice options do (Bshary and Schäffer, 2002; Bshary and Grutter, 2005). Our model is, however, limited to intraspecific interactions, and thus it remains to be investigated how much our results would be affected if interacting individuals belong to two separate gene pools. In intraspecific contexts, empirical tests of biological market theory focus on individuals actively choosing a partner prior to interactions based on a comparison of offers (Noë et al., 1991; Hammerstein and Noë, 2016), rather than on leaving a partner that has defected. Investigating active choice rather than partner switching would be another interesting avenue for future research.

Our result that partner switching does not perform well in small groups (and hence for low behavioural variation) is of potential importance for empirical research on cooperation in stable groups, as is often found in primates. It has been proposed that various trades of investments in primates (e.g., grooming, tolerance, and support in agonistic encounters) are stable against defection partially because of partner switching (Schino and Aureli, 2016). However, it has also been argued that social bonds in primate groups are highly differentiated where individuals form long-term social bonds with particular individuals in the group (Silk et al., 2009). In such groups, partner switching may be highly restricted. Hence, our model suggests that partner switching cannot be accepted as a default partner control mechanism in stable groups without convincing empirical evidence.

The most surprising result of our analyses is the success of punishment in sizable groups, as the evidence for this partner control mechanism in symmetric two-player interactions is rather rare (Raihani et al., 2012). One reason for its success is the assumption that any player can use punishment in a relative cost-efficient way, i.e., the cost of punishing is lower than the cost of being punished. In nature, cost efficiency is likely linked to asymmetries between players and hence asymmetric games. Fittingly, experimental evidence for punishment has been reported for asymmetric games in interspecific interactions (Bshary and Grutter, 2005; Bshary and Bshary, 2010), and the most important intraspecific context involves the “pay-to-stay” concept where helpers help and show appeasement apparently to avoid aggression by dominant breeders (Fischer et al., 2014). A major problem with asymmetric strength is that it may turn a cooperation game in which punishment stabilizes cooperation into an exploitation game in which dominants coerce subordinates (Clutton-Brock and Parker, 1995), i.e., defect while forcing the partner to cooperate. For example, only larger male cleaner wrasse punish their smaller female partners for cheating a joint client, a game akin to an iterated Prisoner’s Dilemma (Bshary et al., 2008; Raihani et al., 2010). To fully understand the effect of asymmetries between individuals on the relative effectiveness of punishment over other partner control mechanisms, this will need to be modelled explicitly, however. In addition, further work is needed to determine how factors such as asymmetries or relatedness between interacting individuals may change the adaptiveness of each partner control mechanism.

References

André, J.-B. and Day, T. Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemma. *Journal of Theoretical Biology*, 247(1):11–22, 2007.

Ashlock, D., Smucker, M. D., Stanley, and Tesfatsion, L. Preferential partner selection in an evolutionary study of Prisoner's Dilemma. *Biosystems*, 37(1-2):99–125, 1996.

Axelrod, R. and Hamilton, W. D. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

Boyd, R. and Richerson, P. J. The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132(3):337–356, 1988.

Bshary, A. and Bshary, R. Self-Serving Punishment of a Common Enemy Creates a Public Good in Reef Fishes. *Current Biology*, 20(22):2032–2035, 2010.

Bshary, R. and Bronstein, J. L. A General Scheme to Predict Partner Control Mechanisms in Pairwise Cooperative Interactions Between Unrelated Individuals. *Ethology*, 117(4):271–283, 2011.

Bshary, R. and Grutter, A. S. Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biology Letters*, 1(4):396–399, 2005.

Bshary, R. and Schäffer, D. Choosy reef fish select cleaner fish that provide high-quality service. *Animal Behaviour*, 63(3):557–564, 2002.

Bshary, R., Grutter, A. S., Willener, A. S. T., and Leimar, O. Pairs of cooperating cleaner fish provide better service quality than singletons. *Nature*, 455(7215):964–966, 2008.

Clutton-Brock, T. H. and Parker, G. A. Punishment in animal societies. *Nature*, 373(6511): 209–216, 1995.

Cresswell, J. E. The influence of nectar and pollen availability on pollen transfer by individual flowers of oil-seed rape (Brassica napus) when pollinated by bumblebees (Bombus lapidarius). *Journal of Ecology*, 87(4):670–677, 1999.

Dugatkin, L. A. The Evolution of Cooperation. *BioScience*, 47(6):355–362, 1997.

Enquist, M. and Leimar, O. The evolution of cooperation in mobile organisms. *Animal Behaviour*, 45(4):747–757, 1993.

Ewens, W. J. *Mathematical Population Genetics.* Springer-Verlag, New York, 2004.

Fischer, S., Zöttl, M., Groenewoud, F., and Taborsky, B. Group-size-dependent punishment of idle subordinates in a cooperative breeder where helpers pay to stay. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1789):20140184+, 2014.

Fudenberg, D. and Tirole, J. *Game Theory.* MIT Press, Massachusetts, 1996.

García, J. and Traulsen, A. Leaving the loners alone: Evolution of cooperation in the presence of antisocial punishment. *Journal of Theoretical Biology*, 307:168–173, 2012.

Hammerstein, P. *Genetic and Cultural Evolution of Cooperation.* MIT Press, Cambridge, MA, 2003.

Hammerstein, P. and Noë, R. Biological trade and markets. *Phil. Trans. R. Soc. B*, 371(1687): 20150101+, 2016.

Izquierdo, L. R., Izquierdo, S. S., and Vega-Redondo, F. Leave and let leave: A sufficient condition to explain the evolutionary emergence of cooperation. *Journal of Economic Dynamics and Control*, 46:91–113, 2014.

Izquierdo, S. S., Izquierdo, L. R., and Vega-Redondo, F. The option to leave: Conditional dissociation in the evolution of cooperation. *Journal of Theoretical Biology*, 267(1):76–84, 2010.

Janssen, M. A. Evolution of cooperation in a one-shot Prisoner's Dilemma based on recognition of trustworthy and untrustworthy agents. *Journal of Economic Behavior & Organization*, 65(3-4): 458–471, 2008.

Johnstone, R. A. and Bshary, R. Mutualism, market effects and partner control. *Journal of Evolutionary Biology*, 21(3):879–888, 2008.

Joyce, D., Kennison, J., Densmore, O., Guerin, S., Barr, S., Charles, E., and Thompson, N. S. My Way or the Highway: a More Naturalistic Model of Altruism Tested in an Iterative Prisoners' Dilemma. *Journal of Artificial Societies & Social Simulation*, 9(2), 2006.

Kreps, D. M., Milgrom, P., Roberts, J., and Wilson, R. Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27(2):245–252, 1982.

Leimar, O. Repeated Games: A State Space Approach. *Journal of Theoretical Biology*, 184(4): 471–498, 1997.

Maynard Smith, J. Group Selection and Kin Selection. *Nature*, 201(4924):1145–1147, 1964.

McNamara, J. M. and Leimar, O. Variation and the response to variation as a basis for successful cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553): 2627–2633, 2010.

McNamara, J. M., Barta, Z., Fromhage, L., and Houston, A. I. The coevolution of choosiness and cooperation. *Nature*, 451(7175):189–192, 2008.

Nakamaru, M. and Iwasa, Y. The coevolution of altruism and punishment: Role of the selfish punisher. *Journal of Theoretical Biology*, 240(3):475–488, 2006.

Noë, R., Schaik, C. P., and Hooff, J. A. R. A. M. The Market Effect: an Explanation for Pay-off Asymmetries among Collaborating Animals. *Ethology*, 87(1-2):97–118, 1991.

Powers, S. T., Taylor, D. J., and Bryson, J. J. Punishment can promote defection in group-structured populations. *Journal of Theoretical Biology*, 311:107–116, 2012.

Raihani, N. J. and Bshary, R. Resolving the iterated prisoner's dilemma: theory and reality. *Journal of Evolutionary Biology*, 24(8):1628–1639, 2011.

Raihani, N. J., Grutter, A. S., and Bshary, R. Punishers Benefit From Third-Party Punishment in Fish. *Science*, 327(5962):171, 2010.

Raihani, N. J., Thornton, A., and Bshary, R. Punishment and cooperation in nature. *Trends in Ecology & Evolution*, 27(5):288–295, 2012.

Rubinstein, A. Finite automata play the repeated prisoner's dilemma. *Journal of Economic Theory*, 39(1):83–96, 1986.

Schino, G. and Aureli, F. Reciprocity in group-living animals: partner control versus partner choice. *Biol Rev*, page n/a, 2016.

Schwagmeyer, P. L. Partner switching can favour cooperation in a biological market. *Journal of Evolutionary Biology*, 27(9):1765–1774, 2014.

Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., Wittig, R. M., Seyfarth, R. M., and Cheney, D. L. The benefits of social capital: close social bonds among female baboons enhance offspring survival. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1670):3099–3104, 2009.

Taborsky, M., Frommen, J. G., and Riehl, C. Correlated pay-offs are key to cooperation. *Phil. Trans. R. Soc. B*, 371(1687):20150084+, 2016.

Table and figures

Table 1: List of parameters

Parameter	**Meaning**
B_{h}	benefit to the recipient of a cooperative act
C_{h}	cost of a cooperative act
D_{p}	payoff reduction for target of punishment
C_{p}	cost of punishment
C_{l}	cost of switching partner
z	duration of punishment
d	number of groups
n	group size
T	number of rounds in one generation
μ	mutation rate
N	population size
a, k	used to calculate the cost of switching in eq. 1

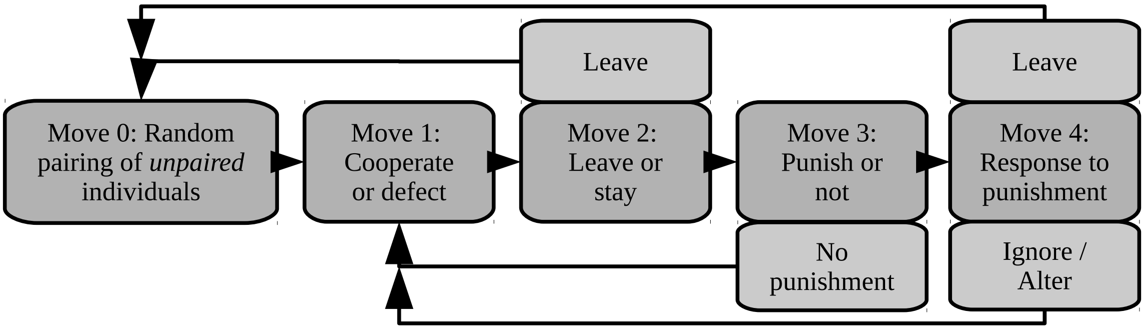


Figure 1: Chain of events per generation. The five moves are repeated for T rounds. After looping T times, the next generation is produced and the parental one dies.

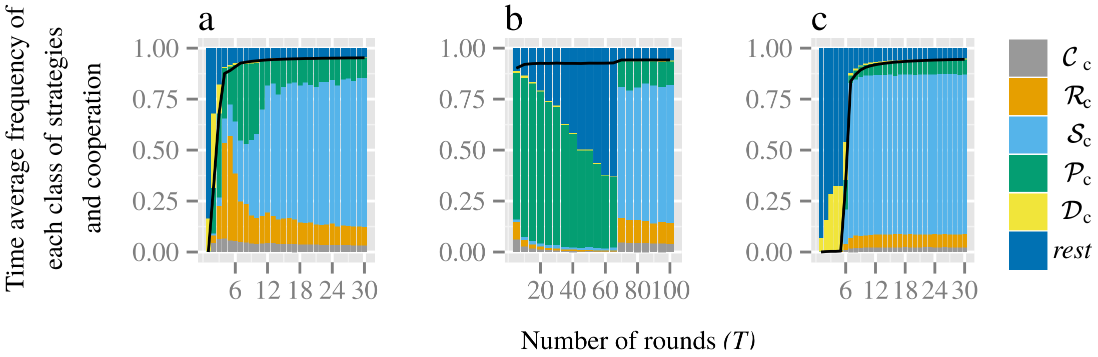


Figure 2: Time average frequency (over 10^6 generations) of the frequency of cooperation (black line) and the six classes of strategies ($\mathcal{R}_c$, $\mathcal{S}_c$, $\mathcal{P}_c$, $\mathcal{C}_c$, $\mathcal{D}_c$, and "*rest*") plotted as a function of the number of rounds T of the repeated game. Parameter values: $B_\text{h} = 2, C_\text{h} = 1, D_\text{p} = 2, C_\text{p} = 1, z = T, \mu = 0.01, d = 1, n = 10000$. Panel specific parameters: $C_\text{l} = 1$ (panel a), $C_\text{l} = 5$ (panel b), using eq. 1 with $a = 100, k = 0.9$ (panel c).

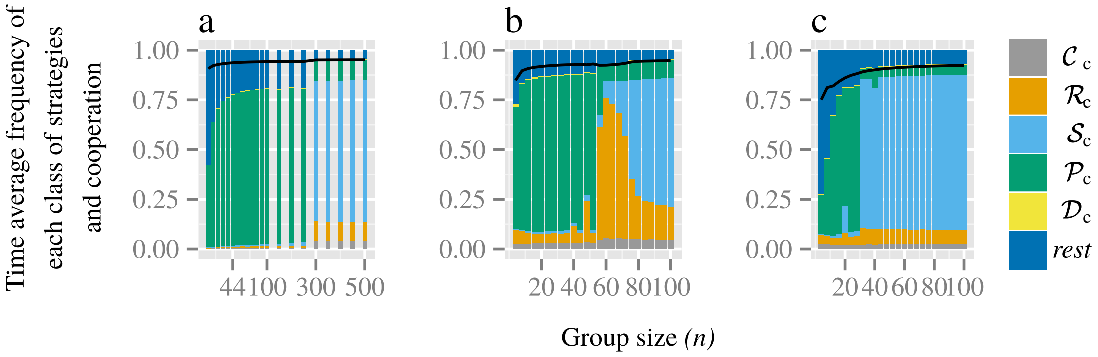


Figure 3: Time average frequency (over 10^6 generations) of the frequency of cooperation (black line) and the six classes of strategies ($\mathcal{R}_c$, $\mathcal{S}_c$, $\mathcal{P}_c$, $\mathcal{C}_c$, $\mathcal{D}_c$, and "*rest*") plotted as a function of group size n. Parameter values: $B_\text{h} = 2, C_\text{h} = 1, D_\text{p} = 2, C_\text{p} = 1, T = 30, \mu = 0.01, d = 250$. Panel specific parameters: $C_\text{l} = 1, z = T$ (panel a), $C_\text{l} = 1, z = 5$ (panel b), $C_\text{l} = 0, z = 5$ (panel c).

Supplementary material

Strategy set

In table S1 the move-wise strategies for each move of the stage game are listed, while table S2 lists the main classes of strategies.

In table S3 the full set of strategies is given. In the "Strategies" section of the main text the coding of the strategies is explained. Although the setup of our model allows for a great number of strategies, we have reduced the strategy space by removing phenotypically indistinguishable strategies (see also main text section "Removing phenotypically indistinguishable strategies"). For example, if the strategy specifies to always leave in move 2, regardless of the partner's action, it can never punish the partner, since the pair will be broken up and thus the actions specified for this move (punish or not punish) will never be played. Some strategies are therefore phenotypically indistinguishable in our model, and per set of phenotypically indistinguishable strategies only one strategy was used in the strategy space.

An X in a move-wise strategy in table S3 is a placeholder for a conditional action that is not played at any point during the game (C or D in move 1; P or N in move 3). Two strategies that are otherwise similar, but differing in this action would therefore never act differently (i.e., they are phenotypically indistinguishable). Only one of these two strategies is included in the strategy space. This can occur in two situations. First, if an individual leaves after the partner cooperates (or defects), then it cannot also punish or conditionally cooperate/defect in the following round since the pair is broken up, and thus an X is shown in place of the conditional action in both the move-wise strategy for move 1 and 3. Second, we assumed that a punishing act could not be followed by a conditional action in move 1 of the following round, and thus the action P in the move-wise strategy for move 3 is always combined with an X in the move-wise strategy for move 1.

Sensitivity analysis

Parameter exploration in a well-mixed population

To test the robustness of our main result (the $\mathcal{S}_c$ class dominates for high T), we explored the parameter space more thoroughly than presented in the main text. The exploration was done as follows. The parameters B_h and D_p were set either to 2, 2.5, or 3, the parameters C_h and C_p were set either to 0.5, 1, or 1.5, and the cost of leaving (C_l) was set either to 0, 1, or 2. For each possible combination of these parameters we ran simulations with the number of rounds T ranging from 1 to 100.

In Fig. S1 the results for four combinations of parameters are shown. If the $B_\mathrm{h}/C_\mathrm{h}$ ratio is low ($B_\mathrm{h}/C_\mathrm{h} = 2/1.5$, panel a), we find that a higher number of rounds is required to reach high levels of cooperation compared to the baseline case (*cf.* main text Fig. 2a). Here, the $\mathcal{S}_c$ class is the dominant class of cooperative strategies. If the $B_\mathrm{h}/C_\mathrm{h}$ ratio is high ($B_\mathrm{h}/C_\mathrm{h} = 3/0.5$, panel b), only a few number of rounds is required for high levels of cooperation. The $\mathcal{S}_c$ dominates here for $T \geq 3$. If there is no cost of leaving ($C_\mathrm{l} = 0$, panel c), the $\mathcal{S}_c$ class is always the dominant cooperative class of strategies. If punishment is very effective in terms of payoff, i.e., the $D_\mathrm{p}/C_\mathrm{p}$ ratio is high ($D_\mathrm{p}/C_\mathrm{p} = 3/0.5$, panel d), the $\mathcal{P}_c$ class dominates for a larger range of T, compared to our baseline case (main text). However, for $T > 40$ the $\mathcal{S}_c$ class is again often found to be dominant.

In short, for any parameter combination we find that, all else being equal, the relative competitiveness of the $\mathcal{S}_c$ class increases with increasing number of rounds.

Dynamic cost of leaving: In this section we investigate additional parameter values for a and k when using eq. 1 from the main text to calculate the cost of leaving. The results are plotted in Fig. S2. Again we find that, although the exact number of rounds where the $\mathcal{S}_c$ class becomes dominant depends both on a and k, the $\mathcal{S}_c$ class will always dominate for high T. This suggests that the $\mathcal{S}_c$ class can evolve irrespective of the underlying cost function.

Large T: To check to what extent the $\mathcal{S}_c$ class dominates for large T, we ran simulations with T up to 1000, while otherwise using our baseline set of parameters. We find that, although the $\mathcal{S}_c$ class is by far the most prevalent, other classes of cooperative strategies are maintained above frequencies higher than what would be expected from the mutation rate alone (Fig. S3). This results from the fact that the classes of cooperative strategies are nearly neutral in a population consisting of mainly those classes ($\mathcal{C}_c$, $\mathcal{R}_c$, $\mathcal{S}_c$, and $\mathcal{P}_c$), since these strategies will gain exactly the same payoff when paired with one another. Therefore, through genetic drift the different cooperative strategies may invade one another. However, uncooperative individuals will continue to enter the population via mutation. When paired with uncooperative individuals the cooperative strategies will respond differently, and thus gain different payoffs. Here, strategies of the $\mathcal{C}_c$ class will then be strongly selected against, but also strategies of the $\mathcal{R}_c$ and $\mathcal{P}_c$ class gain less payoff on average than those of the $\mathcal{S}_c$ class. Thus, through a mutation-selection-drift balance this polymorphism is maintained in the population.

Parameter exploration in a group-structured population

In this section we present additional results for the group-structured case.

Group-structure with small T: In this section, we tested if the $\mathcal{P}_c$ class is also relatively more favoured by selection in a structured population if T is small. For $T = 4$ in the well-mixed case the $\mathcal{R}_c$ class dominates the population (Fig. 2a). Using the same parameters as in the well-mixed case (except $T = 4, d = 250$, varying n), we find again that the $\mathcal{P}_c$ class is relatively more favoured by selection (Fig. S4a). Only in very large groups the $\mathcal{R}_c$ class outcompetes the $\mathcal{P}_c$ class ($n \geq 350$).

Interestingly, reducing D_p (the payoff reduction of being punished) did not affect the frequency of the $\mathcal{P}_c$ class, but instead negatively affected the frequency of the $\mathcal{R}_c$ class ($D_\mathrm{p} = 1.2$, Fig. S4b). This was due to $\mathcal{P}_c$ strategies having less impact on various defector strategies, which consequently increased in frequency, which in turn negatively affected the $\mathcal{R}_c$ class, but not the $\mathcal{P}_c$ class. However, when increasing the cost of punishment (C_p) the $\mathcal{P}_c$ class disappeared from the population and the $\mathcal{R}_c$ class dominated for all but the smallest groups ($C_\mathrm{p} = 2, n \geq 16$, Fig. S4c). This discrepancy with D_p is due to C_p affecting the payoff of the $\mathcal{P}_c$ class directly, while changing D_p affects recipients of punishment instead.

Further parameter exploration: In this section we present results from a larger parameter space for the group-structured population.

First, we determined the minimum group size where the $\mathcal{S}_c$ class dominates in conditions where the $\mathcal{P}_c$ is not dominant. To achieve this we used the same parameters as in Fig. 3c from the main text with $z = 3$. Using these parameters we find that the $\mathcal{P}_c$ class no longer dominates, and instead the "rest" class and the $\mathcal{R}_c$ class dominate in small groups (Fig. S5a). The $\mathcal{S}_c$ class is now dominant in the population for $n \geq 20$ (compared to $n \geq 32$ for $z = 5$ in Fig. 3c in the main text). Thus, even in unfavourable conditions for the $\mathcal{P}_c$ class, the $\mathcal{S}_c$ class does not dominate in populations where interactions occur in small groups.

In Fig. S5b we use the same parameters as in panel a, but with $C_\mathrm{l} = 1, T = 7$ in order to determine if the $\mathcal{S}_c$ class can dominate in a group-structured population when T is low. However, we find here, similarly to our well-mixed population (Fig. 2a main text), that if T is low the $\mathcal{S}_c$

class is outcompeted by the $\mathcal{P}_c$ and $\mathcal{R}_c$ classes for group size smaller than 120, and by the $\mathcal{D}_c$ class in larger groups. This confirms that the $\mathcal{S}_c$ class needs a critical number of rounds in order to be favoured by selection over the other classes.

In Fig. S5c we use the same parameters as Fig. 4b from the main text, with $D_{\mathrm{p}} = 1$. This shows that, although reducing D_{p} initially affects only the frequency of the $\mathcal{S}_c$ and $\mathcal{R}_c$ classes (main text), if D_{p} is too low then the $\mathcal{P}_c$ disappears completely from the population and the $\mathcal{D}_c$ class is dominant for all group sizes.

In Fig. S5d we use the same parameters as in Fig. 4c from the main text, but with $T = 7$. In the well-mixed population we have found that a higher number of rounds generally favours the $\mathcal{P}_c$ class over the $\mathcal{R}_c$ class. Similarly, in Fig. 4c (main text) we find the $\mathcal{R}_c$ class dominates if punishment is costly to the punisher ($C_{\mathrm{p}} = 2$). However, if the number of rounds is increased, then we find that the $\mathcal{P}_c$ is dominant for group size $n \leq 60$. This confirms that also in the a group-structured population, all else being equal, a higher number of rounds increases the relative competitiveness of the $\mathcal{P}_c$ class over the $\mathcal{R}_c$ class.

Coevolution of the $\mathcal{P}_c$ class and response to punishment

In this section we show that the response to punishment by altering behaviour ($x_4 = \mathrm{A}$) co-evolves with the $\mathcal{P}_c$ class. To demonstrate this we used the same data as our baseline case (main text Fig. 2a), but pooled the time average frequency of all strategies into three groups, based on their response to punishment (ignore, alter, or leave) (Fig. S6). The frequency of each group is plotted together with the frequency of the $\mathcal{P}_c$ class. The results show that if $\mathcal{P}_c$ strategies are frequent, then strategies that alter behaviour after punishment ($x_4 = \mathrm{A}$) are also selected for.

Table S 1: The sets of move-wise strategies for each move of the stage game. The coding of strategies is explained in the "Strategies" section.

Move 1:	$\{\mathrm{CCC}, \mathrm{CCD}, \mathrm{CDC}, \mathrm{CDD}, \mathrm{DCC}, \mathrm{DCD}, \mathrm{DDC}, \mathrm{DDD}\}$
Move 2:	$\{\mathrm{LL}, \mathrm{SS}, \mathrm{LS}, \mathrm{SL}\}$
Move 3:	$\{\mathrm{PP}, \mathrm{NN}, \mathrm{PN}, \mathrm{NP}\}$
Move 4:	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$

Table S 2: Main classes of strategies.

Name	Move 1	Move 2	Move 3	Move 4
Always cooperate $\mathcal{C}_c$	CCC	SS	NN	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$
Positive reciprocity $\mathcal{R}_c$	CCD	SS	NN	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$
Partner switching $\mathcal{S}_c$	CCC	SL	NN	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$
Punishment $\mathcal{P}_c$	CCC	SS	NP	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$
Always defect $\mathcal{D}_c$	DDD	SS	NN	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$

Table S 3: Full set of strategies used in all simulations. The coding is explained in the main text.

Strategy	**Move 1**	**Move 2**	**Move 3**	**Move 4**	**Strategy**	**Move 1**	**Move 2**	**Move 3**	**Move 4**
1	CCC	SS	NN	I	47	DCC	SS	NN	I
2	CCC	SS	NN	A	48	DCC	SS	NN	A
3	CCC	SS	NN	L	49	DCC	SS	NN	L
4	CCD	SS	NN	I	50	DCD	SS	NN	I
5	CCD	SS	NN	A	51	DCD	SS	NN	A
6	CCD	SS	NN	L	52	DCD	SS	NN	L
7	CCX	SL	NX	I	53	DCX	SL	NX	I
8	CCX	SL	NX	A	54	DCX	SL	NX	A
9	CCX	SL	NX	L	55	DCX	SL	NX	L
10	CCX	SS	NP	I	56	DCX	SS	NP	I
11	CCX	SS	NP	A	57	DCX	SS	NP	A
12	CCX	SS	NP	L	58	DCX	SS	NP	L
13	CDC	SS	NN	I	59	DDC	SS	NN	I
14	CDC	SS	NN	A	60	DDC	SS	NN	A
15	CDC	SS	NN	L	61	DDC	SS	NN	L
16	CDD	SS	NN	I	62	DDD	SS	NN	I
17	CDD	SS	NN	A	63	DDD	SS	NN	A
18	CDD	SS	NN	L	64	DDD	SS	NN	L
19	CDX	SL	NX	I	65	DDX	SL	NX	I
20	CDX	SL	NX	A	66	DDX	SL	NX	A
21	CDX	SL	NX	L	67	DDX	SL	NX	L
22	CDX	SS	NP	I	68	DDX	SS	NP	I
23	CDX	SS	NP	A	69	DDX	SS	NP	A
24	CDX	SS	NP	L	70	DDX	SS	NP	L
25	CXC	LS	XN	I	71	DXC	LS	XN	I
26	CXC	LS	XN	A	72	DXC	LS	XN	A
27	CXC	LS	XN	L	73	DXC	LS	XN	L
28	CXD	LS	XN	I	74	DXD	LS	XN	I
29	CXD	LS	XN	A	75	DXD	LS	XN	A
30	CXD	LS	XN	L	76	DXD	LS	XN	L
31	CXX	LL	XX	I	77	DXX	LL	XX	I
32	CXX	LS	XP	I	78	DXX	LS	XP	I
33	CXX	LS	XP	A	79	DXX	LS	XP	A
34	CXX	LS	XP	L	80	DXX	LS	XP	L
35	CXC	SS	PN	I	81	DXC	SS	PN	I
36	CXC	SS	PN	A	82	DXC	SS	PN	A
37	CXC	SS	PN	L	83	DXC	SS	PN	L
38	CXD	SS	PN	I	84	DXD	SS	PN	I
39	CXD	SS	PN	A	85	DXD	SS	PN	A
40	CXD	SS	PN	L	86	DXD	SS	PN	L
41	CXX	SL	PX	I	87	DXX	SL	PX	I
42	CXX	SL	PX	A	88	DXX	SL	PX	A
43	CXX	SL	PX	L	89	DXX	SL	PX	L
44	CXX	SS	PP	I	90	DXX	SS	PP	I
45	CXX	SS	PP	A	91	DXX	SS	PP	A
46	CXX	SS	PP	L	92	DXX	SS	PP	L

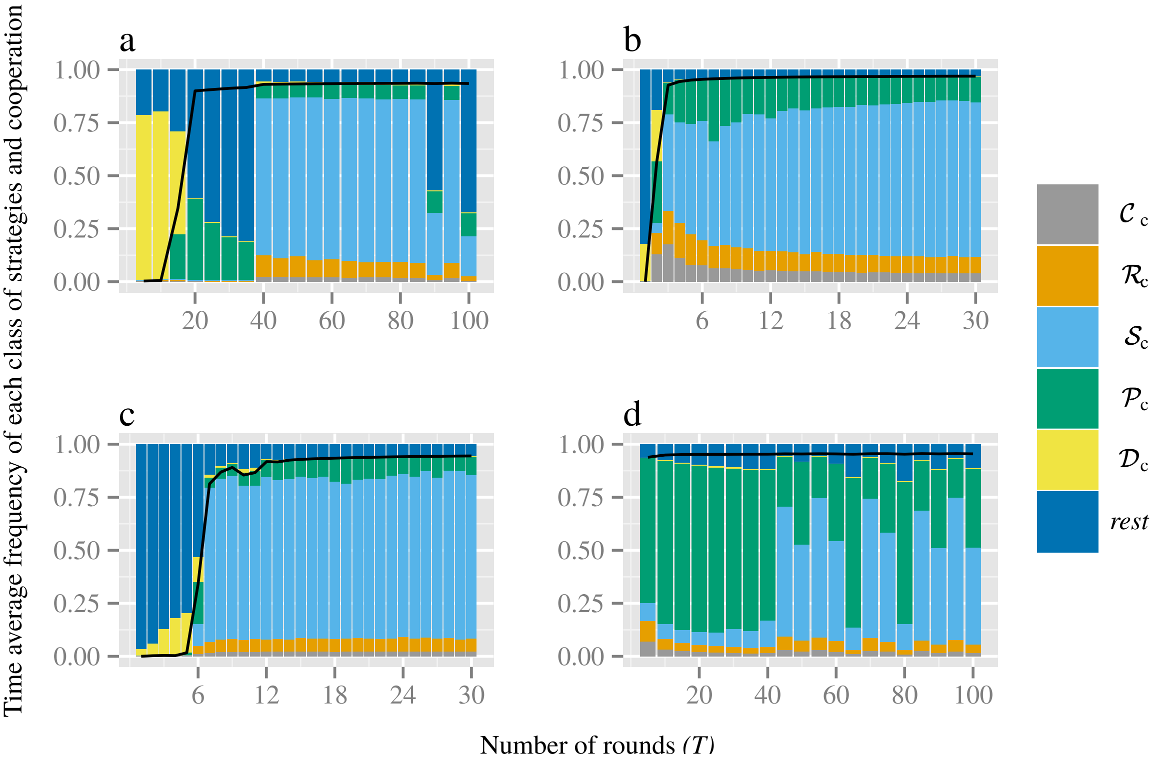


Figure S 1: Time average frequency (over 10^6 generations) of the frequency of cooperation (black line) and the six classes of strategies ($\mathcal{R}_c$, $\mathcal{S}_c$, $\mathcal{P}_c$, $\mathcal{C}_c$, $\mathcal{D}_c$, and "*rest*") plotted as a function of the number of rounds T of the repeated game. Parameter values: $B_\mathrm{h} = 2, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, z = T, \mu = 0.01, d = 1, n = 10000$. Panel specific parameters: $C_\mathrm{h} = 1.5$ (panel a), $B_\mathrm{h} = 3, C_\mathrm{h} = 0.5$ (panel b), $C_\mathrm{l} = 0$ (panel c), $D_\mathrm{p} = 3, C_\mathrm{p} = 0.5$ (panel d).

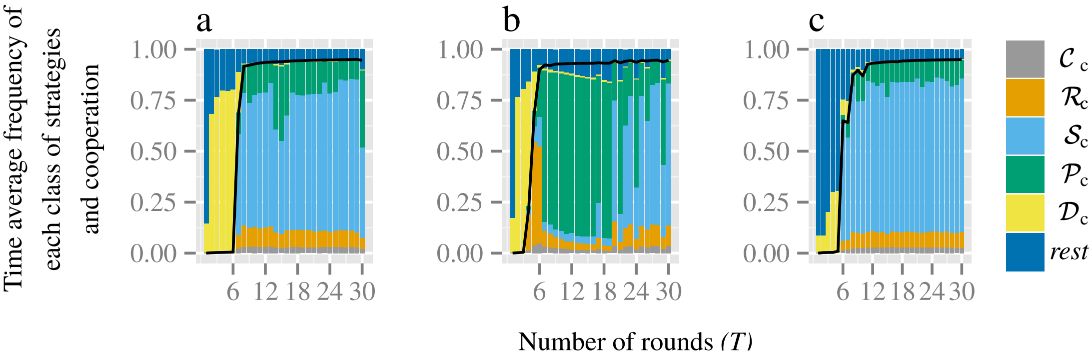


Figure S 2: Time average frequency (over 10^6 generations) of the frequency of cooperation (black line) and the six classes of strategies ($\mathcal{R}_c$, $\mathcal{S}_c$, $\mathcal{P}_c$, $\mathcal{C}_c$, $\mathcal{D}_c$, and "*rest*") plotted as a function of the number of rounds T of the repeated game. Parameter values: $B_\mathrm{h} = 2, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, z = T, \mu = 0.01, d = 1, n = 10000$, using eq. 1 to calculate the cost of switching. Panel specific parameters: $a = 50, k = 0.8$ (panel a), $a = 100, k = 0.8$ (panel b), $a = 100, k = 1$ (panel c).

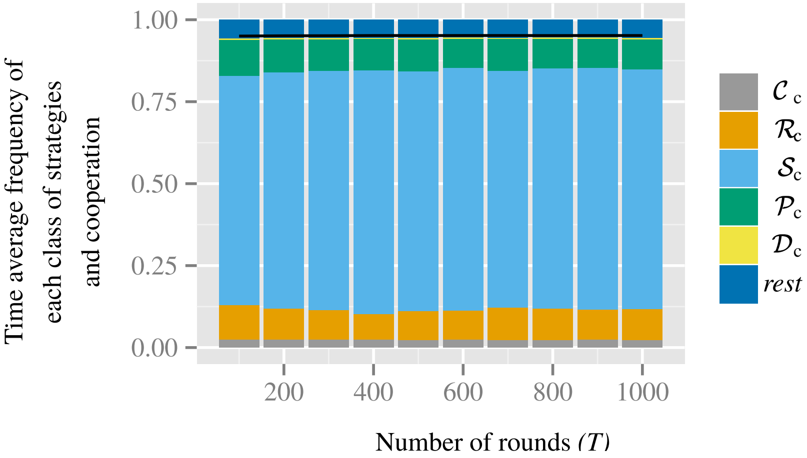


Figure S 3: Same as the baseline case (Fig. 2a, main text) but with a high number of rounds T.

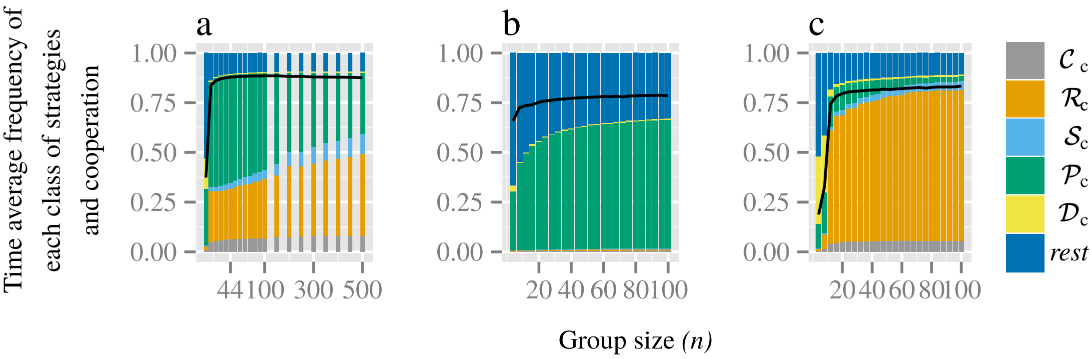


Figure S 4: Time average frequency (over 10^6 generations) of the frequency of cooperation (black line) and the six classes of strategies ($\mathcal{R}_c$, $\mathcal{S}_c$, $\mathcal{P}_c$, $\mathcal{C}_c$, $\mathcal{D}_c$, and "*rest*") plotted as a function of group size n. Parameter values: $B_\mathrm{h} = 2, C_\mathrm{h} = 1, C_\mathrm{l} = 1, T = 4, z = T, \mu = 0.01, d = 250$. Panel specific parameters: $D_\mathrm{p} = 2, C_\mathrm{p} = 1$ (panel a), $D_\mathrm{p} = 1.2, C_\mathrm{p} = 1$ (panel b), $D_\mathrm{p} = 2, C_\mathrm{p} = 2$ (panel c)

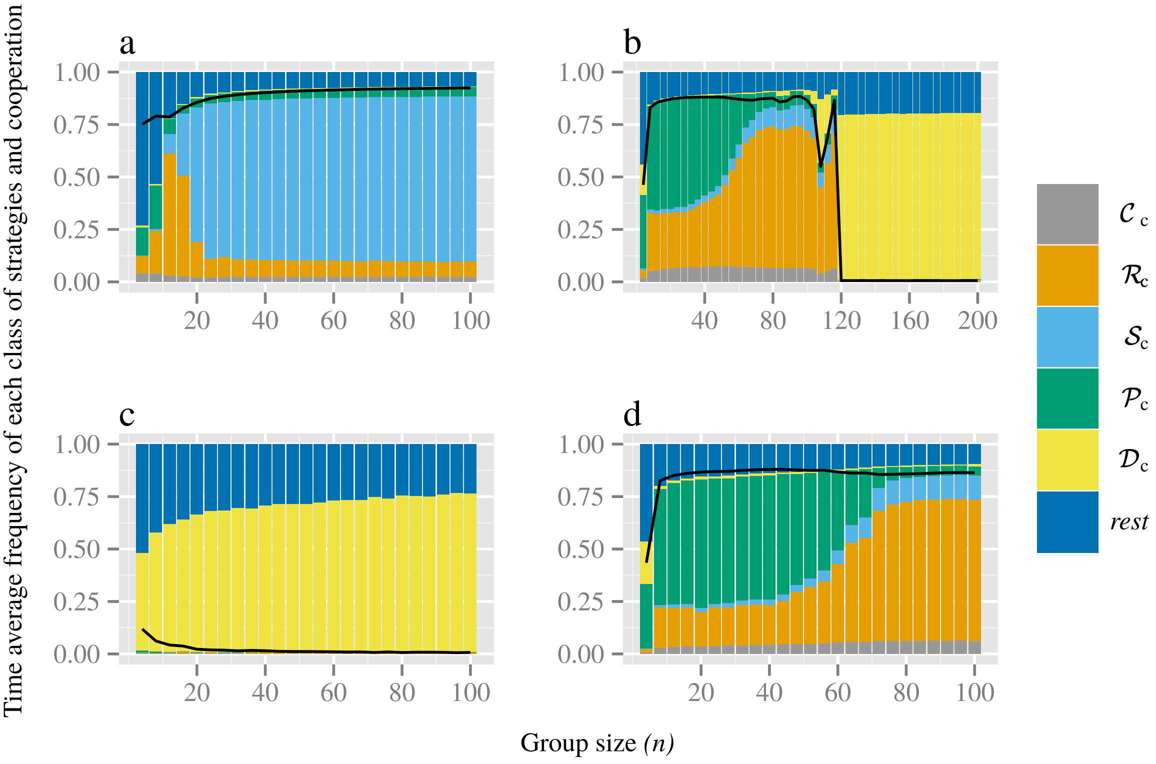


Figure S 5: Time average frequency (over 10^6 generations) of the frequency of cooperation (black line) and the six classes of strategies ($\mathcal{R}_c$, $\mathcal{S}_c$, $\mathcal{P}_c$, $\mathcal{C}_c$, $\mathcal{D}_c$, and "*rest*") plotted as a function of group size n. Parameter values: $B_\mathrm{h} = 2, C_\mathrm{h} = 1, \mu = 0.01, d = 250$. Panel specific parameters: $D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 0, T = 30, z = 3$ (panel a), $D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, T = 7, z = 3$ (panel b), $D_\mathrm{p} = 1, C_\mathrm{p} = 1, C_\mathrm{l} = 1, T = 4, z = T$ (panel c), $D_\mathrm{p} = 2, C_\mathrm{p} = 2, C_\mathrm{l} = 1, T = 7, z = T$ (panel d).

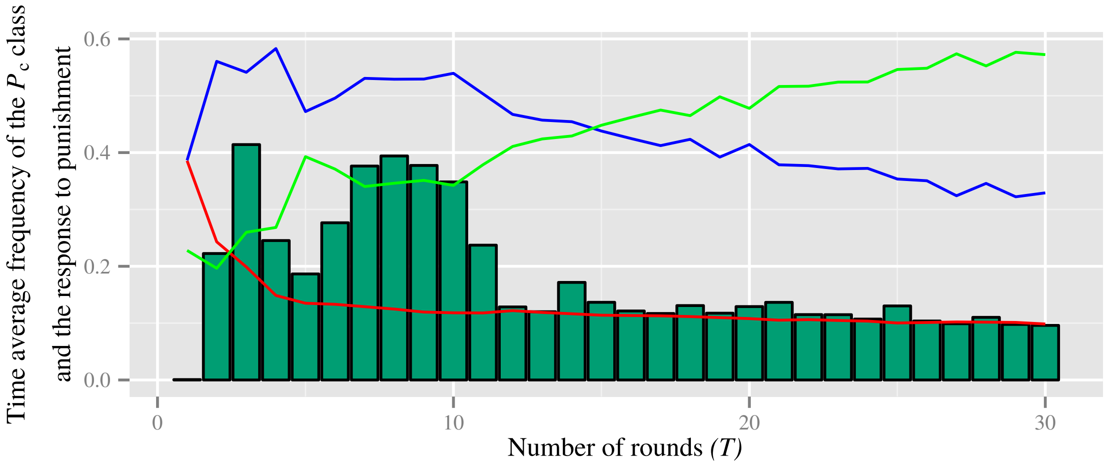


Figure S 6: Same as the baseline case (Fig. 2a, main text), but only showing the frequency of the $\mathcal{P}_c$ class (bars), together with three sets of strategies based on the response to punishment (x_4). The red line is the frequency of all strategies for which $x_4 = \mathrm{I}$, the blue line for which $x_4 = \mathrm{A}$, and the green line for which $x_4 = \mathrm{L}$.

Chapter two

Coevolution of positive reciprocity, punishment, partner switching, and unconditional helping in a kin structured population

Status: Preparing to submit to Scientific Reports

Wubs M, Lehmann L, Bshary R

Contributions: All authors contributed to the conceptual design and the writing of the manuscript. MW wrote the code and performed the analysis.

Abstract

Helping other individuals is often an investment: an act that reduces the helper's current payoff. Models trying to explain the evolution and maintenance of helping either investigate how population structure may promote unconditional helping via kin selection and/or interdependencies, or they investigate how partner control mechanisms like positive reciprocity, punishment, or partner switching may yield conditional helping in repeated interactions in panmictic populations. It is currently unclear how population structure would affect the various helping strategies, in particular if more complex conditional strategies incur a cognitive cost compared to simpler strategies, such as always help or always defect. Here, we address these questions using simulations. We first show that our approach reproduces key results from previous research, making our extension directly comparable to existing literature. We find that regular changes in group composition due to overlapping generations promotes the competitiveness of partner switching. Punishment is most strongly negatively affected by a cost for strategic complexity. Kin structure increases the range of conditions where conditional helping strategies are favoured over defection, while the conditions where unconditional helping is favoured are rather stringent. Overall, our study highlights how combining ecological and game theoretical modelling may yield unexpected and interesting outcomes.

Introduction

Nature is full of examples where an individual helps another, i.e., increases the direct fitness of the recipient (Carlisle and Zahavi, 1986; Russell and Hatchwell, 2001; Bergmüller et al., 2005). Helping often takes the form of an investment, defined as an act that reduces the helper's current payoff and increases the recipient's current payoff (Bshary and Bergmüller, 2008). Mutual helping based on investments is traditionally modelled with the Prisoner's Dilemma game (Rapoport and Chammah, 1965). In this game, two players can either help or defect, where the payoffs are such that mutual helping yields higher payoffs than mutual defection. The dilemma arises because defecting yields a higher payoff than helping in each single round, no matter how the partner behaves (Luce and Raiffa, 1957). Thus, the basic conclusion is that mutual defection is an evolutionarily stable strategy in the game, and the only one if the game consists of a single interaction.

Two rather separate approaches have been developed to explain the evolution and persistence of helping despite the apparent superiority of defection. The first approach (termed "ecological approach" by Bshary and Bergmüller, 2008) assumes that a population consists of unconditional helpers and unconditional defectors only, and then investigates in how far specific aspects of population structure may promote the persistence of helpers. Limited migration of offspring between demes of interacting individuals causes relatedness to build up within demes. Although the benefits of a helping act are then likely to be received by a related individual, these benefits may be cancelled out due to the increased competition between the offspring (Taylor, 1992a). Taylor and Irwin have shown that this balance is disrupted if generations overlap, favouring the evolution of helping behaviours (Taylor and Irwin, 2000). Various other assumptions such as the timing of life cycle events, migration, competition, and more have been identified that can change the selective gradient for helping behaviours (Lehmann and Rousset, 2010). Here, helping is altruistic in the Hamiltonean sense, i.e., it reduces the helper's lifetime direct fitness but is under positive selection due to indirect fitness benefits (Hamilton, 1964a,b). Alternatively, between-group competition may cause strong interdependency between individuals and hence helping may yield direct fitness benefits to both helper and recipient, which we term cooperation (Lehmann and Keller, 2006; Bshary and Bergmüller, 2008). The latter type of models is often called multi-level selection models and emphasise selective forces on the group level. However, it should be noted that in real life, group structuring is typically linked to genetic structuring, leading to a combination of direct and indirect benefits of helping (Okasha, 2006; Lehmann and Keller, 2006).

The second approach to explain helping based on mutual investments, termed "game theoretical approach" by Bshary & Bergmüller (Bshary and Bergmüller, 2008), emphasises the observation that many social interactions in nature are repeated and hence offer the potential for cooperative solutions based on conditional strategies (Kreps et al., 1982; Fudenberg and Maskin, 1986). Game theoretic models typically assume a genetically panmictic population and a probabilistic function concerning the likelihood that partners will interact again. Under such circumstances, helping can only evolve and persist if it yields direct fitness benefits. Direct benefits are achieved if cooperators can use conditional responses, called partner control mechanisms, to push the payoff of defectors below the population average, resulting in defectors being selected against (Bshary and Bronstein, 2011). For example, positive reciprocity is a partner control mechanism where investments are conditioned on the investments made by the partner, and the most well-known strategy using positive reciprocity is tit-for-tat (Axelrod and Hamilton, 1981). This strategy invariably starts with helping, but in subsequent rounds it expresses the action its partner has taken in the previous rounds, and thus a cooperative partner will be rewarded while a defecting partner will not (Axelrod and Hamilton, 1981; Linster, 1992). Other partner control mechanisms may also result in cooperators gaining more payoff than defectors. Through partner switching cooperators can avoid prolonged interactions with defectors and instead assort with other cooperators, leaving defectors end up interacting with each other. Although switching partner may initially be costly as it necessitates finding a new partner, this can be overcome if the individual finds a cooperator to interact with for sufficiently long to negate these costs (Aktipis, 2004; McNamara et al., 2008). A third important partner control mechanism is punishment, where the focal punishes the defecting partner thus decreasing the partner's payoff to such extent that the partner is better off cooperating (Clutton-Brock and Parker, 1995; Hilbe and Sigmund, 2010).

Both the effects of population structure and repeated interactions are frequently studied, but the two approaches are rarely properly combined. Game theoretic models often assume that the individuals are unrelated, or a fixed relatedness is assumed between interaction partners (e.g., Roberts, 2005), rather than letting a distribution of relatedness coefficients emerge as a function of specific life cycles. Conversely, ecological models often investigate scenarios in which individuals act unconditionally (Taylor, 1992a; Taylor and Irwin, 2000), excluding conditional strategies from the strategy space. Several models have investigated a combination of the two approaches, showing that the effects are not additive (Lehmann and Keller, 2006; Akçay and Cleve, 2012; Van Cleve and Akçay, 2014). Here, we attempt further integration of the two approaches. Integration warrants a simple terminology to avoid confusion about the relative importance of direct versus indirect fitness benefits. In the remainder of the manuscript, we will invariably use the term "helping" as the phenomenon we wish to explain, keeping in mind that helping can be either altruistic (yielding indirect fitness benefits) or cooperative (yielding direct fitness benefits) or a combination of both. Our starting point is our recent study on the relative competitiveness of positive reciprocity, punishment and partner switching in genetically unstructured populations (Wubs et al., 2016, chapter one). Using simulations, the main results of that study were that positive reciprocity may be favoured under rather restrictive conditions, as small deme sizes or few rounds of interactions favoured punishment while large deme sizes and many rounds of interactions favoured partner switching as partner control mechanisms (see also Izquierdo et al., 2010). Unconditional helping was never frequent in the simulations. Here we expand the previous study by introducing three factors that may potentially influence the relative competitiveness of helping strategies in important ways. The first two factors affect the genetic structure of the population and hence the scope for interactions between relatives: limited migration of offspring between demes, and overlapping generations due to selective survival of the parental generation. As a third factor, we introduced a cost that correlates with the complexity of a strategy, i.e., the amount of information that needs to be processed to make appropriate decisions. This assumption appears to be justified based on both theoretical research allowing for the evolution of the size of artificial intelligence networks (McNally et al., 2012) and empirical research involving selection on cognitive performance (Mery and Kawecki, 2003, 2004; Kotrschal et al., 2013). Such research shows that strategic sophistication comes at a cost that is only under positive selection under conditions in which sophistication yields overall net benefits. We analysed the effect of each variable first separately and then the combination of all three.

We were interested in three main questions. First, we asked in how far limited migration, overlapping generations, and complexity costs influence either alone or in combination the relative competitiveness of our conditionally helpful strategies, i.e., positive reciprocity, punishment and partner switching. In previous work we have characterized the conditions where each partner control mechanism is likely to be dominant in genetically unstructured populations (Wubs et al., 2016, chapter one). It remains unclear, however, how interactions among related individuals, overlapping generations, and complexity costs affect the relative competitiveness of each partner control mechanism. Second, we asked whether unconditional helping is ever favoured over conditional strategies and over defective strategies outside the single interaction case (in which conditional strategies offer no advantage), and if so which of our three factors would be particularly important. The coevolution of conditional and unconditional strategies is rarely studied, and it would thus be relevant to investigate what conditions may favour one type of strategies over the other. Third, we investigated how the level of helping in the population is affected by the interaction between conditional strategies and a kin structure. Recent work has demonstrated that the combination of repeated interactions and a kin structured population will increase the range of conditions where helping is favoured by selection compared to either repeated interactions or a kin structure in isolation (Lehmann and Keller, 2006; Akçay and Cleve, 2012; Van Cleve and Akçay, 2014), and it would thus be interesting to find how levels of helping in our model are affected by the interaction. We present iconic results in the main paper and more systematic variation of parameter values in the supporting material. It is important to keep in mind that several idiosyncratic key strategy parameters may vary greatly between species and/or individuals of the same species. For example, the cost of switching may vary in nature from "virtually absent" to "almost certain death", the cost of punishing may depend on a species' weaponry and/or an individual's relative strength, and the benefits of punishing may depend on the target's memory abilities. As each parameter may vary independently of the others, it is impossible to explore the full parameter space of all possible parameter state combinations. Thus, the main insights of our simulations are conclusions about how limited migration, overlapping generations, and complexity costs may affect the relative competiveness of competing strategies.

The model

Population

We consider a haploid population of constant size with a total number of $N = n \times d$ adult individuals, distributed on d demes of equal size n. The population advances in discrete time steps. Each time step $t = 1, 2, ..., T$ is marked by the following events. First, individuals within a deme interact socially with each other and accumulate payoffs. Next, each individual produces a large number of offspring proportionally to the payoff gained in the current time step. Each individual offspring either disperses with probability m or remains on its native deme with probability $1 - m$. Each dispersing offspring selects any deme other than its natal deme with equal probability $1/(d-1)$. After dispersal the next generation is formed as follows. Each of the N individuals of the parental generation survives with probability s and remains on the deme. Adults that have died are replaced by randomly selected offspring from within the deme (both native and immigrant) until deme size equals n, while the remaining offspring die.

This model is an extension of our previous work (Wubs et al., 2016, chapter one), with two changes in the life cycle: 1) limited migration of the offspring resulting in relatedness building up within demes (instead of random mixing), 2) generational overlap through survival of adult individuals over successive time steps such that individuals born at different time may coexist and reproduce at the same time (instead of the whole population dying each step).

Social interactions

During the social interaction phase the individuals play an extensive-form game that is repeated for N_{R} rounds before reproduction occurs. A single round of the game consists of five sequential moves. The first move is a so-called move by nature (a random move), followed by four moves played by the individuals, where the individuals interact in pairs and choose their actions simultaneously as follows.

Move 0: random pairing. Each unpaired individual is randomly paired "by nature" to another unpaired individual. Individuals cannot choose their interaction partner.

Move 1: Prisoner's Dilemma. Each individual chooses whether to help (action H) or to defect (action D). A helping individual pays a personal payoff cost C_{h}, while contributing a payoff benefit B_{h} to its partner. Choosing to defect does not affect the payoff of either individual.

Move 2: leaving. Each individual decides whether to leave its partner (action L) or to stay (action S). One individual leaving is sufficient for the pairbond to break. Both individuals pay a payoff cost C_l if the pairbond breaks and will not engage in the next two moves. Instead, they are added to a pool of unpaired individuals that will be re-matched in move 0 in the following round.

Move 3: punishment. Each individual decides whether to punish its partner (action P) or not (action N). A punishing individual pays a personal payoff cost C_p while reducing the payoff of the partner by D_p. Individuals that are not punished will skip the fourth move.

Move 4: response to punishment. A punished individual can respond to the punishment in three different ways. (1) The individual ignores the punishment (action I), meaning it will not change any future action whilst playing with the current partner unless specified by its own strategy. (2) The individual leaves the partner (action L as in move 2), breaking the pairbond and both individuals paying a cost C_l. (3) The individual alters its behaviour in the Prisoner's Dilemma (action A), meaning that if it played action C (resp. D) in move 1 of this round, it will switch to playing action D (resp. C) in move 1 for the next z rounds. If a punished individual is punished again, it will again change its action for z rounds, thus "overwriting" previous punishment. If the pairbond breaks, this effect of punishment disappears.

Strategies

We assume that strategies are pure and the actions are chosen deterministically. The strategy of an individual consists of a decision rule for each move, except move 0. For move 1 the rule specifies the action to take if the partner defected in the previous round, if the partner has helped and an initial action the first time it interacts with a new partner, yielding 8 ($= 2^3$) decision rules. For move 2 and 3 the decision rule specifies whether to leave or stay (and punish or not) if the partner defected and if the partner helped in move 1 of the current round, yielding four decision rules for each move. For move 4 the response to punishment is conditioned on being punished in the current round, yielding three decision rules (corresponding to each action that can be taken in move 4). To compose the strategy space we allow for any combination of decision rules for the separate moves. This generates a set of 384 ($= 8 \times 4 \times 4 \times 3$) strategies from which remove phenotypically indistinguishable strategies, i.e., strategies that cannot be distinguished by observing the actions only. The final set of strategies then contains 92 unique strategies, which are listed in the online supplementary material. See (Wubs et al., 2016, chapter one) or the online supplementary material for further details regarding the strategy space.

For the analysis of the results we pool the strategies into six distinct classes. Each strategy of the *positive reciprocity* class (denoted $\mathcal{R}_c$; we use the subscript $_c$ to indicate that it concerns a class of strategies) helps the first time it interacts with a new partner, while in subsequent rounds it copies the action of the partner, and it does not leave or punish the partner. Each strategy of the *partner switching* class (denoted $\mathcal{S}_c$) will always help in move 1, it does not punish, but it leaves as soon as the partner defects. Each strategy of the *punishment* class (denoted $\mathcal{P}_c$) will always help in move 1, it does not leave, but it punishes a partner that defects. Each strategy of the *unconditional helping* (denoted $\mathcal{C}_c$)and *unconditional defecting* class (denoted $\mathcal{D}_c$) always helps (defects if $\mathcal{D}_c$), and does not express any conditional play in move 1, 2, and 3. Each of these classes of strategies consists of three strategies that differ only in how they respond to being punished. The remaining $92 - 5 \times 3 = 77$ strategies are pooled in "*rest*".

Strategic complexity

Each round individuals suffer a reduction in payoff correlated to the complexity of the strategy used. We use a naive measure to compute the cost of a given strategy based on the assumption that conditional strategies are more costly than unconditional ones. The cost per strategy is determined as follows. We define a total strategic complexity score as the sum of the complexity scores of its decision rules for each move. The complexity score of the decision rules is determined as follows. For each move, the actions taken are either conditional on what the partner did, or fixed. Any conditional decision rule gets a complexity score of 1, an unconditional/fixed rule receives a score of 0. The sum of scores determines how complex a strategy is. Thus as an example, the strategy always help in move 1 (fixed behaviour), leave defectors only in move 2 (conditional), never punish in move 3 (fixed), and alter behaviour after receiving punishment (conditional) has a complexity score of 2 $(0 + 1 + 0 + 1)$. The total strategic complexity score is multiplied by a complexity scalar $C_\mathrm{s} \geq 0$ to compute the payoff cost an individual pays each round. Therefore, C_s determines the weight of the total strategic complexity score in relation to the other parameters of the model, and thus if $C_\mathrm{s} = 0$ the individual does not pay any cost, regardless of the complexity of its strategy.

Analyses

The model is analysed using individual-based simulation where we track the frequencies of the six classes of strategies in the population. The number of offspring an individual produces per time step is proportional to a baseline payoff (guaranteeing there can be no negative payoff) plus the effects of the social interaction during that time step and the effect of the complexity of its strategy. Offspring adopt with probability $1 - \mu$ the strategy of the parent, but with probability μ the offspring mutates to any other strategy with equal probability.

We define one generation as the number of time steps it takes on average for the entire population to be replaced. Thus if the survival rate (s) is 0.95, then one generation is equal to $1/(1 - 0.95) = 20$ time steps. We ran simulations for 10^6 generations and computed the average frequency of each class of strategies over an entire run, where each strategy was equally abundant in the population at the first generation. Finally, we computed the frequency of helping in a simulation, which is defined as the average frequency over the entire population and the repeated game where the individual helped its partner in the Prisoner's Dilemma game (move 1).

We note here that throughout the results we maintain the expected number of rounds ($= \frac{N_\mathrm{R}}{1-s}$) the individuals play at 30. In our previous work we found that the number of rounds can have a profound impact on the outcome. Focusing on 30 rounds is suitable for our current analyses because the three partner control mechanisms of interest coexisted to varying degrees in our previous analyses, depending on the deme size (Wubs et al., 2016, chapter one). We can hence investigate how adding migration, survival, and complexity costs affect the relative competitiveness of these control mechanisms, and how far the unconditional helping strategy may replace conditional ones.

Results

The results are structured as follows. We start by recovering two known results. 1) We reproduce similar results to our previous work, thus excluding the three new factors that are introduced in the current model: overlapping generations, limited migration between demes, and a complexity cost for strategies. This will be our baseline case. 2) We also recover analytical predictions made by Taylor and Irwin, who have shown when unconditional helpers are favoured over defectors in a kin structured population (Taylor and Irwin, 2000). We then proceed by introducing each new factor separately to our baseline model before we include all factors in a single model. For each analysis we delineate the conditions where each helping class of strategies ($\mathcal{C}_c$, $\mathcal{R}_c$, $\mathcal{P}_c$, and $\mathcal{S}_c$) is favoured by selection. We do this for both a low benefit to cost ratio (2 to 1) and a high one (10 to 1) to ensure that our model fits in the Taylor and Irwin framework. For the high benefit to cost ratio case we present results with all factors included in a single model here, while referring to the supplementary material for the results where each factor is introduced in isolation, as these were found to be mostly similar to the low benefit to cost ratio case. Finally, we investigate how the level of helping is affected due to the interaction of a kin structured population and conditional strategies.

Recovering baseline results

The panmictic case

In figure 1 we have reproduced one of our main previous findings (Fig. 3b in Wubs et al., 2016, chapter one), which shows the $\mathcal{P}_c$ class as the dominant class of helping strategies for deme size up to 40 for the following parameters: $B_\mathrm{h} = 2, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, \mu = 0.01, d = 250, z = 5$. Furthermore, to match the previous model, $N_\mathrm{R} = 30$ and $s = 0$ (thus generations do not overlap), the migration rate is set to 0.996 ($= 1 - 1/d$) such that there is random mixing of the offspring generation (no genetic structure), and $C_\mathrm{s} = 0$ such that there is no cost of complexity for the strategies. Note hat for $n = 2$ (which was not tested in Wubs et al., 2016, chapter one) the rest class dominates instead, which consists mostly of strategies that initiate the interaction with defection, punish other defectors, but switch to helping when being punished. From this baseline case we introduce the various new factors in the section "New results".

The structured case

We now reproduce the analytical predictions made by Taylor and Irwin showing when helpers are favoured over defectors in pairwise interactions in an infinite island model (equations A9 and A10 Taylor and Irwin, 2000; note that there is no cost for dispersal in our model). With $m = 0.2, s = 0.96667, B_\mathrm{h} = 10, C_\mathrm{h} = 1$, helpers are favoured by selection for $n \leq 6$ (we use $s = 0.96667$ to set the expected number of interactions to 30 in order to match the conditions from our previous work). In figure 2 our results are plotted using those parameter values together with only one always help and one always defect strategy in the model. We find that the helpers are relatively favoured over defectors up until deme size 6, thus verifying that our model fits within the Taylor and Irwin framework.

New results

Migration

Here we introduce a limited migration rate to our baseline case in figure 1. In figure 3 we show the results of simulations with the same parameter values as in figure 1, except for $m = 0.5$ in panel a and $m = 0.2$ in panel b. We find that a limited rate of migration for the offspring has little effect on the relative proportion of the different classes of strategies, which is in line with various findings that the benefit of interacting with related individuals may often be cancelled out by the increased competition among relatives (Taylor, 1992a).

Overlapping generations

Introducing overlapping generations to our baseline model negatively impacts the frequency of the $\mathcal{P}_c$ class. In figure 4a we set $N_\mathrm{R} = 10, s = 0.66667$ (thus keeping the expected number of rounds an individual plays at 30) with otherwise identical parameter values as in figure 1, while in panel b we set $N_\mathrm{R} = 1, s = 0.96667$. We find that especially the $\mathcal{S}_c$ class is relatively more favoured by selection when generations strongly overlap, as the $\mathcal{S}_c$ class dominates for $n \geq 10$ (Fig. 4b). The $\mathcal{S}_c$ class is relatively more favoured here because of the increased variation in the deme composition due to the overlap in generations: when generations do not overlap a partner switcher in a deme full of defectors will switch each round of the game, but end up interacting with a defector each time. If generations overlap, then at each time step there is a non-zero probability of new individuals arriving in the deme (from any deme, including the focal one). Due to the extra variation introduced the $\mathcal{S}_c$ class is more likely to get paired with another helping strategy eventually and is thus favoured by selection over the $\mathcal{P}_c$ class.

Strategic complexity

We now introduce a cost for strategic complexity to our baseline case (Fig. 1). In figure 5a the results are shown for $C_\mathrm{s} = 0.1$ with otherwise identical parameter values as in figure 1. We find that the $\mathcal{P}_c$ class has almost entirely disappeared from the population, being replaced by the $\mathcal{D}_c$ class in small demes ($n \leq 10$), the $\mathcal{R}_c$ class in intermediate sized demes ($12 \leq n \leq 32$), and $\mathcal{S}_c$ class in large demes ($n \geq 34$). This strong reduction in frequency for the $\mathcal{P}_c$ class is due to the fact that the recipient switching to helping after being punished will have a reduction in payoff (as it pays the cost for conditionally responding to punishment). Thus, defectors that opportunistically switch to help punisher are more strongly selected against if C_s increases and as a consequence the competitiveness of the $\mathcal{P}_c$ class decreases, while the $\mathcal{S}_c$ and $\mathcal{R}_c$ classes do not necessarily require their partner to use conditional strategies. For $C_\mathrm{s} = 0.2$ (panel b) we find that the $\mathcal{D}_c$ class is favoured for a larger range of deme sizes ($n \leq 26$) as the strategies in this class are less costly than conditional strategies. Therefore for higher values of C_s the $\mathcal{D}_c$ class is more likely to be favoured by selection, since in the absence of relatedness the $\mathcal{C}_c$ class will similarly be selected against.

Full model - low benefit/cost ratio

We now present the results of the full model, thus combining a limited migration rate, a cost for strategic complexity, and overlapping generations. We use the following parameter values: $B_\mathrm{h} = 2, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, N_\mathrm{R} = 1, \mu = 0.01, d = 250, z = 5, s = 0.96667, m = 0.5, C_\mathrm{s} = 0.1$. We find that through the combination of limited migration, overlapping generations and a cost for strategic complexity the $\mathcal{P}_c$ class is very much selected against compared to our baseline case (Fig. 6a). The $\mathcal{S}_c$ class dominates for $n \geq 12$, which is a considerably smaller deme size compared to our previous model ($n \geq 76$, Fig. 3b in Wubs et al., 2016, chapter one). However, in smaller demes the $\mathcal{S}_c$ class is still outcompeted by the $\mathcal{R}_c$ class ($4 \leq n \leq 12$), which shows that in small demes a switching strategy is unlikely to dominate. Finally, only for $n = 2$ the $\mathcal{D}_c$ and rest class are instead dominant in the population, where also the main strategy in the rest class always defects in move 1, and thus the level of helping in the population here is low.

A higher cost of complexity ($C_\mathrm{s} = 0.2$, panel b) results in the $\mathcal{D}_c$ dominating for $n \leq 20$ and the $\mathcal{S}_c$ for larger deme sizes, while if the cost of switching is increased as well ($C_\mathrm{s} = 0.2, C_\mathrm{l} = 5$, panel c) the $\mathcal{D}_c$ class is dominant for any deme size. This shows that as the costs for conditional strategies increases, the unconditional $\mathcal{D}_c$ class will end up dominating since it is also favoured over the $\mathcal{C}_c$ class for this benefit to cost ratio (Taylor and Irwin, 2000).

We also find, however, that other classes than the $\mathcal{C}_c$ class may benefit from interacting with more closely related individuals. In panel d the migration rate is reduced to 0.2 while otherwise using the same parameter values as in panel c. As a result, the $\mathcal{R}_c$ class now dominates for $n \leq 8$. This together shows that depending on the exact combination of parameter values, various classes of strategies may dominate.

Full model - high benefit/cost ratio

In order to identify conditions where the $\mathcal{C}_c$ class is favoured over conditional helping classes, we performed simulations with a high benefit from a helping act (B_h), which is needed for unconditional helpers to outcompete unconditional defectors (Taylor and Irwin, 2000). For the high benefit to cost ratio case we first show that, also when including all 92 strategies in the model, we can still recover the analytical predictions by Taylor and Irwin, and then note a few subtle differences between the low and high benefit to cost ratio case. In figure 7a the results of the following parameter values are shown: $B_\mathrm{h} = 10, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 10, C_\mathrm{l} = 10, N_\mathrm{R} = 1, \mu = 0.01, d = 250, z = 5, s = 0.96667, m = 0.2, C_\mathrm{s} = 10$, thus with a high benefit for helping, but large costs for strategic complexity, punishing, and switching partners, together leading to strong selection against any strategy other than the simplest always help and always defect strategy. As such, we find that the $\mathcal{C}_c$ class (which contains the simplest strategy that always helps) dominates for $n \leq 6$ and the $\mathcal{D}_c$ class (which contains the simplest strategy that always defects) beyond that point, which shows that the model is consistent with the results of Taylor and Irwin (Fig. 2).

Then, in order to compare the high and low benefit to cost ratio cases, in panels b and c (of Fig. 7) the parameter values are as in figure 6a and b, respectively, except for $B_\mathrm{h} = 10$. This shows, first of all, that a higher benefit from a helpful act results in more helping in the population as the $\mathcal{D}_c$ class is virtually absent (Fig. 7c) while it dominated for $n \leq 20$ for $B_\mathrm{h} = 2$ (Fig. 6b). Furthermore we now find that for $C_\mathrm{s} = 0.2$ the $\mathcal{C}_c$ class increases in frequency. However, the $\mathcal{C}_c$ class is also here never found to be the dominant class of helping strategies, and thus a high benefit to cost ratio does not result in unconditional helping being relatively favoured over conditional strategies. Therefore, the conditions where unconditional helping is favoured over conditional helping appear to be rather stringent (e.g., with unlikely high C_s).

Helping

We now turn to the level of helping itself. We analyse the level of helping in three main cases: in a kin structured population with only the always help and always defect strategies, in a genetically unstructured population with the full strategy set, and in a kin structured population with the full strategy set. For a migration rate of 0.996 ($= 1 - 1/d$) the population is genetically unstructured, and we thus expect and find that in the absence of conditional strategies the population almost entirely exists of defectors (Fig. 8, $m = 0.996$, grey bars). Reducing the migration rate results in relatedness within demes building up, and we consequently find higher levels of helping, where especially for $n = 2$ (where relatedness is relatively high) high levels of helping are achieved. Futhermore, we find that if individuals use conditional strategies, helping can evolve in a genetically unstructured population (Fig. 8, $m = 0.996$, orange bars). However, the highest levels of helping are achieved through a combination of conditional strategies and a kin structured population (orange bars, $m = 0.5$ and $m = 0.2$), which is in line with predictions from previous work (Lehmann and Keller, 2006; Akçay and Cleve, 2012; Van Cleve and Akçay, 2014).

Interestingly, we do not necessarily find that higher levels of relatedness (as achieved through a reduction in migration rate) increases the overall level of helping. Obviously the level of helping cannot increase much further, but the slight decrease in the level of helping was found for a variety of parameter combinations. On closer inspection, we find that the $\mathcal{R}_c$ class is relatively more favoured than the $\mathcal{P}_c$ class when the migration rate is decreased. Thus, while the $\mathcal{P}_c$ class always helps their partner, the $\mathcal{R}_c$ class switches to play defect when faced with a defecting partner. Cooperation between strategies from the $\mathcal{R}_c$ class and other strategies is thus slightly less likely to occur, and as such the level of helping slightly decreases when the $\mathcal{R}_c$ class replaces the $\mathcal{P}_c$ class.

Testing the robustness of results

In the supplementary material we provide additional simulations that test the robustness of the various results presented here. First, we show that a high benefit to cost ratio does not affect the relative competitiveness of helping strategies, though levels of helping are increased (Figs. S1 to S4). Furthermore, we show that fewer numbers of interactions (by varying survival) do not favour unconditional helping (Fig. S5). Finally, we show that even if the effect of being punished (z) lasts indefinitely, the $\mathcal{P}_c$ class is still largely outcompeted by other strategies for $C_\mathrm{s} = 0.1, B_\mathrm{h} = 2$ (Fig. S6).

Discussion

Our model and results build upon previous work where we have investigated the coevolution of three partner control mechanisms (positive reciprocity, punishment, and partner switching) in a genetically unstructured population. In the current work we added three new factors: 1) limited migration of the offspring individuals; 2) overlapping generations; 3) cost for strategic complexity. The first two will ensure that relatedness within a deme will build up, while the latter acknowledges that unconditional strategies are cognitively less demanding than conditional ones. We will first discuss the effect of each factor on the relative competitiveness of the different partner control mechanism, followed by a section on unconditional helping versus conditional helping, and finally discuss the effects of relatedness and conditionality on the level of helping.

Partner control mechanisms

With a migration rate below $1-(1/d)$ relatedness will build up within demes. However, our life cycle assumptions are such that individuals first interact socially, followed by reproduction, mortality of the parental generation (complete mortality in our baseline case), and then migration. Therefore, with non-overlapping generations (Fig. 3) the benefits of interacting with related individuals are exactly cancelled out by the increased competition for resources (i.e., the breeding spots on a deme) among relatives (Taylor, 1992a,b; West, 2002). In line with this logic, levels of helping and the relative competitiveness of our strategies were hardly affected by migration rates alone (Fig. 3). In principle, it could have been possible that the changes in migration rate may affect relative competitiveness of strategies by changing variation in deme composition: if offspring randomly mix over demes then the composition of a deme will reflect the composition of the entire population, while if migration is limited then on average the majority of individuals will be sampled from the focal deme resulting in less variation within demes. Such reduced variation within a deme should reduce the likelihood that a switching individual finds another helping partner and hence reduce the strategy's relative competitiveness. Apparently, this effect was not strong enough to alter the outcome in our baseline case.

The main effect of having generations overlap in our model is that it increases the competitiveness of the switching strategy (Fig. 4). Variation is a key factor for the evolution of helping (McNamara et al., 2004; McNamara and Leimar, 2010), and our results highlight that the probability of finding a helping partner can depend on generational overlap rather than on standing variation in a population, as typically assumed. The extent to which generations overlap influences the number of potential partners a focal individual can have during its lifetime: while it is fixed at $n-1$ when there is no overlap, the number will be higher for overlapping generations. For example, for demes with $n = 40, s = 0.96667$, on average 1.33 individuals die each time step in a deme, and thus a focal individual living for 30 time steps will share the deme with 78 $(= n - 1 + 29 * 1.33)$ individuals during its lifetime. More importantly, any death will lead to two new potential partners available for pairing up for the next round: the "widow" and the replacement. Due to the higher number of potential partners a switching strategy benefits greatly from generational overlap (Fig. 4).

In the economic literature the fact that more complex strategies may be more costly and that this will influence the outcome of a model has long been recognised (Piccione and Rubinstein, 1993; Abreu and Rubinstein, 1988; Binmore and Samuelson, 1992), but in biological models it is often neglected. Here we have applied a complexity cost by penalizing conditionality, which was found to increase the relative competitiveness of the $\mathcal{R}_c$ and $\mathcal{S}_c$ classes compared to the $\mathcal{P}_c$ class as the $\mathcal{P}_c$ class relies on a conditional response in its partner (Fig. 5). Rather similarly, in our previous model we found that positive reciprocity was often outcompeted by alternative partner control mechanisms due to the absence of defectors that opportunistically switch to helping when paired with a positive reciprocity player (Wubs et al., 2016, chapter one). Thus the evolutionary success of both positive reciprocity and punishment (also called negative reciprocity) rely on the presence of "smart" defectors.

This result exposes an important issue with punishment in general. That is, the recipient of a punishing act has to learn to associate the punishment as a result of its own behaviour, and, additionally, alter its behaviour accordingly. It may be (relatively) easy for a focal individual to associate receiving punishment each time it performs a specific action (say, stealing food) and thus learn to stop performing that action. However, if the focal is punished for defecting (which technically is an action in our model, but in nature would mean doing nothing), the punished individual is less likely to learn to express the "right" behaviour (Raihani et al., 2012). As such, the evolution of punishment as a partner control mechanism to stabilize cooperation may be strongly impaired due to cognitive constraints, which in turn may exist due to the costs of cognitive strategies. Our model verifies that, due to this extra hurdle that needs to be taken for punishment to be effective, the likelihood of punishment evolving as the dominant partner control mechanism is severely decreased.

Unconditional helping

In the full model, due to combining a limited migration rate with overlapping generations, unconditional helping can be favoured by selection over defection. Although unconditional helpers are often part of the strategy set in models testing partner control mechanisms (García and Traulsen, 2012; Izquierdo et al., 2010), rarely is it investigated when they would outcompete conditional strategies. This can occur when conditional strategies are being penalised for their conditionality and the relatedness within demes is high enough to favour the $\mathcal{C}_c$ class over the $\mathcal{D}_c$ class. This is exemplified in figure 7a, where the parameters are set to select against any other strategy than the most simple always help and always defect strategy and thus we recover the analytical result from Taylor and Irwin here (Taylor and Irwin, 2000). However, we would argue that such extreme conditions are extremely rare in natural populations. Without extreme costs we find for both a low and a high benefit to cost ratio conditional helping strategies are relatively favoured over unconditional ones (Figs. 6 and 7). One could argue that other mechanisms, such as kin recognition, may still allow helping to be stable against defection without the need for partner control mechanisms (Antal et al., 2009). However, recognizing kin may be cognitively costly as well or could be combined with a partner control mechanism, and thus further modelling is required to determine how such a mechanism will affect the competitiveness of unconditional helping. Additionally, in species where a kin recognition mechanism has been identified, the kin are not necessarily treated preferentially (Mateo, 2002). Finally, one may expect that unconditional helping could gain a relative advantage over conditional helping when the number of rounds of interaction during a lifetime is reduced, since conditional strategies depend on multiple rounds of interaction. However, in the online supplementary material we show that when varying the survival, such that individual go from 1 to 10 number of rounds of interaction, the conditional strategies are still favoured over unconditional helping (Fig. S5).

Helping

Theoretical research on social evolution can be roughly divided into two main approaches. First, the evolution of altruistic traits due to assortment of genetically related individuals, and second, the evolution and maintenance of helping due to the evolution of conditional strategies. In their seminal paper, Axelrod and Hamilton already realised that clustering of individuals would increase the range of conditions under which the tit for tat strategy can invade the always defect strategy (Axelrod and Hamilton, 1981). Despite this recognition, much of the literature focuses on a single approach. Although this may ease the analysis, it is now increasingly being recognized that for a proper understanding of the evolution of a social trait the two approaches need to be combined. In several recent papers the fitness effects of interacting with both behaviourally responsive unrelated individuals (e.g., conditional strategies) and genetically related individuals are combined into a single equation (Lehmann and Keller, 2006; Akçay and Cleve, 2012; Van Cleve and Akçay, 2014). From these equations it is clear that the effects of relatedness and responsiveness cannot be subsumed into a single parameter, and thus both must be considered when studying the evolution of a social trait (Akçay and Cleve, 2012; Van Cleve and Akçay, 2014).

We have tested several analytical predictions in our model. First, we tested whether our model can reproduce the predictions made by Taylor and Irwin regarding the evolution of helping in a kin structured population (Taylor and Irwin, 2000). There are two differences between their analytical model and our simulations: we use a fixed number of demes and allow mutants to enter the population at every time step. Despite this, we find that helping evolves under the same conditions as predicted by their equations (Fig. 2).

Second, we also show here that when conditional strategies evolve in a kin structured population, the level of helping can be significantly higher compared to unconditional strategies in a kin structured population or conditional strategies in an unstructured population (Fig. 8). However, our results highlight the fact that the presence of conditional (behaviourally responsive) strategies may affect the level of helping in counter intuitive ways. Increasing relatedness, which increases the level of helping in the absence of conditional strategies, can decrease the level of helping under certain conditions (Fig. 8). This is due to helping strategies differing in the likelihood that they induce a partner into helping. Although the level of helping is only marginally affected, one should be aware that the interaction between relatedness and behavioural responsive strategies may yield unexpected outcomes.

Conclusion

In conclusion, we have made an attempt here to combine an ecological modelling approach with a game-theoretic one. By letting conditional strategies evolve in a kin structured population we studied the interaction of the two. Although strategies may evolve in our model both because of indirect effects (helping related individuals) or direct effects (using a partner control mechanism), we cannot easily partition direct and indirect effects on the fitness of an individual in the model. We were, however, able to study the isolated effect of several factors (migration rate, survival rate, complexity cost; Figs. 3, 4, and 5). Furthermore, while unconditional helping is unlikely to be favoured by selection over conditional helping strategies (Figs. 6 and 7), we showed that a kin-structure increased the range of conditions where conditional strategies are favoured over defection (Fig. 6c, d and supplementary material).

One of the most striking results is that both positive reciprocity and partner switching are much more likely to dominate in the population compared to our previous work where punishment was found to be the dominant partner control mechanism for a large part of the parameter space (Wubs et al., 2016, chapter one). However, we argue that the three factors introduced here add more realism to the model, as they apply to many species where also direct benefits are documented, such as many primate species (Schino and Aureli, 2010). It is therefore not surprising that most work on primates document either positive reciprocity or partner choice as the mechanism by which helping is stabilized (Schino and Aureli, 2016), the same mechanisms that are found to be dominant in our model. The current model therefore better represents what is seen in nature.

References

Abreu, D. and Rubinstein, A. The Structure of Nash Equilibrium in Repeated Games with Finite Automata. *Econometrica*, 56(6):1259–1281, 1988.

Akçay, E. and Cleve, J. V. Behavioral Responses in Structured Populations Pave the Way to Group Optimality*. *The American Naturalist*, 179(2):257–269, 2012.

Aktipis, C. A. Know when to walk away: contingent movement and the evolution of cooperation. *Journal of Theoretical Biology*, 231(2):249–260, 2004.

Antal, T., Ohtsuki, H., Wakeley, J., Taylor, P. D., and Nowak, M. A. Evolution of cooperation by phenotypic similarity. *Proceedings of the National Academy of Sciences*, 106(21):8597–8600, 2009.

Axelrod, R. and Hamilton, W. D. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

Bergmüller, R., Heg, D., and Taborsky, M. Helpers in a cooperatively breeding cichlid stay and pay or disperse and breed, depending on ecological constraints. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1560):325–331, 2005.

Binmore, K. G. and Samuelson, L. Evolutionary stability in repeated games played by finite automata. *Journal of Economic Theory*, 57(2):278–305, 1992.

Bshary, R. and Bergmüller, R. Distinguishing four fundamental approaches to the evolution of helping. *Journal of Evolutionary Biology*, 21(2):405–420, 2008.

Bshary, R. and Bronstein, J. L. A General Scheme to Predict Partner Control Mechanisms in Pairwise Cooperative Interactions Between Unrelated Individuals. *Ethology*, 117(4):271–283, 2011.

Carlisle, T. and Zahavi, A. Helping at the nest, allofeeding and social status in immature arabian babblers. 18(5):339–351, 1986.

Clutton-Brock, T. H. and Parker, G. A. Punishment in animal societies. *Nature*, 373(6511): 209–216, 1995.

Fudenberg, D. and Maskin, E. The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54(3):533–554, 1986.

García, J. and Traulsen, A. Leaving the loners alone: Evolution of cooperation in the presence of antisocial punishment. *Journal of Theoretical Biology*, 307:168–173, 2012.

Hamilton, W. D. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, 7 (1):1–16, 1964a.

Hamilton, W. D. The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, 7 (1):17–52, 1964b.

Hilbe, C. and Sigmund, K. Incentives and opportunism: from the carrot to the stick. *Proceedings of the Royal Society B: Biological Sciences*, 277(1693):2427–2433, 2010.

Izquierdo, S. S., Izquierdo, L. R., and Vega-Redondo, F. The option to leave: Conditional dissociation in the evolution of cooperation. *Journal of Theoretical Biology*, 267(1):76–84, 2010.

Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., Immler, S., Maklakov, A. A., and Kolm, N. Artificial Selection on Relative Brain Size in the Guppy Reveals Costs and Benefits of Evolving a Larger Brain. *Curr Biol*, 23(2):168–171, 2013.

Kreps, D. M., Milgrom, P., Roberts, J., and Wilson, R. Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27(2):245–252, 1982.

Lehmann, L. and Keller, L. The evolution of cooperation and altruism a general framework and a classification of models. *Journal of Evolutionary Biology*, 19(5):1365–1376, 2006.

Lehmann, L. and Rousset, F. How life history and demography promote or inhibit the evolution of helping behaviours. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553):2599–2617, 2010.

Linster, B. G. Evolutionary stability in the infinitely repeated prisoners' dilemma played by two-state Moore machines. *Southern Economic Journal*, 58:880–903, 1992.

Luce, R. D. and Raiffa, H. *Games and Decisions.* Whiley, New York, NY, USA, 1957.

Mateo, J. M. Kin-recognition abilities and nepotism as a function of sociality. *Proceedings of the Royal Society of London B: Biological Sciences*, 269(1492):721–727, 2002.

McNally, L., Brown, S. P., and Jackson, A. L. Cooperation and the evolution of intelligence. *Proceedings of the Royal Society B: Biological Sciences*, 279(1740):3027–3034, 2012.

McNamara, J. M. and Leimar, O. Variation and the response to variation as a basis for successful cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553): 2627–2633, 2010.

McNamara, J. M., Barta, Z., and Houston, A. I. Variation in behaviour promotes cooperation in the Prisoner's Dilemma game. *Nature*, 428(6984):745–748, 2004.

McNamara, J. M., Barta, Z., Fromhage, L., and Houston, A. I. The coevolution of choosiness and cooperation. *Nature*, 451(7175):189–192, 2008.

Mery, F. and Kawecki, T. J. An operating cost of learning in Drosophila melanogaster. *Animal Behaviour*, 68(3):589–598, 2004.

Mery, F. and Kawecki, T. J. A fitness cost of learning ability in Drosophila melanogaster. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1532):2465–2469, 2003.

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

Piccione, M. and Rubinstein, A. Finite Automata Play a Repeated Extensive Game. *Journal of Economic Theory*, 61(1):160–168, 1993.

Raihani, N. J., Thornton, A., and Bshary, R. Punishment and cooperation in nature. *Trends in Ecology & Evolution*, 27(5):288–295, 2012.

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

Roberts, G. Cooperation through interdependence. *Animal Behaviour*, 70(4):901–908, 2005.

Russell, A. F. and Hatchwell, B. J. Experimental evidence for kin-biased helping in a cooperatively breeding vertebrate. *Proceedings of the Royal Society of London B: Biological Sciences*, 268 (1481):2169–2174, 2001.

Schino, G. and Aureli, F. The relative roles of kinship and reciprocity in explaining primate altruism. *Ecology Letters*, 13(1):45–50, 2010.

Schino, G. and Aureli, F. Reciprocity in group-living animals: partner control versus partner choice. *Biol Rev*, page n/a, 2016.

Taylor, P. D. Altruism in viscous populations an inclusive fitness model. *Evolutionary Ecology*, 6 (4):352–356, 1992a.

Taylor, P. D. and Irwin, A. J. Overlapping generations can promote altruistic behavior. *Evolution*, 54(4):1135–1141, 2000.

Taylor, P. D. Inclusive Fitness in a Homogeneous Environment. *Proceedings of the Royal Society of London B: Biological Sciences*, 249(1326):299–302, 1992b.

Van Cleve, J. and Akçay, E. Pathways to social evolution: Reciprocity, relatedness, and synergy. *Evolution*, 68(8):2245–2258, 2014.

West, S. A. Cooperation and Competition Between Relatives. *Science*, 296(5565):72–75, 2002.

Wubs, M., Bshary, R., and Lehmann, L. Coevolution between positive reciprocity, punishment, and partner switching in repeated interactions. *Proc. R. Soc. B*, 283(1832):20160488+, 2016.

Table and figures

Table 1: Meaning of used symbols.

Symbol	**Meaning**
N	Population size
n	Deme size
d	Number of demes
T	Time steps
m	Migration rate
s	Survival rate
N_{R}	Number of rounds
C_{h}	Payoff cost when helping
B_{h}	Payoff benefit for the recipient of a helping act
C_{l}	Payoff cost when the pairbond breaks
C_{p}	Cost of punishing
D_{p}	Payoff reduction when being punished
z	Duration of behavioural change after being punished
C_{s}	Scalar for cost of strategic complexity
μ	Mutation rate
$\mathcal{C}_c$	Unconditionally helping class
$\mathcal{R}_c$	Positive reciprocity class
$\mathcal{P}_c$	Punishment class
$\mathcal{S}_c$	Switching class
$\mathcal{D}_c$	Unconditionally defect class
"rest"	Remaining strategies

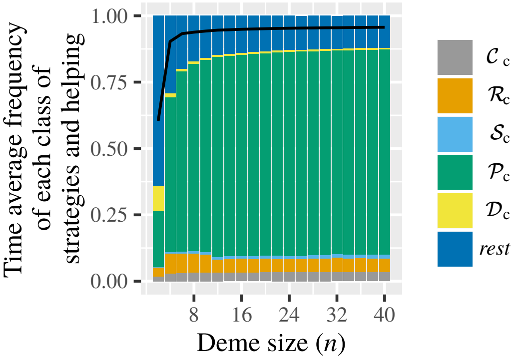


Figure 1: Time average frequency (over 10^6 generations) of the frequency of helping (black line) and the six classes of strategies ($\mathcal{C}_c, \mathcal{R}_c, \mathcal{S}_c, \mathcal{P}_c, \mathcal{D}_c$, and *rest*) plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 2, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, N_\mathrm{R} = 30, z = 5, \mu = 0.01, d = 250, m = 0.996, s = 0, C_\mathrm{s} = 0$.

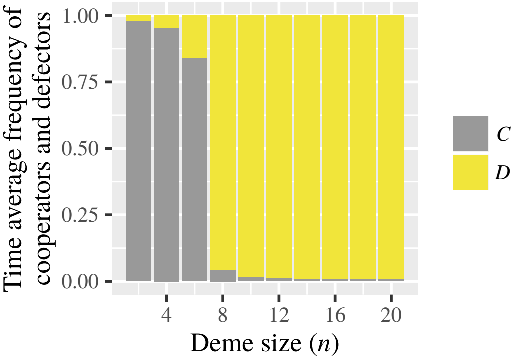


Figure 2: Time average frequency (over 10^6 generations) of the frequency of cooperators and defectors plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 10, C_\mathrm{h} = 1, \mu = 0.001, d = 250, m = 0.2, s = 0.96667$. Irrelevant parameters: $D_\mathrm{p}, C_\mathrm{p}, C_\mathrm{l}, N_\mathrm{R}, z, C_\mathrm{s}$.

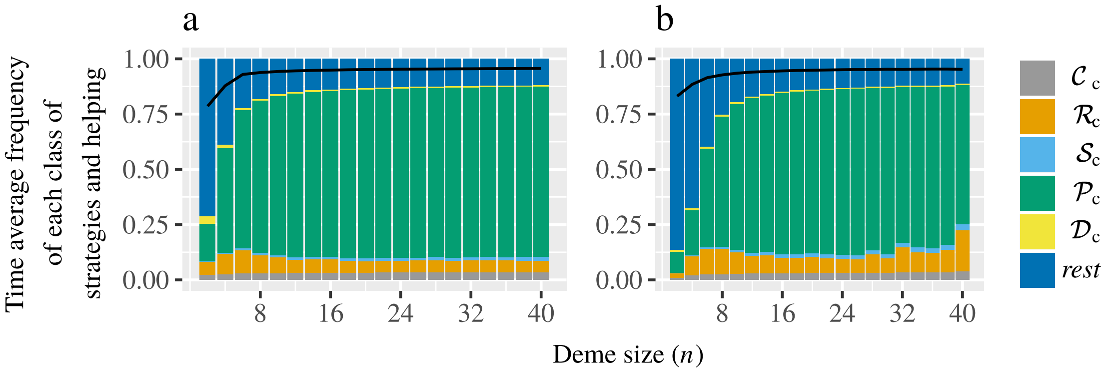


Figure 3: Same as in figure 1, except the migration rate $m = 0.5$ in panel a and $m = 0.2$ in panel b.

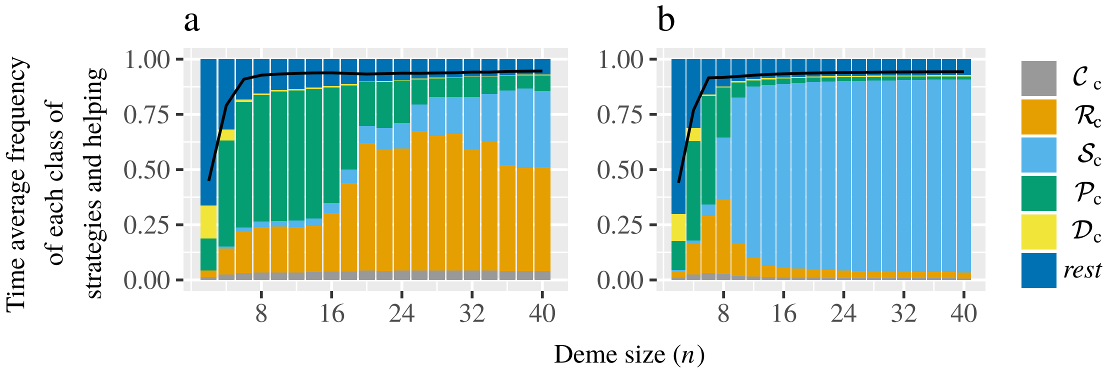


Figure 4: Same as in figure 1, except the number of rounds $N_\mathrm{R} = 10$ and the survival rate $s = 0.66667$ in panel a and $N_\mathrm{R} = 1, s = 0.96667$ in panel b.

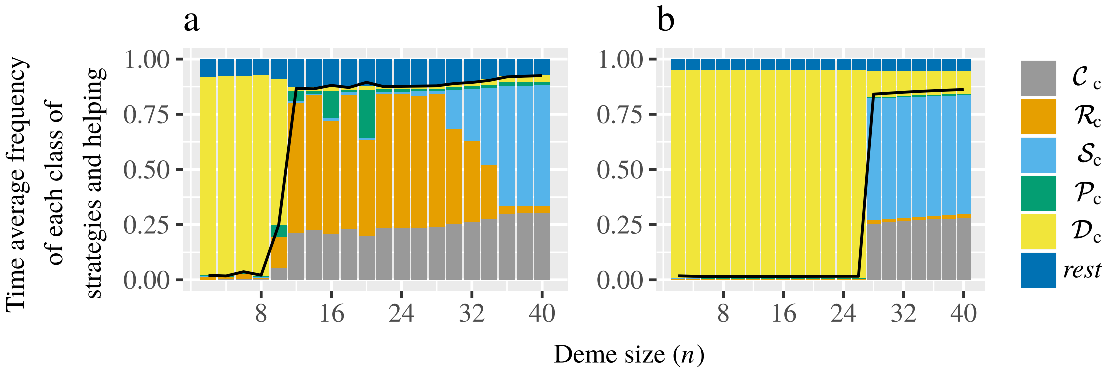


Figure 5: Same as in figure 1, except complexity scalar $C_\mathrm{s} = 0.1$ in panel a and $C_\mathrm{s} = 0.2$ in panel b.

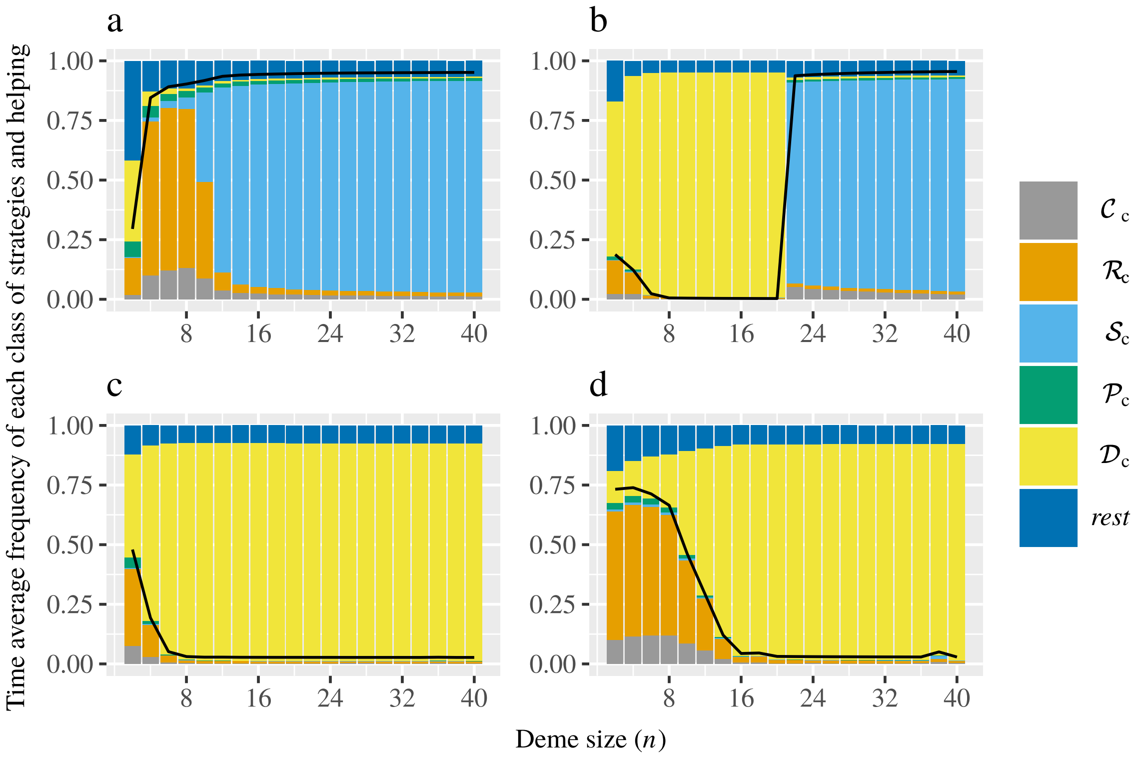


Figure 6: Time average frequency (over 10^6 generations) of the frequency of helping (black line) and the six classes of strategies ($\mathcal{C}_c, \mathcal{R}_c, \mathcal{S}_c, \mathcal{P}_c, \mathcal{D}_c$, and *rest*) plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 2, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, N_\mathrm{R} = 1, z = 5, \mu = 0.01, d = 250, s = 0.96667$. Panel specific parameters: panel a $C_\mathrm{s} = 0.1, m = 0.5, C_\mathrm{l} = 1$; panel b $C_\mathrm{s} = 0.2, m = 0.5, C_\mathrm{l} = 1$; panel c $C_\mathrm{s} = 0.2, m = 0.5, C_\mathrm{l} = 5$; panel d $C_\mathrm{s} = 0.2, m = 0.2, C_\mathrm{l} = 5$.

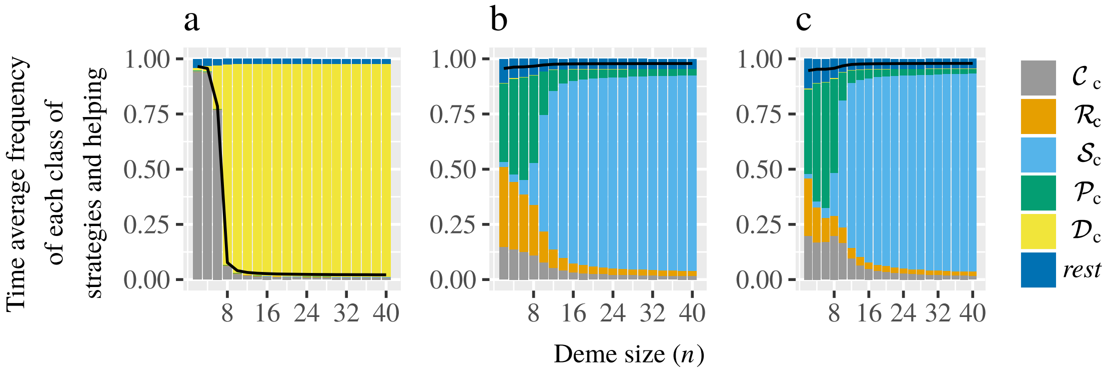


Figure 7: Time average frequency (over 10^6 generations) of the frequency of helping (black line) and the six classes of strategies ($\mathcal{C}_c, \mathcal{R}_c, \mathcal{S}_c, \mathcal{P}_c, \mathcal{D}_c$, and *rest*) plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 10, C_\mathrm{h} = 1, D_\mathrm{p} = 2, N_\mathrm{R} = 1, z = 5, \mu = 0.01, d = 250, s = 0.96667$. Panel specific parameters: panel a $m = 0.2, C_\mathrm{s} = 10, C_\mathrm{l} = 10, C_\mathrm{p} = 10$; panel b $m = 0.5, C_\mathrm{s} = 0.1, C_\mathrm{l} = 1, C_\mathrm{p} = 1$; panel c $m = 0.5, C_\mathrm{s} = 0.2, C_\mathrm{l} = 1, C_\mathrm{p} = 1$.

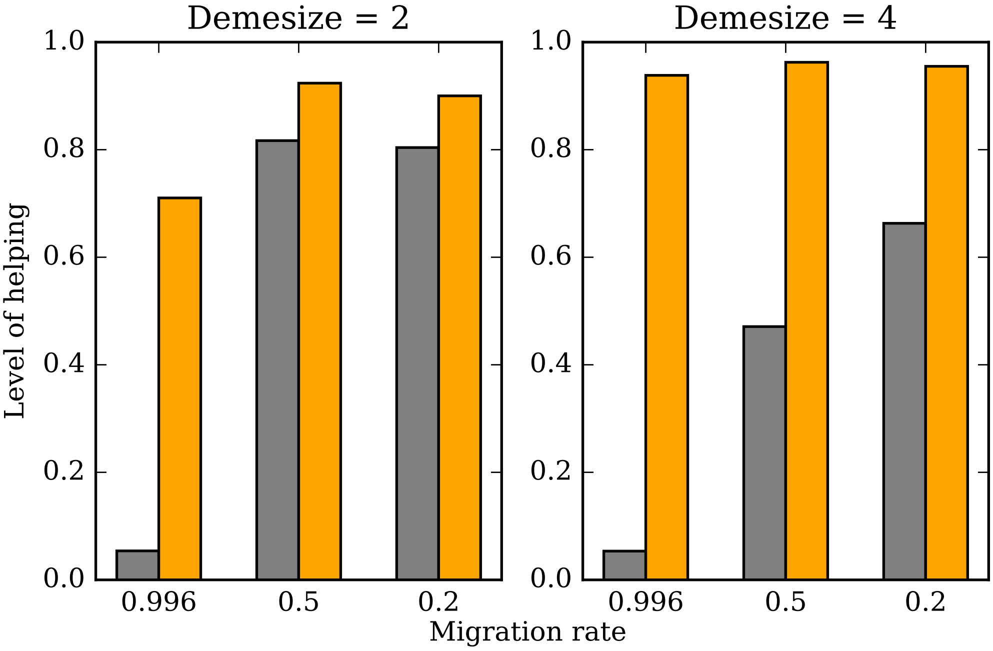


Figure 8: Level of helping for different migration rates. Parameter values: $B_{\mathrm{h}} = 10, C_{\mathrm{h}} = 1, D_{\mathrm{p}} = 2, C_{\mathrm{p}} = 1, C_{\mathrm{l}} = 1, N_{\mathrm{R}} = 1, z = 5, \mu = 0.01, d = 250, s = 0.96667, C_{\mathrm{s}} = 0$. Grey columns show results for simulations using only one always help and one always defect strategy, orange columns include the full strategy set.

Supplementary material

Introducing migration rate, overlapping generations, and complexity costs with a high benefit to cost ratio

We start by showing how each factor (migration rate, overlapping generations, complexity cost) affects the results when introduced in the high benefit to cost ratio case. For this, we first show the distribution of classes of strategies in our baseline case with a high benefit, $B_\mathrm{h} = 10$. The results are plotted in figure 1. From these results it is clear that changing the benefit of a helping act (B_h) has little impact on the frequency of the various classes of strategies, since the $\mathcal{P}_c$ class is also here the dominant helping class of strategies for any deme size.

Migration rate

Introducing a limited migration rate to our baseline case with $B_\mathrm{h} = 10$ (figure 1) distorts the frequencies of the various classes more than compared to the baseline with $B_\mathrm{h} = 2$ as in the manuscript. The results, plotted in figure 2, show that a limited migration rate increases the frequency of the $\mathcal{R}_c$ class for small deme sizes. However, it should be noted that only for $n = 8$ the $\mathcal{R}_c$ class is actually dominant, while for all other deme sizes the $\mathcal{P}_c$ class is.

Overlapping generations

When introducing generational overlap to the baseline case we find again that the $\mathcal{S}_c$ class is more strongly favoured by selection (figure 3). Compared to the results shown in the manuscript we do not find the $\mathcal{R}_c$ class with intermediate generational overlap (for $N_\mathrm{R} = 10, s = 0.66667$) since the $\mathcal{S}_c$ class will benefit more from a higher B_h as soon as a helping partner has been found.

Complexity cost

Regarding the cost of complexity we find that the transition from the $\mathcal{P}_c$ class to the other helping classes to the $\mathcal{D}_c$ class occurs for higher values of C_s (figure 4). With $C_\mathrm{s} = 0.2$ the $\mathcal{S}_c$ and $\mathcal{C}_c$ classes dominate for large n, while the $\mathcal{P}_c$ class is still dominant up to $n = 34$. Undoubtedly the $\mathcal{D}_c$ class will eventually end up dominating with large enough C_s since the $\mathcal{C}_c$ class will be outcompeted in a genetically unstructured population.

Fewer interactions

In this section we test whether fewer number of interactions favours unconditional helping over conditional helping as conditional strategies are obviously dependent on repeated interactions, and higher number of interactions may relatively increase the payoff gained by conditional strategies. To test this we vary the survival from 0 to 0.9 in figure 5. With $s = 0$ individuals interact only once before reproducing and dying. As such the population consists of various strategies that defect on the first interaction, since subsequent interaction are irrelevant. If survival is increased to $s = 0.3$ the $\mathcal{D}_c$ class replaces other defecting strategies: e.g. strategies that would defect on the first interaction, but then start to help unconditionally. We see here that for $n = 2$ a small proportion of the population consists of the $\mathcal{R}_c$ class already, which has a much larger frequency than the $\mathcal{C}_c$ class. Increasing survival even further, to $s = 0.6$ and $s = 0.9$, will recover almost similar results as are found in the main text (figure 7b; although $C_\mathrm{s} = 0.1$ there). As such, for $C_\mathrm{s} = 0$ we are unable to find conditions where the $\mathcal{C}_c$ class outcompetes other helping classes of strategies.

Additional parameter variation for full model

Here we show results of various conditions where the $\mathcal{P}_c$ class may be relatively more favoured compared to what is found in the main text. In the first panel in figure 6 we set $C_\mathrm{s} = 0.1$ but with a high cost of switching ($C_\mathrm{l} = 5$) and the behavioural change after receiving punishment lasts indefinitely ($z = 0$), where both changes should relatively favour the $\mathcal{P}_c$ class. However, even under such favourable conditions we find that for small demes the $\mathcal{P}_c$ class is largely outcompeted by the $\mathcal{R}_c$ class, while in larger dems it is outcompeted by the $\mathcal{S}_c$ class. As such, with $C_\mathrm{s} = 0.1$ the $\mathcal{P}_c$ class is very unlikely to dominate.

Setting $C_\mathrm{s} = 0$ while keeping $C_\mathrm{l} = 5, z = 0$, we do find that the $\mathcal{P}_c$ class is dominant for any deme size (up until at least 40), which again verifies how strongly the class is affected by the parameter C_s. In the third and fourth panel we set $z = 5$ and $C_\mathrm{l} = 1$, respectively. This shows that in the absence of a cognitive cost we can find various conditions where the $\mathcal{P}_c$ class will still outcompete alternative conditional helping strategies.

Strategy set

In table S1 the decision rules for each move of the stage game are listed, while table S2 lists the main classes of strategies. In table S3 the full set of strategies is given. The coding of the decision rules is as follows.

We assume that the strategy for move 1 specifies an action taken when the individual first interacts with its partner, and an action taken in subsequent rounds is conditioned on what the partner did in the previous round in move 1. This decision rule can thus be written as $a_1 a_\mathrm{C} a_\mathrm{D}$, where $a_1, a_\mathrm{C}, a_\mathrm{D} \in \{\mathrm{C}, \mathrm{D}\}$. Here, a_1 is the action taken the first time the two individuals in a pair interact, a_C is the action taken if the partner cooperated in the previous round, and a_D is the action taken if the partner defected in the previous round. We thus have a total of 8 (2^3) decision rules for move 1: {CCC, CCD, CDC, CDD, DCC, DCD, DDC, DDD}.

For move 2, the decision to leave or stay is assumed to be conditional on the action taken by the partner in move 1 of the current round. Hence, the decision rule can be written as $a_\mathrm{C} a_\mathrm{D}$, where $a_\mathrm{C} \in \{\mathrm{L}, \mathrm{S}\}$ ($a_\mathrm{D} \in \{\mathrm{L}, \mathrm{S}\}$) gives the action take when the partner cooperated (defected). The set of decision rules for move 2 is thus {LL, LS, SL, SS}.

Likewise, for move 3, the decision to punish or not to punish the partner is assumed to be conditional on the action taken by the partner in move 1, so that the decision rule is $a_\mathrm{C} a_\mathrm{D}$, where $a_\mathrm{C} \in \{\mathrm{P}, \mathrm{N}\}$ ($a_\mathrm{D} \in \{\mathrm{P}, \mathrm{N}\}$) is the action taken when the partner cooperated (defected). The set of decision rules for move 3 is thus {PP, PN, NP, NN}.

Finally, the response to punishment in move 4 is simply given by $\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$.

An X in a decision rule in table S3 is a placeholder for an action that is not played at any point during the game (C or D in move 1; P or N in move 3). Two strategies that are otherwise similar, but differing in this action would therefore never act differently (i.e. they are phenotypically indistinguishable). Only one of these two strategies is included in the strategy space. This can occur in two situations. First, if an individual leaves after the partner cooperates (or defects), then it cannot also punish or conditionally cooperate/defect in the following round since the pair is broken up, and thus an X is shown in place of the action in both the decision rule for move 1 and 3. Second, we assumed that a punishing act could not be followed by a conditional action in move 1 of the following round, and thus the action P in the decision rule for move 3 is always combined with an X in the decision rule for move 1.

Table S 1: The sets of decision rules for each move of the game.

Move 1:	$\{\mathrm{CCC}, \mathrm{CCD}, \mathrm{CDC}, \mathrm{CDD}, \mathrm{DCC}, \mathrm{DCD}, \mathrm{DDC}, \mathrm{DDD}\}$
Move 2:	$\{\mathrm{LL}, \mathrm{SS}, \mathrm{LS}, \mathrm{SL}\}$
Move 3:	$\{\mathrm{PP}, \mathrm{NN}, \mathrm{PN}, \mathrm{NP}\}$
Move 4:	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$

Table S 2: Main classes of strategies.

Name	Move 1	Move 2	Move 3	Move 4
Always cooperate $\mathcal{C}_c$	CCC	SS	NN	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$
Positive reciprocity $\mathcal{R}_c$	CCD	SS	NN	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$
Partner switching $\mathcal{S}_c$	CCC	SL	NN	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$
Punishment $\mathcal{P}_c$	CCC	SS	NP	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$
Always defect $\mathcal{D}_c$	DDD	SS	NN	$\{\mathrm{I}, \mathrm{A}, \mathrm{L}\}$

Table S 3: Full set of strategies used in all simulations.

Move 1	**Move 2**	**Move 3**	**Move 4**	**Complexity**	**Move 1**	**Move 2**	**Move 3**	**Move 4**	**Complexity**
CCC	SS	NN	I	0	DCC	SS	NN	I	1
CCC	SS	NN	A	1	DCC	SS	NN	A	2
CCC	SS	NN	L	1	DCC	SS	NN	L	2
CCD	SS	NN	I	1	DCD	SS	NN	I	1
CCD	SS	NN	A	2	DCD	SS	NN	A	2
CCD	SS	NN	L	2	DCD	SS	NN	L	2
CCX	SL	NX	I	1	DCX	SL	NX	I	2
CCX	SL	NX	A	2	DCX	SL	NX	A	3
CCX	SL	NX	L	2	DCX	SL	NX	L	3
CCX	SS	NP	I	1	DCX	SS	NP	I	2
CCX	SS	NP	A	2	DCX	SS	NP	A	3
CCX	SS	NP	L	2	DCX	SS	NP	L	3
CDC	SS	NN	I	1	DDC	SS	NN	I	1
CDC	SS	NN	A	2	DDC	SS	NN	A	2
CDC	SS	NN	L	2	DDC	SS	NN	L	2
CDD	SS	NN	I	1	DDD	SS	NN	I	0
CDD	SS	NN	A	2	DDD	SS	NN	A	1
CDD	SS	NN	L	2	DDD	SS	NN	L	1
CDX	SL	NX	I	2	DDX	SL	NX	I	1
CDX	SL	NX	A	3	DDX	SL	NX	A	2
CDX	SL	NX	L	3	DDX	SL	NX	L	2
CDX	SS	NP	I	2	DDX	SS	NP	I	1
CDX	SS	NP	A	3	DDX	SS	NP	A	2
CDX	SS	NP	L	3	DDX	SS	NP	L	2
CXC	LS	XN	I	1	DXC	LS	XN	I	2
CXC	LS	XN	A	2	DXC	LS	XN	A	3
CXC	LS	XN	L	2	DXC	LS	XN	L	3
CXD	LS	XN	I	2	DXD	LS	XN	I	1
CXD	LS	XN	A	3	DXD	LS	XN	A	2
CXD	LS	XN	L	3	DXD	LS	XN	L	2
CXX	LL	XX	I	0	DXX	LL	XX	I	0
CXX	LS	XP	I	2	DXX	LS	XP	I	2
CXX	LS	XP	A	3	DXX	LS	XP	A	3
CXX	LS	XP	L	3	DXX	LS	XP	L	3
CXC	SS	PN	I	1	DXC	SS	PN	I	2
CXC	SS	PN	A	2	DXC	SS	PN	A	3
CXC	SS	PN	L	2	DXC	SS	PN	L	3
CXD	SS	PN	I	2	DXD	SS	PN	I	1
CXD	SS	PN	A	3	DXD	SS	PN	A	2
CXD	SS	PN	L	3	DXD	SS	PN	L	2
CXX	SL	PX	I	2	DXX	SL	PX	I	2
CXX	SL	PX	A	3	DXX	SL	PX	A	3
CXX	SL	PX	L	3	DXX	SL	PX	L	3
CXX	SS	PP	I	0	DXX	SS	PP	I	0
CXX	SS	PP	A	1	DXX	SS	PP	A	1
CXX	SS	PP	L	1	DXX	SS	PP	L	1

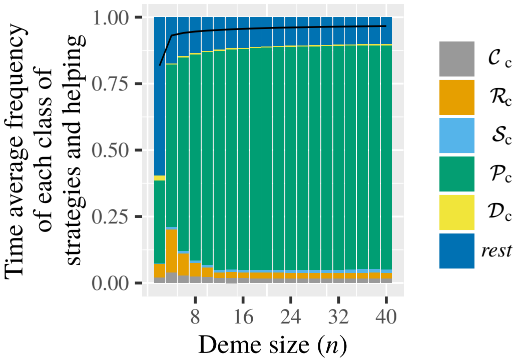


Figure S 1: Time average frequency (over 10^6 generations) of the frequency of helping (black line) and the six classes of strategies ($\mathcal{C}_c, \mathcal{R}_c, \mathcal{S}_c, \mathcal{P}_c, \mathcal{D}_c$, and *rest*) plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 10, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, N_\mathrm{R} = 30, z = 5, \mu = 0.01, d = 250, m = 0.996, s = 0, C_\mathrm{s} = 0$.

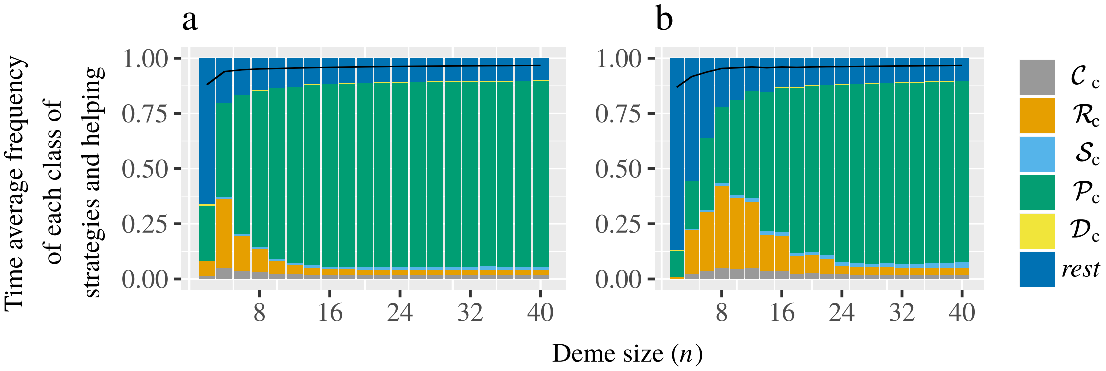


Figure S 2: Time average frequency (over 10^6 generations) of the frequency of helping (black line) and the six classes of strategies ($\mathcal{C}_c, \mathcal{R}_c, \mathcal{S}_c, \mathcal{P}_c, \mathcal{D}_c$, and *rest*) plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 10, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, N_\mathrm{R} = 30, z = 5, \mu = 0.01, d = 250, s = 0, C_\mathrm{s} = 0$. Panel a: $m = 05$, panel b: $m = 0.2$.

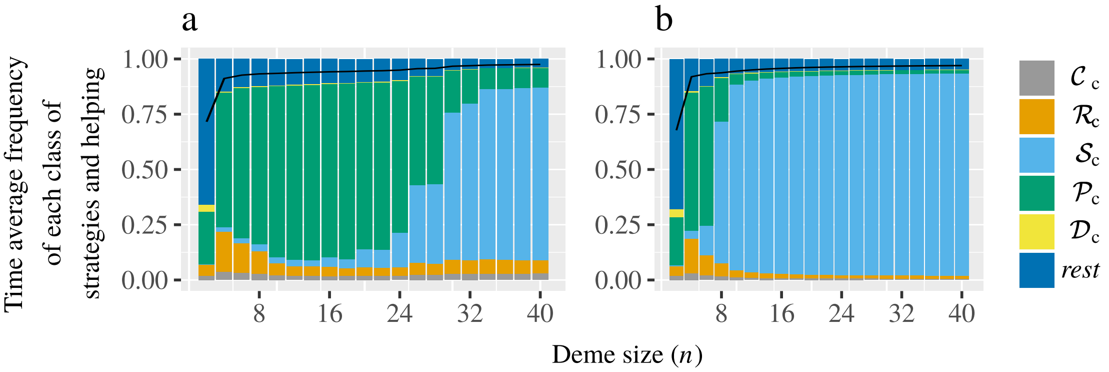


Figure S 3: Time average frequency (over 10^6 generations) of the frequency of helping (black line) and the six classes of strategies ($\mathcal{C}_c, \mathcal{R}_c, \mathcal{S}_c, \mathcal{P}_c, \mathcal{D}_c$, and *rest*) plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 10, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, z = 5, \mu = 0.01, d = 250, m = 0.996, C_\mathrm{s} = 0$. Panel a: $N_\mathrm{R} = 10, s = 0.66667$, panel b: $N_\mathrm{R} = 1, s = 0.96667$.

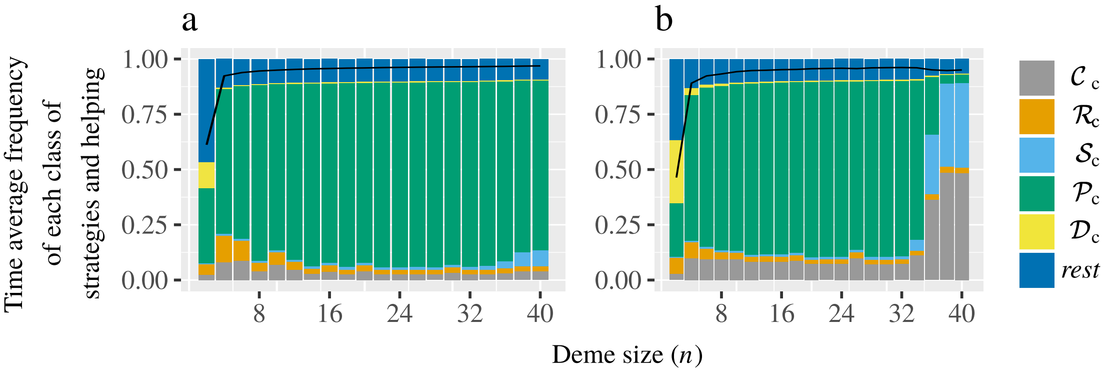


Figure S 4: Time average frequency (over 10^6 generations) of the frequency of helping (black line) and the six classes of strategies ($\mathcal{C}_c, \mathcal{R}_c, \mathcal{S}_c, \mathcal{P}_c, \mathcal{D}_c$, and *rest*) plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 10, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, N_\mathrm{R} = 30, z = 5, \mu = 0.01, d = 250, m = 0.996, s = 0$. Panel a: $C_\mathrm{s} = 0.1$, panel b: $C_\mathrm{s} = 0.2$.

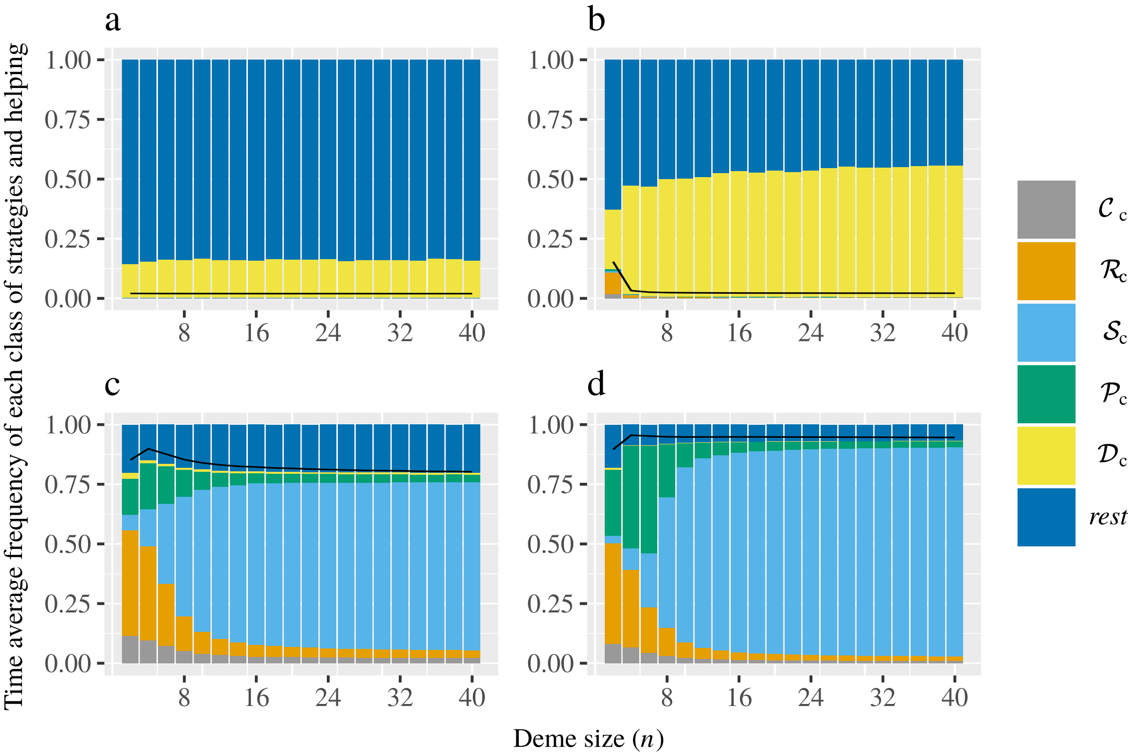


Figure S 5: Time average frequency (over 10^6 generations) of the frequency of helping (black line) and the six classes of strategies ($\mathcal{C}_c, \mathcal{R}_c, \mathcal{S}_c, \mathcal{P}_c, \mathcal{D}_c$, and *rest*) plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 10, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, C_\mathrm{l} = 1, N_\mathrm{R} = 1, z = 5, \mu = 0.01, d = 250, m = 0.5, C_\mathrm{s} = 0$. Panel a: $s = 0$, panel b: $s = 0.3$, panel c: $s = 0.6$, panel d: $s = 0.9$.

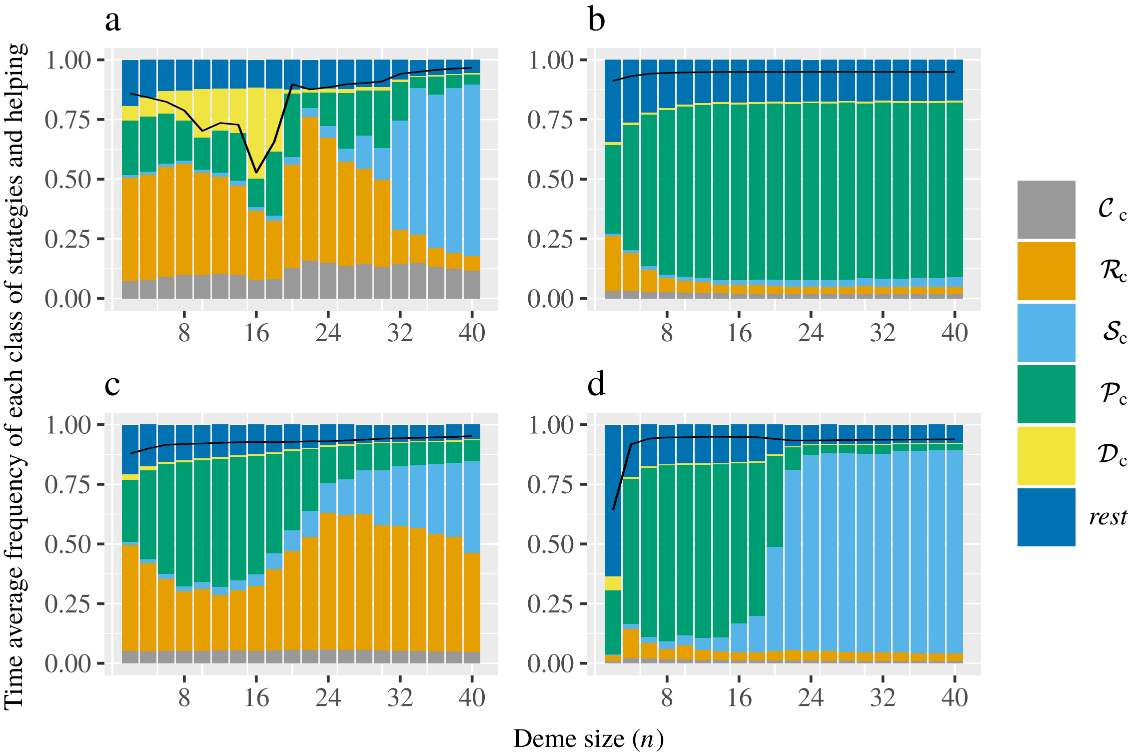


Figure S 6: Time average frequency (over 10^6 generations) of the frequency of helping (black line) and the six classes of strategies ($\mathcal{C}_c, \mathcal{R}_c, \mathcal{S}_c, \mathcal{P}_c, \mathcal{D}_c$, and *rest*) plotted a function of deme size n. Parameter values: $B_\mathrm{h} = 2, C_\mathrm{h} = 1, D_\mathrm{p} = 2, C_\mathrm{p} = 1, N_\mathrm{R} = 1, \mu = 0.01, d = 250, m = 0.996, s = 0.96667$. Panel a: $C_\mathrm{l} = 5, z = 5, C_\mathrm{s} = 0.1$, panel b: $C_\mathrm{l} = 5, z = 0, C_\mathrm{s} = 0$, panel c: $C_\mathrm{l} = 5, z = 5, C_\mathrm{s} = 00$, panel d: $C_\mathrm{l} = 1, z = 0, C_\mathrm{s} = 0$.

Chapter three
A model on grooming up the hierarchy in primates

Status: Preparing to submit to Animal Behaviour
Wubs M, Bshary R, Lehmann L

Contributions: All authors contributed to the conceptual design and the writing of the manuscript. MW wrote the code and performed the analysis.

Abstract

Primates spend a considerable amount of time grooming each other. Grooming is regularly found to be traded reciprocally (for grooming) or for rank-related benefits in the presence of food competition. It has been suggested that if food sources are clustered and monopolizable, then lower ranked individuals will groom higher ranked ones in order to be tolerated on food patches. This leads to grooming being directed up the hierarchy. However, the conditions where grooming up the hierarchy is expected to occur are based on verbal reasoning alone, and no quantitative analysis of the conditions favouring grooming up the hierarchy appear in the literature. Here, we develop a model to investigate if food competition can result in grooming up the hierarchy. Individuals are assumed to take actions pertaining to whom to groom, where to feed, and whom to tolerate on food patches. By allowing individuals to choose actions according to reinforcement learning, we delineate conditions where groups of individuals will express reciprocal grooming and grooming up the hierarchy depending on environmental conditions (e.g., quality, number of food patches). In particular, we show that conditions of intense food competition may lead to less grooming up the hierarchy. We suggest that quantifying the intensity of food competition and the aggressiveness of individuals on a food source will lead to a better understanding of observed grooming patterns in natural populations.

Introduction

Primates spend a significant proportion of their daily time budget grooming other individuals: picking through fur to remove ectoparasites and clean the skin (social grooming, hereafter grooming) (Dunbar, 1991). This observation comes maybe at no surprise since being groomed is beneficial due to the removal of ectoparasites (Tanaka and Takefushi, 1993). Furthermore, grooming has an important social function, where social bonds between individuals are reinforced through the act of grooming. It has been argued that grooming behaviour has evolved because of these benefits (Dunbar, 1991). However, there are also costs associated with giving grooming. These come in the form of opportunity costs (not being able to use the time or energy for other activities), such as reduced vigilance (Maestripieri, 1993; Barrett and Henzi, 2006). Thus, grooming behaviour can potentially be exploited, where in a pair of individuals only one makes the investment to groom and the other reaps the benefit without reciprocating. Understanding the causal proximate and ultimate factors influencing individual grooming decisions is a long standing goal in primatology (Seyfarth, 1980; Schino, 2001).

Grooming can be thought of as a commodity that can be traded, where individuals give grooming in order to receive something back. If grooming is solely traded for grooming, then pairs of individuals are expected to trade equal amounts of grooming. Numerous studies do indeed report a so-called time-matching of grooming between individuals (Rowell et al., 1991; Barrett et al., 1999; Leinfelder et al., 2001; Pazol and Cords, 2005). Here, the individuals in a dyad are found to spend an approximately equal amount of time grooming each other, either within a single grooming bout or over a longer period of time. However, primate troops are nearly always characterized by a dominance hierarchy, and grooming interactions occur between individuals of different ranks. In a significant proportion of studies, grooming is observed to be directed up the hierarchy, i.e., higher ranked individuals receive more grooming than lower ranked ones. Hence, grooming decisions are likely to depend here on other factors than just the exchange of grooming. In other words, grooming of high ranked individuals by low ranked individuals is likely to be traded for other benefits than grooming itself (Seyfarth, 1980; Ventura et al., 2006; Wei et al., 2012; Roubová et al., 2015).

What are the benefits that can explain grooming up the hierarchy? The socioecological model of primatology is concerned with the causes and consequences of food competition and its effect on social relationships (Kappeler and van Schaik, 2002). Within this framework it is argued that under conditions where individuals compete over food resources low ranked individuals may trade grooming for non-grooming benefits provided by high ranked ones (van Schaik, 1989; Sterck et al., 1997; Kappeler and van Schaik, 2002). If food sources (or patches) are clumped and monopolizable, then individuals are expected to defend food sources, or attempt to displace one another from these. Here, coalitions may be formed between individuals, either to defend a food source or to displace others from it. Then, low ranking individuals may groom higher ranking individuals in order to gain coalitionary support in agonistic interactions (Seyfarth, 1980; Seyfarth and Cheney, 1984; Schino, 2007; Carne et al., 2011). Experimental evidence for grooming being traded for coalitionary support can be found in baboons and vervet monkeys (Cheney et al., 2010; Borgeaud and Bshary, 2015). Due to the trading of grooming by low ranked individuals for such support from high ranked individuals, on average grooming should be directed up the hierarchy.

Grooming can also be traded for increased tolerance (reduced aggression) for a low ranked individual by a high ranked one on a food source (Henzi and Barrett, 1999). This relies on broadly the same ecological conditions as the grooming under the coalitionary support hypothesis. If food sources are monopolizable then, instead of attempting to displace a higher ranked individual, the low ranked individual may groom the higher ranked one in order to be tolerated by that individual on the food source. Like the coalitionary support hypothesis, there is empirical evidence showing that grooming for tolerance occurs (Ventura et al., 2006; Carne et al., 2011; Tiddi et al., 2011; Xia et al., 2012). Furthermore, while most primate studies rely on correlational data, it has been shown experimentally that vervet monkeys trade grooming for short term tolerance (Borgeaud and Bshary, 2015).

There exists ample empirical support that grooming may be directed up the hierarchy, either when it is traded for coalitionary support or for tolerance, but the conditions under which this is expected to occur are mainly based on verbal models (Wrangham, 1980; van Schaik, 1989; Sterck et al., 1997). Therefore, there are few quantitative predictions as to the conditions under which grooming will be directed up the hierarchy and what grooming patterns will be associated with it (which individuals receive extra grooming) (Dunbar, 2002). One of the few models investigating grooming up the hierarchy is the one by Seyfarth (Seyfarth, 1977). In this quantitative model individuals make grooming decisions based on the "attractiveness" of potential recipients. The attractiveness of an individual as grooming partner depends on the rank of that individual, since rank influences the likelihood of successfully supporting the focal individual in an agonistic interaction, and thus the highest ranked individual is the most attractive grooming partner. The model managed to replicate two features that are typically found in various primate species: high ranking individuals receive more grooming than low ranking ones and individuals of adjacent rank groom each other more than expected from chance alone (Seyfarth, 1977). In an extension of this model, various types of competition for grooming access were investigated (Sambrook et al., 1995). However, both models explicitly assumed that high ranked individuals are attractive grooming partners, and thus the grooming decisions are not linked directly to the intensity of food competition itself. It would thus be relevant to have a quantitative model linking endogenously grooming decisions to the mode of food competition and determined conditions under which grooming up the hierarchy occurs as an equilibrium behaviour.

Here, we present a reinforcement learning model in which a group of individuals have grooming and feeding interactions, whose payoff consequences ("rewards") affect grooming, feeding, and tolerance decisions. This allows us to determine conditions under which food competition can lead to grooming being directed up the hierarchy under the grooming for tolerance hypothesis. Although grooming occurs frequently between kin, a meta-analysis demonstrated a significant effect of grooming reciprocity among non-kin (Schino and Aureli, 2008). Since our main interest here is the effect of food competition, we assume that individuals in the model are unrelated. However, we further assume that individuals spend their entire life together, which corresponds to the philopatric sex, and this typically means the females in primates (Pusey and Packer, 1987). Our main questions are: (i) can food competition result in grooming up the hierarchy as an equilibrium behaviour when individuals decide whom to groom, where to feed and whom to evict, dynamically according to payoffs, (ii) what kind of grooming patterns are expected if grooming is directed up the hierarchy (e.g., do all individuals groom the alpha individual, or do they mostly groom one rank up the hierarchy), and (iii) in how far does the strength of competition (due to the number of patches or the aggressiveness) modulates the grooming up the hierarchy.

The model

The biological setting

We consider a group of N individuals ranked in a stable, linear dominance hierarchy where individual $i \in \{1, 2, ..., N\}$ has a unique rank $r_i = i \in \{1, 2, ..., N\}$. Individual $i = 1$ is the lowest ranked and $i = N$ is the highest ranked or alpha individual. Individuals within this group interact with each others for T discrete time steps, which can be thought of as the number interactions on a daily basis, monthly basis, yearly basis, etc. Each interaction time step $t = 1, 2, ..., T$ is characterized by four sequential behavioural stages: grooming, food patch selection, aggressive interactions, and feeding, which occur in this order and that we now detail.

Stage 1 - Grooming

Each individual either selects an individual that it grooms or it grooms no one. The set of actions available to individual i in this stage is denoted $\mathcal{A}_i = \{0, 1, 2, ..., i-1, i+1, ..., N\}$, where $j \in \mathcal{A}_i$ is the rank of the recipient of the act of grooming, and where $j = 0$ means that individual i grooms no one. Grooming incurs a payoff cost c to the groomer, while contributing a payoff benefit b to the groomee (see table 1 for a list of parameters).

Stage 2 - Selecting a food patch

Each individual selects a food patch out of N_P available patches to feed on. We denote by $\mathcal{P} = \{1, 2, ..., N_\mathrm{P}\}$ the set of food patches so that action $k \in \mathcal{P}$ means an individual has selected food patch k.

Stage 3 - Aggressive interactions

After all individuals have chosen a patch, each individual makes a binary decision, to engage in a contest or not, for every other individual on the same patch. Because we assume that individuals can only engage others on the same food patch, we denote by $\mathcal{E}_{i,t} \subseteq \mathcal{A}_i$ the set of individuals that are on the same patch as i at time t.

After each individual has selected their targets (or no one) to engage, a contest occurs between the aggressor and the target, and the loser of this contest gets evicted from the patch. We adhere to the following rules for contest. (i) Individual i engaging j is sufficient for a contest to occur between i and j (it is not required that j also wants to engage i). (ii) To each individual we associate an array of target individuals. The order in which the contests occur is determined by random sampling (without replacement) from this array. (iii) Individuals that have lost a single contest, will not engage in any further contests in the current time step, as they are evicted from the patch. This thus means that selecting a target to engage does not necessarily mean that the focal individual and the target will have a contest, as one (or both) of them may be evicted by another individual before the contest occurs.

We assume that the contest does not incur any costs, as these interactions are often settled without physical contact (covert interactions). The outcome of each interaction is given by a contest success function (e.g., Hirshleifer, 1989) so that individual i engaged with j wins the contest with probability

$$\frac{\exp(dr_i)}{\exp(dr_i) + \exp(dr_j)}, \tag{1}$$

where $d \geq 0$ is the dominance gradient, which determines the influence of the ranks of the individuals on the outcome of the contest. If $d = 0$, then rank has no influence on the outcome and both individuals are equally likely to win, but if $d \to \infty$ the higher ranked individual is almost certain to win the contest.

Stage 4 - Feeding

After aggression occurred, individuals feed. We use the interference model to compute the payoff individuals gain on a given patch (Parker and Sutherland, 1986; Sutherland and Parker, 1992). The payoff per individual on patch k at time t is given by

$$\pi^\mathrm{P}_{k,t} = \frac{Q_k}{N^m_{k,t}}, \tag{2}$$

where $Q_k > 0$ is the quality of patch k, $N_{k,t}$ is the number of individuals on patch k at time t during the feeding stage and thus *after contests and eviction* occurred. The parameter $m \geq 0$ describes the interference between individuals on a food patch (a larger m resulting in a greater payoff decrease when an extra individual feeds on the patch). Thus an individual benefits from evicting others as its payoff increases when there are less competitors. After all individuals have fed, the group mixes again so that in the following time step each individual can have a grooming interaction with any other individual of the group again.

We consider three different scenarios regarding how the quality of patches is determined. (1) All patches are of equal quality, so that $Q_k = q_0$ for all k and q_0 is used as a baseline quality. (2) Half the patches are of high quality (equal to q_0), while the other half are low quality (equal to $0.5q_0$). In cases where the number of patches is odd, there is one extra high quality patch. (3) The quality of patch k is given by

$$Q_k = q_0(1/1.2)^{(k-1)}, \tag{3}$$

thus all patches differ in quality, where the quality asymptotically decreases with increasing k.

The learning process

Individuals take actions in three stages during an interaction step: whom to groom, select a food patch, and whom to evict. We assume that individuals learn which action to play by way of reinforcement learning. We follow standard reinforcement learning models as used in biology and game theory (e.g., Fudenberg and Levine, 1998; Camerer and Hua Ho, 1999; Achbany et al., 2006; Harley, 1981; McNamara and Houston, 1985; Hamblin and Giraldeau, 2009; Arbilly et al., 2010, 2011a,b; Dridi and Lehmann, 2014, 2015), and assume that each individual has internal motivations (or attractions) for each action, which get updated as a function of time according to payoff received and allows to express actions in a probabilistic way (e.g., Fudenberg and Levine, 1998; Camerer and Hua Ho, 1999; Dridi and Lehmann, 2014, 2015).

Action choice and motivation dynamics

Action choice for grooming (stage 1): The probability $p^{\mathrm{G}}_{i,t}(j)$ that individual i takes action $j \in \mathcal{A}_i$ during the grooming stage (i.e., grooms individual $j \neq i$ or grooming no one if $j = 0$) at time step t is given by

$$p^{\mathrm{G}}_{i,t}(j) = \frac{\exp[\lambda M^{\mathrm{G}}_{i,t}(j)]}{\sum\limits_{q \in \mathcal{A}_i} \exp[\lambda M^{\mathrm{G}}_{i,t}(q)]}, \tag{4}$$

where $M^{\mathrm{G}}_{i,t}(j)$ is the motivation that individual i has at time t for action $j \in \mathcal{A}_i$ (and can be thought of as the cumulative reward up to t of taking that action), and is assumed to be a real valued number ($M^{\mathrm{G}}_{i,t}(j) \in \mathbb{R}$). Eq. 4 is the standard logit choice rule for reinforcement learning of decision theory and neuroscience (e.g., Sutton and Barto, 1998; Fudenberg and Levine, 1998; Camerer and Hua Ho, 1999), and which has been used in behavioural ecology (e.g., Arbilly et al., 2010, 2011a,b; Dridi and Lehmann, 2015), see (Dridi and Lehmann, 2014) for more details and justifications on the framework we use here). The parameter $\lambda \geq 0$ therein can be seen as the sensitivity to motivations or exploration rate: if $\lambda = 0$ each action is chosen with equal probability regardless of motivation; if $\lambda \to \infty$ the action with the highest motivation is chosen almost deterministically: $p^{\mathrm{G}}_{i,t}(j) \to 1$.

The change $\Delta M^{\mathrm{G}}_{i,t}(j) = M^{\mathrm{G}}_{i,t+1}(j) - M^{\mathrm{G}}_{i,t}(j)$ in motivation (or attraction) of individual i for action $j \in \mathcal{A}_i$ at time step $t \geq 1$ is assumed to take the form

$$\Delta M^{\mathrm{G}}_{i,t}(j) = -\delta M^{\mathrm{G}}_{i,t}(j) + \pi^{\mathrm{G}}_{ij,t} + \pi^{\mathrm{F}}_{ij,t}. \tag{5}$$

The first term on the right-hand-side is a decay of motivation (for instance due to loss of memory) of i for action j, with discount factor δ. The second term, $\pi^{\mathrm{G}}_{ij,t}$, reflects the change in motivation due to grooming and is given by

$$\pi^{\mathrm{G}}_{ij,t} = \begin{cases} b\mathrm{I}^{\mathrm{G}}_{j\to i,t} - c\mathrm{I}^{\mathrm{G}}_{i\to j,t}, & \text{if } j \in \mathcal{A}_i \setminus 0 \\ \gamma \mathrm{I}^{\mathrm{G}}_{i\to j,t}, & \text{if } j = 0, \end{cases} \tag{6}$$

where $\mathrm{I}^{\mathrm{G}}_{i\to j,t}$ is an indicator function equal to 1 if i grooms j at time t, 0 otherwise. The payoff γ that i obtains if it grooms no one can be interpreted as the psychological reward of not grooming. Thus in words, assuming $b, c > 0$, the motivation of individual i to groom j increases with receiving grooming from j, but decreases if i grooms j.

The third term in eq. 5 is the (motivational) payoff

$$\pi^{\mathrm{F}}_{ij,t} = \mathrm{I}^{\mathrm{C}}_{ij,t}\pi^{\mathrm{P}}_{k(ij),t}\left(\frac{1}{2} - \frac{\exp(dr_i)}{\exp(dr_i) + \exp(dr_j)}\right) \tag{7}$$

due to co-feeding. This depends on $\mathrm{I}^{\mathrm{C}}_{ij,t}$, which is equal to 1 if i and j feed on the same patch at time t, zero otherwise. When i and j feed on the same patch individual i obtains material payoff $\pi^{\mathrm{P}}_{k(ij),t}$ (eq. 2), where $k(ij)$ is the patch on which i and j co-feed. We assume that the motivation reinforcement due to co-feeding (eq. 7) is proportional to this payoff and $(1/2 - \exp(r_i D)/[\exp(r_i D) + \exp(r_j D)])$. This is positive if i has a lower ranked than j, negative otherwise. Hence, we assume that a high ranked individual will decrease its motivation to groom a lower ranked one after feeding together on a patch, while the low ranked individual will increase its motivation to groom the high ranked one.

We note that the reinforcement learning rule for motivation implemented by eq. 5 (and the forthcoming ones) is an example of the the so-called *relative payoff sum* rule (Harley, 1981; Hamblin and Giraldeau, 2009; Dridi and Lehmann, 2014), where payoff over time is simply accumulated, and where the cumulated payoff decays at rate δ.

Action choice for patch selection (stage 2): The probability that at time t individual i selects patch $k \in \mathcal{P}$ for feeding is assumed to be given by

$$p^{\mathrm{P}}_{i,t}(k) = \frac{\exp[\lambda M^{\mathrm{P}}_{i,t}(k)]}{\sum\limits_{q\in\mathcal{P}} \exp[\lambda M^{\mathrm{P}}_{i,t}(q)]}, \tag{8}$$

which is similar to stage 1 (eq. 4) and where the change in the motivation $M^{\mathrm{P}}_{i,t}(k)$ that individual i has at time t for patch $k \in \mathcal{P}$ is updated as follows

$$\Delta M^{\mathrm{P}}_{i,t}(k) = -\delta M^{\mathrm{P}}_{i,t}(k) + \pi^{\mathrm{P}}_{k,t}\mathrm{I}^{\mathrm{P}}_{ik,t}. \tag{9}$$

The first term is again the discounted motivation and the second term is the patch payoff to an individual multiplied by the indicator for the individual on the patch, as in eq. 7. Note that, individuals that have been kicked out of a patch do not positively reinforce their patch motivations.

Action choice for aggression (stage 3): To determine action choice for this stage, we let $M^{\mathrm{T}}_{i,t}(j)$ represent the motivation of individual i to tolerate individual j ("the tolerance"). The change $\Delta M^{\mathrm{T}}_{i,t}(j) = M^{\mathrm{T}}_{i,t+1}(j) - M^{\mathrm{T}}_{i,t}(j)$ in tolerance is assumed to occur as follows:

$$\Delta M^{\mathrm{T}}_{i,t}(j) = -\delta M^{\mathrm{T}}_{i,t}(j) + b\mathrm{I}^{\mathrm{G}}_{j\to i,t} - \epsilon \mathrm{I}^{\mathrm{A}}_{ij,t}\mathrm{I}^{\mathrm{E}}_{ij,t}, \tag{10}$$

where the first term is the discounted motivation, the second term is the increase in tolerance if i has been groomed by j in round t (as in eq. 6), and the third term is the change in motivation as a result of the outcome of aggressive interactions, which depends on the parameter ϵ giving the impact of eviction on the motivation. In this third term, $\mathrm{I}^{\mathrm{A}}_{ij,t}$ takes value 1 if a contest occurred between i and j in round t, and 0 otherwise, and $\mathrm{I}^{\mathrm{E}}_{ij,t}$ takes value 1 if i won the contest, otherwise it takes value -1 (where probabilities of wining a contest is given by eq. 1). Hence, the winner of the contest decreases its motivation to tolerate the loser of the contest, while the loser increases it.

In this stage, each individual takes a sequence of eviction decisions, one for each individual in its patch (one for each $j \in \mathcal{E}_{i,t}$). We assume that the probability $p^{\mathrm{T}}_{i,t}(j)$ that individual i wants to evict j is given by

$$p^{\mathrm{T}}_{i,t}(j) = \frac{1}{1 + \exp[(M^{\mathrm{T}}_{i,t}(j) - \beta)]}. \tag{11}$$

This is a logistic function bounded between 0 and 1 (see Fig. 1), where the probability of engagement decreases with increasing motivation. As such, the function ensures that a higher tolerance motivation for individual j results in a lower probability of engaging that individual. The parameter β can be seen as the intrinsic aggressiveness of the individuals, as it regulates the probability that i will engage j with for a given value of its motivation to engage j, since with the same motivation a larger value for β increases the probability that i engages j (Fig. 1).

We note that motivations can become negative in the model (the last term in eq. 10 may induce negative motivation), and so β can be negative. But what matters for decision making is not the sign of the motivations, only their ranking and the model could be rescaled, such that motivations are only positive, but we refrain from doing so for simplicity of presentation and model analysis.

Descriptors of interactions

To recapitulate, the model allows to track the actions and the outcome of actions of each individual i for each round t of interaction ; that is, whether or not it grooms individual j ($\mathrm{I}^{\mathrm{G}}_{i\to j,t} \in \{0,1\}$), whether or not it feeds on patch k ($\mathrm{I}^{\mathrm{P}}_{ik,t} \in \{0,1\}$), whether or not it tries to evict j ($\mathrm{I}^{\mathrm{A}}_{ij,t} \in \{0,1\}$), whether it wins or looses a contest against j ($\mathrm{I}^{\mathrm{E}}_{ij,t} \in \{1,-1\}$), and thus to determine with whom it co-feeds ($\mathrm{I}^{\mathrm{C}}_{ij,t} \in \{0,1\}$). This in turn allows to describe the networks of interactions, which are of two kind. First, the grooming network (how much each individual grooms every other individual in the group), and second the feeding association network (proportion of time that pairs of individuals feed together), which can both be represented as graphs (Wasserman and Faust, 1994).

To described these networks, we first let the learning process run for $T_{\mathrm{eq}} < T$ time steps so that it reaches a quasi-equilibrium state and then characterize the networks in this state as follows. For the grooming network we computed the fraction g_{ij} of time that individual i spends grooming j at a quasi-equilibrium of the learning process. This is defined as

$$g_{ij} = \frac{1}{(T - T_{\mathrm{eq}})} \sum_{t=T_{\mathrm{eq}}+1}^{T} \mathrm{I}^{\mathrm{G}}_{i\to j,t}, \tag{12}$$

where $T - T_{\mathrm{eq}}$ is the amount of time we record the grooming interactions at quasi-equilibrium, whereby g_{i0} is the fraction of time i grooms no one. From g_{ij}, we also computed the average fraction of time that i receives grooming as

$$g_i = \frac{1}{N} \sum_{j=1}^{N} g_{ji}, \tag{13}$$

whereby $1 - g_i$ gives the fraction of time i is not groomed.

Since we are interested in characterizing the grooming patterns, we quantify the proportion of the grooming actions that are directed up the hierarchy as

$$h_{\mathrm{u}} = \frac{1}{(T - T_{\mathrm{eq}})} \sum_{t=T_{\mathrm{eq}}+1}^{T} \frac{\sum_{i=1}^{N} \sum_{j>i}^{N} \mathrm{I}_{i\to j,t}^{\mathrm{G}}}{\sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{I}_{i\to j,t}^{\mathrm{G}}}, \tag{14}$$

which varies between zero and one ($0 \leq h_{\mathrm{u}} \leq 1$). The numerator in the second ratio is the number of grooming actions in the group at time t that are directed towards an individual with a higher rank, while the denominator is the total number of grooming actions at that time. As such, if $h_{\mathrm{u}} = 0.5$ the amount of grooming received by higher ranked individuals is equal to the amount of grooming received by lower ranked individuals, while if $h_{\mathrm{u}} = 1$ then all grooming actions are directed towards an individual with a higher rank. Furthermore, to describe in more details the pattern of grooming up the hierarchy, we also consider

$$h_{\alpha} = \frac{1}{(T - T_{\mathrm{eq}})} \sum_{t=T_{\mathrm{eq}}+1}^{T} \frac{\sum_{i=1}^{N-1} \mathrm{I}_{i\to N,t}^{\mathrm{G}}}{\sum_{i=1}^{N-1} \sum_{j=1}^{N} \mathrm{I}_{i\to j,t}^{\mathrm{G}}}, \tag{15}$$

where the numerator in the second ratio is the total number of grooming actions directed towards the highest ranked individual (the "alpha individual" which has rank N), and the denominator is the total number of grooming actions. Hence, $h_{\alpha} \in [0, 1]$ gives the fraction of grooming (among all grooming interactions) directed towards the alpha individual, while

$$h_{1} = \frac{1}{(T - T_{\mathrm{eq}})} \sum_{t=T_{\mathrm{eq}}+1}^{T} \frac{\sum_{i=1}^{N-1} \mathrm{I}_{i\to i+1,t}^{\mathrm{G}}}{\sum_{i=1}^{N-1} \sum_{j=1}^{N} \mathrm{I}_{i\to j,t}^{\mathrm{G}}}, \tag{16}$$

gives the fraction of grooming that is is directed one rank up the hierarchy (we also have $h_1 \in [0, 1]$).

In order to characterize the feeding association network, we compute the frequency of feeding association between individuals i and j (a_{ij}) as,

$$a_{ij} = \frac{1}{(T - T_{\mathrm{eq}})} \sum_{t=T_{\mathrm{eq}}+1}^{T} \mathrm{I}_{ij,t}^{\mathrm{C}}, \tag{17}$$

where $\mathrm{I}_{ij,t}^{\mathrm{C}}$ is the co-feeding indicator function. We also compute the (quasi-equilibrium) frequency that individual i is not evicted and thus feeds on a patch (f_i) as,

$$f_i = 1 - \frac{1}{(T - T_{\mathrm{eq}})} \sum_{t=T_{\mathrm{eq}}+1}^{T} \sum_{j=1}^{N} -\mathrm{I}_{ij,t}^{\mathrm{A}} 0.5(\mathrm{I}_{ij,t}^{\mathrm{E}} - 1), \tag{18}$$

where the second term is the fraction of time individual i has been evicted. We subtract 1 to $\mathrm{I}_{ij,t}^{\mathrm{E}}$ to avoid counting the instances where individual i has won a contest (where $\mathrm{I}_{ij,t}^{\mathrm{E}} = 1$). Finally, we computed (at the quasi-stationary equilibrium), the average level of tolerance $\tau \in [0, 1]$ over all individuals in the population, and which is defined as the fraction of time that an individual has not engaged a partner, given that it is not alone on its patch. If $\tau = 0$, an individual never tries to evict a partner, while if $\tau = 1$ an individual always try to evict a partner, given that it is not alone on a patch.

Because the model is stochastic and highly non-linear we implemented it via individual-based simulations and tracked the actions and statistical descriptors numerically (simulation file can be obtained on request). For all reported results (see next section) we set all motivations of all individuals initially to 0 (i.e., $M^{\mathrm{G}}_{i,0} = M^{\mathrm{P}}_{i,0} = M^{\mathrm{T}}_{i,0} = 0$), which implies actions are uniformly distributed, and then let the model run for $T = 5000$ time steps while using a period of $T_{\mathrm{eq}} = 4000$ to reach the quasi-equilibrium state. We verified that $T_{\mathrm{eq}} = 4000$ time steps were generally enough to reach a quasi-equilibrium state, as the pattern of variation in the hierarchy score (h_{u}) between replicates remains stable after 500-2000 time steps (see the supplementary material). Assuming about 7 grooming interactions per day (e.g., rhesus monkeys, Lindburg, 1971), a period of 4000 time steps means a learning period of 1.5 years. To assess sensitivity to initial conditions we run 10 replicates for each set of parameter values, and thus when we report hierarchy scores ($h_{\mathrm{u}}, h_{\alpha}, h_1$) and tolerance level ($\tau$), we report their means (and sometimes standard errors).

The outcome of the model will be different grooming and association networks, depending on the intensity of food competition and other model parameters. Although the model uses a large number of parameters, our main interest is in the parameters that regulate the intensity of the competition and the value of the rewards (table 1). The main parameters varied here are the number of patches (N_{P}), the aggressiveness (β), the benefit of grooming (b), the dominance gradient (d), and the effect of patch quality (q_0), while we investigate the effect of varying group size (N) and the level of interference (m) in the supplementary material.

Results

In order to gain intuition about the outcomes of the model, we first consider two special cases: grooming without food competition (that is, no patch selection, aggressive interactions, and feeding) and food competition without grooming. Unless stated otherwise, all forthcoming results are based on the following baseline parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \beta = 1, \epsilon = 1, d = 2, q_0 = 6, m = 0.5, N = 10$ (table 1).

Uncoupled model

Grooming for grooming

In order to understand the incentive structure in our model, we start with the simplest case of grooming interactions for group size $N = 2$ (skipping stage 2, 3, and 4 of an interaction round). In this case, grooming motivations are only reinforced by receiving grooming (eq. 5 with $\mathrm{I}^{\mathrm{C}}_{ij,t} = 0$ always). This allows to investigate the conditions under which grooming as an equilibrium behaviour emerges (as opposed to choosing the non-grooming action "0"). For this case, we show in the supplementary material that when λ dominates δ ($\lambda \gg \delta$ so that reinforcement dominates stochastic choice), the change in the probabilities $p^{\mathrm{G}}_{1,t}(2)$ and $p^{\mathrm{G}}_{1,t}(2)$ that individual 1 grooms 2 and 2 grooms 1, respectively, can be approximated by the following differential equations

$$\begin{aligned}
\frac{\mathrm{d}p^{\mathrm{G}}_{1,t}(2)}{\mathrm{d}t} &= \lambda\left(1 - p^{\mathrm{G}}_{1,t}(2)\right)\left[p^{\mathrm{G}}_{2,t}(1)b - \left(p^{\mathrm{G}}_{1,t}(2)c + \left(1 - p^{\mathrm{G}}_{1,t}(2)\right)\gamma\right)\right] \\
\frac{\mathrm{d}p^{\mathrm{G}}_{2,t}(1)}{\mathrm{d}t} &= \lambda\left(1 - p^{\mathrm{G}}_{2,t}(1)\right)\left[p^{\mathrm{G}}_{1,t}(2)b - \left(p^{\mathrm{G}}_{1,t}(2)c + \left(1 - p^{\mathrm{G}}_{1,t}(2)\right)\gamma\right)\right]. \qquad (19)
\end{aligned}$$

This shows that an individual will increase the tendency to groom its partner proportionally to the level of grooming of the partner multiplied by the benefit b. The individual will decrease its grooming according to a term (second one in parenthesis in each equation) that can be thought off as the average of the cost c of grooming and the reward γ of not grooming. Eq. 19 makes explicit that if both interacting partner have an initial tendency to groom ($p^{\mathrm{G}}_{1,0}(2) > 0$ and $p^{\mathrm{G}}_{2,0}(1) > 0$), then it is possible that the learning dynamic converges to both individuals grooming each other. This is due to grooming being "reciprocated" and thus reinforced over time. On the other hand, if an individual interacts with another that has a very low grooming tendency, then it will not get exploited and will reduce its level of grooming.

Eq. 19 thus encapsulates the core behavioural assumptions behind our model and from a quantitative point of view, the global dynamics of learning depends on the parameter values. For instance, while fixing $b = 3$ and $c = 1$, we find that for $\gamma = 0.5$ pairs always end up grooming each other as the reinforcement when choosing the non-grooming action is very low (Fig. 2). For $\gamma = 2$ (which is equal to $b - c$) we find that half the pairs end up grooming, while the other half ends up not grooming. The reinforcement that individuals receive when grooming while receiving grooming ($b - c$) or when both do not groom (γ) is sufficiently high to converge on playing that action only (Fig. 2). For $\gamma = 5$ the reinforcement for not grooming is larger than grooming an individual that grooms back ($\gamma > b - c$), and most pairs end up not grooming.

For the remainder of the results we set $\gamma = 0.5$, as we want grooming to emerge, and then determine when it will be directed up the hierarchy. For instance, for $N = 10$ with $\gamma = 0.5$ we find that (while still skipping stage 2, 3, and 4) there exists virtually no variation in grooming propensity between individuals, and all individuals spent at least 98% of their time grooming others. Although giving grooming decreases the motivation to repeat that action, giving grooming to an individual that gives grooming back results in an overall increase in the motivation to groom that individual (since we always assume $b > c$ in eq. 5). As such, we expect and find that all individuals converge on grooming a single individuals predominantly, that also almost exclusively groom them back (Fig. 3a). Pairs of individuals thus trade equal amounts of grooming in the absence of food competition. This corresponds to the time-matching of grooming that is regularly found in primates (e.g., Schino et al., 2003).

Tolerance without grooming

If the grooming stage is skipped, individuals can still select a patch to feed, evict targets, and feed, but there can be no grooming for tolerance. In this case, tolerance motivation no longer depends on grooming (eq. 10 with $\mathrm{I}^{\mathrm{G}}_{j \to i,t} = 0$ always), but can still depend on the contests that have occurred, and where the number of contests depends on the intrinsic aggressiveness of the individuals (eq. 11). We thus find that a larger β (higher aggressiveness) results in less tolerance (Fig. 3b). Additionally, the number of patches (N_{P}) influences the average number of competitors on a patch, which in turn influences the likelihood of engaging at least one other individual. Thus fewer patches means more competitors on the same patch and therefore less tolerance (Fig. 3b).

Full model

Grooming up the hierarchy for tolerance

We now analyze how grooming for tolerance occurs in the full model (allowing all four interaction stages to occur). We find that if there is food competition, not all individuals manage to feed on a patch and several individuals get evicted. Regularly no more than two individuals feed on the same patch (Fig. 4a, c). A third (lower ranked) individual on the same patch would not to be tolerated by at least one other individual and thus gets evicted. Hence, individuals tend to feed as dyads.

Although the distribution of the highest ranked individuals over the available food patches differs between replicates of simulations, the highest ranked individuals consistently segregate over the patches in order to avoid competition with one another (Fig. 4a, c). Between the individuals that feed together on a patch, the lower ranked individual always grooms the higher ranked one more than the higher ranked individual grooms back, and is the only individual that is tolerated by the higher ranked one (Fig. 4b, d). This is quantitatively captured by our hierarchy score, h_{u}, which is systematically above 0.5 (Figs. 5, 7). For individuals that do not co-feed on a patch, the grooming between them is either balanced or absent. Thus only when individuals feed together an imbalance of grooming within that pair is created where the higher ranked individual receives additional grooming. Due to the imbalance of grooming between individuals sharing a patch (the lower ranked grooming more), we can thus conclude that our model generates grooming for tolerance; namely, grooming is exchanged for tolerance during feeding.

We also find that grooming up the hierarchy follows a systematic qualitative pattern in which the more there is grooming up the hierarchy, the more this is directed towards the alpha individual. We indeed find a strong positive correlation between h_u and h_α (Pearson's $r = 0.48$, Fig. 5). Conversely, the grooming just one rank up the hierarchy (h_1) is negatively correlated with h_u (Pearson's $r = -0.15$). From the feeding associations, like in Fig. 4, this is not surprising, since closely ranked individuals segregate over the available food patches and thus do not need to groom each other in order to be tolerated on a patch. Occasionally neighbour ranked individuals may share a patch, but it is clear from our results that grooming up the hierarchy is a result of the highest ranked individuals (in particular the alpha) receive additional grooming and not because individuals groom one rank up the hierarchy.

Conditions favouring grooming for tolerance

We now delineate how grooming up the hierarchy depends on various parameters, like patch number and quality (affecting competition), aggressiveness and dominance gradient, and grooming costs and benefits.

We find that grooming up the hierarchy depends markedly on competition for resources. Indeed, the hierarchy score depends strongly on the number of available food patches (Fig. 5b). If there is just a single patch to feed on ($N_\mathrm{P} = 1$), competition for feeding is high, and thus there is little tolerance ($\tau = 0.04$). Most individuals are evicted from the patch, and will not groom for tolerance, resulting in a hierarchy score close to 0.5 ($h_\mathrm{u} = 0.52$). The slight amount of grooming that does occur up the hierarchy (since $h_\mathrm{u} > 0.5$) is due to the second highest ranked individual being tolerated and grooming the alpha.

If the number of food patches is increased, individuals will segregate over the available patches (see Fig. 4). Fewer individuals per patch means there will be less competition within each patch and thus more tolerance (Fig. 3b). The highest ranked individual on each patch is likely to tolerate one other individual in exchange for grooming (Fig. 4), and since these exchanges of grooming for tolerance occur more frequently if there are more patches, then there will consequently be more grooming up the hierarchy (Fig. 5b). Most grooming up the hierarchy occurs for $N_\mathrm{P} = 4$ ($h_\mathrm{u} = 0.61$). A further increase in the number of patches results in alleviating some of the competition, as individuals may sometimes find themselves alone on a patch, and therefore less tolerance and thus grooming up the hierarchy is required. We thus observe that the hierarchy score is non-monotonic in number of food patches, first increasing and then decreasing for $N_\mathrm{P} > 4$ (Fig. 5b). However, even for high N_P there remains an overall tendency to groom up the hierarchy ($h_\mathrm{u} = 0.58$ for $N_\mathrm{P} = 10$), since the extra patches are of such low quality that individuals avoid feeding on these (quality of the best patch: $Q_1 = 6$, worst patch: $Q_{10} = 1.16$). In the online supplementary material we show that if there are as many patches as individuals and all patches are of equal quality, then grooming is completely reciprocal ($h_\mathrm{u} = 0.50$), since there is no co-feeding and grooming for tolerance is not required (Fig. S2). Thus, in the absence of competition for resources we recover the grooming for grooming results (section "Grooming for grooming").

For the effect of patch quality on grooming, we find that when this quality is ranked (eq. 3) lowering the baseline quality, q_0, results in less grooming up the hierarchy. A lower patch quality means that individuals gain less material (and thus motivational) payoff on patches, and thus motivations for patch choice are less strongly affected from feeding (eq. 9). Individuals are then more likely to choose different patches at each time step and therefore encounter different individuals on the patch regularly. Strong feeding associations between individuals (i.e., feeding together regularly) will therefore not occur and thus there will be little grooming for tolerance (Fig. 6). Increasing patch quality increases grooming up the hierarchy, because higher ranked individuals will be less inclined to groom lower ranked individuals, while lower ranked individuals will reward higher ranking ones even more (eq. 5). Evicting another individual has a fixed effect on tolerance motivation, and the effect of being groomed outweighed the effect of evicting ($b > \epsilon$). Therefore, if only the patch quality (q_0) changes, dominant individuals would continue to tolerate subordinates.

Grooming up the hierarchy also depends on the intrinsic aggressiveness of individuals (β). Lowering β means individuals are less likely to engage others for eviction (in stage 3 of the interaction stage), and individuals are tolerated more often (Fig. 3b). A consequence of more tolerance is that low ranked individuals are more likely to reward higher ranked individuals with grooming and grooming up the hierarchy increases (Fig. 7a). The effect of changing aggressiveness is most clear under conditions of high food competition (i.e., low number of food patches). For $N_\mathrm{P} = 1$ and $\beta = -3$ the tolerance is greatly increased ($\tau = 0.84$) as well as the hierarchy score ($h_\mathrm{u} = 0.81$). Conversely, higher aggressiveness ($\beta = 3$) increases the number of contests, thus lowering the tolerance. Then, if low ranked individuals are not tolerated on food patches, they do not groom higher ranked ones in return. As such, increasing β results in less tolerance and thus grooming becoming more reciprocal (i.e., $h_\mathrm{u} \to 0.5$), and for $N_\mathrm{P} = 4$ we find $h_\mathrm{u} = 0.56$ (Fig. 7b).

We find that a positive dominance gradient (d) is needed for grooming to be directed up the hierarchy. If $d = 0$, all individuals are symmetrical. In this case, no individual can consistently control a food patch by evicting others, since each individual has a 0.5 probability to win a contest with every other individual (eq. 1). For $d = 0$ we thus find that $h_\mathrm{u} = 0.5$ for any number of patches (N_P), showing that the intensity of food competition has no effect on grooming decisions. If individuals are symmetrical, we thus recover that grooming is exchanged for grooming (Fig. 8). For $d = 2$ the probability to win a contest with an individual one rank below is 0.88 (eq. 1) and thus a higher ranked individual already wins most contests. We find that increasing the dominance gradient has little effect on the hierarchy scores as the contest success function (eq. 1) saturates (e.g., $d = 10$, Fig. 8).

Finally, we investigated how the cost to benefit ratio of grooming affects grooming up the hierarchy by varying the benefit b (but always holding $b > c$). If b is large, grooming motivations are increasingly controlled by whom an individual is groomed (eq. 5), and thus grooming will become increasingly reciprocal (for $b = 4.5$ we find $h_\mathrm{u} = 0.5$ regardless of number of patches; Fig. 9). The tolerance motivation similarly increase with b (eq. 10), and thus we continue to find individuals feeding together, but with reduced asymmetry of grooming between them. Hence, we have grooming for grooming with high tolerance. If b is low, we again observe a high tendency to reciprocal grooming, since tolerance motivations will be less affected by receiving grooming (Fig. 9). Concomitantly one then has less tolerance overall. This shows that the benefit of grooming (b) has a non-monotonic effect on grooming up the hierarchy.

Sensitivity of the results

In the online supplementary material, we present several additional results to test the robustness of our qualitative results. We show that the assumption regarding the distribution of patch quality (whether patches are of equal quality or of low and high quality, see eq. 3) has a surprisingly small effect on the outcome (Fig. S2). We also vary the group size N and the interference while feeding (m) and find that we can reproduce qualitatively similar levels of grooming up the hierarchy (Fig. S3 and Fig. S4). Increasing group size lowers the grooming up the hierarchy slightly, however, since more dyads groom reciprocally, while increasing interference has a similar effect as lowering patch quality (q_0) as both affect the payoff an individual gain on a patch, and thus influence the strength of feeding associations between individuals. Finally, we attempted to create conditions where individuals preferentially groom one rank up the hierarchy (instead of the alpha individual). We do this by having patch payoff decrease dramatically if more than two individual feed on the same patch. Then, by creating substantial differences between patch quality we expect that mostly the highest ranked individuals feed together on the highest quality patch, the third and fourth ranked individuals on the second highest quality patch, and so forth. However, we find even under such conditions that grooming up the hierarchy is preferentially directed towards the alpha individual ($h_\alpha > h_1$, Fig. S5).

Discussion

Grooming is maybe the most documented behaviour in the primate literature, as it is both relatively easy to quantify and understanding its occurrence is a major goal in primatology (Seyfarth, 1980; Schino, 2001). Grooming is regularly found to be directed up the hierarchy, where lower ranked individuals groom higher ranked ones. It has been argued that this occurs when there is competition for food, where low ranked individuals trade grooming for either tolerance or coalitionary support from high ranked individuals. To better understand the conditions under which this is likely to occur, we developed here a quantitative reinforcement learning model, where individuals take decisions on whom to groom, where to feed, and whom to tolerate on a food patch, and where these actions depend dynamically on payoffs. With this model we have investigated if food competition can result in grooming being directed up the hierarchy (qualitatively), what are the grooming patterns that emerge (which individuals receive additional grooming), and under which conditions it is likely to occur (and break down).

Grooming for tolerance

Our main finding is that low ranked individuals are indeed likely to trade grooming for tolerance from high ranked individuals (Fig. 4) leading to grooming being directed up the hierarchy (Fig. 5). This matches the empirical observations made in various primate species where grooming is regularly found to be traded for tolerance (Ventura et al., 2006; Carne et al., 2011; Tiddi et al., 2011; Xia et al., 2012; Borgeaud and Bshary, 2015).

Our model thus demonstrates that grooming can be traded for tolerance as a result of food competition. Seyfarth's simulation model on grooming in female monkeys was seminal in explicitly investigating grooming patterns. The model was relatively simple in the sense that the individuals in the model only made grooming decisions, and were assumed to have an innate preference to groom the highest ranking individuals (Seyfarth, 1977). However, under conditions where high ranked individuals cannot provide an alternative benefit to grooming (e.g., tolerance or coalitionary support), low ranked individuals should adjust their grooming preferences accordingly. With a fixed preference, individuals would always groom up the hierarchy. The current work is thus a logical next step, since we allowed grooming preferences (motivation) to vary between individuals and over time. Our model successfully produced, depending on the state of the environment, both grooming for tolerance and grooming for grooming, the latter not being possible in Seyfarth's model (Seyfarth, 1977). Our results thus demonstrate that dynamic motivations (or preferences) that are updated through learning can capture a wider range of grooming patterns, and that grooming up the hierarchy depends crucially on ecological and behavioural conditions.

The pattern of grooming up the hierarchy

We find that grooming up the hierarchy is mostly the result of the alpha individual receiving additional grooming, but not the individual one rank up the hierarchy (Fig. 5). It has been argued that all individuals in a primate troop should be willing to preferentially groom the alpha individual (followed by the second highest ranked individual, etc.), but that, at the same time, there is competition for grooming access, which means that not all individuals can groom their preferred target. Consequently, most individuals may end up grooming adjacent ranked individuals instead (Seyfarth, 1977, 1980). However, both the evidence for competition for grooming access and more grooming between adjacent ranked individuals is mixed (in favour of competition for grooming access: (Fairbanks, 1980; Schino, 2001), against: (Silk, 1982; Henzi et al., 2003); in favour of grooming adjacent ranked: (Seyfarth, 1977, 1980), against: (Thierry et al., 1990; O'Brien, 1993)). Therefore, it is worth investigating what conditions lead to competition for grooming access or increased grooming of adjacent ranked individuals.

In our model we do not find preferential grooming of individuals one rank above in the hierarchy. Interestingly, we find that adjacent ranked individuals tend to avoid one another in order to be the highest ranked individual on their chosen food patch (Fig. 4). In the absence of feeding associations between adjacent ranked individuals, there is no incentive for the lower ranked individual to groom the higher ranked one for tolerance. Therefore, the factors causing individuals to groom one rank up or down the hierarchy need further investigation, but it seems unlikely that this is directly caused by food competition. An interesting/logical extension of the current model would be to introduce some form of competition for grooming partners in order to investigate if such a restriction will result in individuals grooming adjacent ranked individuals more frequently.

Conditions where grooming up the hierarchy occurs and breaks down

We have quantified how the intensity of food competition (due to the number of food patches, aggressiveness, etc) modulates the grooming up the hierarchy. The so-called socioecological model (Sterck et al., 1997) tries to understand how food competition, predation risk, and infanticide risk determine various aspects of primate societies (e.g., patterns of migration, dominance hierarchies, and grooming behaviour) (van Schaik, 1996; Sterck et al., 1997; Koenig, 2002; Kappeler and van Schaik, 2002). While the socioecological model predicts that food competition may cause grooming to be directed up the hierarchy, it does not give quantitative predictions regarding how the intensity of food competition will modulate the grooming patterns (it only distinguishes between types of competitions, Sterck et al., 1997).

Using the socioecological model as a premise for our own analysis, we find that our model replicates a variety of findings in the primate literature where grooming is directed up the hierarchy due to food competition (Ventura et al., 2006; Carne et al., 2011; Tiddi et al., 2011; Xia et al., 2012). Although it is argued in these studies that grooming is traded for tolerance on food sources, the intensity of food competition is rarely investigated. However, our model shows that the relationship between food competition and grooming up the hierarchy is not necessarily linear (Fig. 5). Therefore, in order to better understand observed patterns of grooming behaviour in primates, it is crucial to quantify the intensity of food competition.

Additionally, we identified conditions under which grooming will be traded for grooming instead of tolerance. For example, the model predicts that when there is a sufficient number of high quality patches available grooming will be reciprocal (Fig. S2). This matches the finding by Pazol and Cords (Pazol and Cords, 2005) where, despite the presence of a dominance hierarchy, blue monkeys actively avoided one another while feeding, and high and low ranking individuals therefore received equal amounts of grooming. Similarly, in baboon troops where individuals did not compete for resources, a strong time-matching pattern of grooming was found (Barrett et al., 1999).

Our model also predicts that high aggressiveness can lead to less grooming up the hierarchy (Fig. 7). This conforms to the finding of Leinfelder (Leinfelder et al., 2001), where grooming is solely traded for grooming, despite the presence of monopolizable food sources, and this is attributed to high food-related aggression (Leinfelder et al., 2001). Obviously, aggressiveness varies between individuals, groups, and species (Isbell, 1991; Fairbanks et al., 2004). However, for the sake of simplicity we used a single parameter to set the aggressiveness of all individuals here. This showed that high aggression results in grooming being traded for grooming. Further work is required to investigate how grooming networks are affected if aggressiveness can vary between individuals.

Finally, there may exist other factors that are not captured in our model, which also can favour reciprocal grooming over grooming up the hierarchy. For example, in a troop of blue monkeys grooming was found to be traded reciprocally, but this appeared to be the result of intense between group competition, which increases the importance of group cohesiveness leading to equal shares of grooming within dyads (Rowell et al., 1991). Since we investigated grooming only as a result of within group competition, such effects could not occur in our model.

Emotional bookkeeping

Although primates reciprocate grooming and trade grooming for other commodities, the mechanistic processes underlying these decisions are difficult to unravel. One of the hypothesized explanations how primates take actions in a social context is through emotional bookkeeping (Aureli and Schaffner, 2002; Schino and Aureli, 2009, 2017). Primates are thought to have partner-specific emotions that guide their decision making process. Receiving benefits, in various currencies, are translated into single variables (emotions) in an individual towards those that provided the benefit. Through social interactions individuals will form differential social relationships. The emotions can thus be seen as a measure of the relationship quality, and will determine the actions an individual takes in a social context. Evidence supporting emotional bookkeeping has been found in various species (Schino et al., 2009; Schino and Pellegrini, 2009).

Our model can be interpreted as an application of emotional bookkeeping, where the motivations can be seen as emotional states, since they reflect the history of interactions between the individuals. Different currencies are translated into a single motivation: both tolerance and receiving grooming affect the grooming motivation (eq. 5). Although emotional bookkeeping has, until now, only been described verbally, our model provides a quantitative example of how partner specific motivations (or emotions) may be updated through interactions with social partners or decay over time. Different types of interactions (grooming, co-feeding, contests) can all affect a single motivation, which consequently will determine the actions an individual takes. Our model shows that through such updating mechanisms stable grooming bonds can be formed between individuals, while at the same time it allows for interchange of currencies (grooming for tolerance) between individuals of different ranks (Fig. 4).

Assumptions of the model

Our model is a proximate one that captures grooming and feeding actions in a group of interacting individuals, and we did not study the evolution of the decision rules here. Therefore, we have made a variety of assumptions for this model, especially regarding how motivations are updated as a result of payoff received.

We use a reinforcement learning model (instead of an evolutionary model) because both food sources and grooming partners widely fluctuate in a primate's life (Borgeaud et al., 2016), which makes it likely that social strategies are learned as it allows individuals to respond to the variation in social partners and the environment it will experience over time. We assumed that the motivation to groom another increases with receiving grooming from that individual, while it decreases with giving grooming to that individual. Such updating matches the evolution of reciprocal strategies in many models on social evolution, the most well-known being tit-for-tat, which can be seen as a simple learning strategy (Dridi and Lehmann, 2014). We choose parameter values such that grooming is favoured (see Fig. 2). The assumptions of our model thus do not favour "cheating" behaviour (choosing not to groom), since grooming acts are likely to be reciprocated. We were interested in patterns of grooming behaviour here, and not in how individuals can avoid being cheated. Regarding tolerance, we assumed that winning a contest would increase the probability to engage the same individual in future time steps (thus becoming less tolerant). Since evicting another from a food patch is likely to increase the payoff of the evicting individual in nature, repeating such an action will increase its payoff similarly in future time steps, and such a learning rule is thus likely to increase the payoff overall. Finally, we assumed that receiving grooming will increase the tolerance towards the groomer, which has been shown experimentally (Borgeaud and Bshary, 2015), and we assumed that tolerating a lower ranked causes the higher ranked to be less likely to groom that individual, which is a logical consequence if grooming is to be traded for tolerance instead of both tolerance and grooming.

Outlook for empiricists

Our model generates patterns of grooming reciprocity (trading grooming for grooming) and grooming up the hierarchy (trading grooming for tolerance), which matches a number of grooming patterns in natural populations. These different patterns emerge from the model by varying its parameters. First, varying the level of food competition, as regulated through the number of food patches, can result in complete reciprocal grooming (given a sufficient number of high quality patches), as well as in strong grooming up the hierarchy if competition increases, although high levels of competition may also decrease the grooming up the hierarchy. Therefore, a proper understanding of the intensity of the food competition in natural populations is crucial in understanding the observed grooming pattern. Second, high aggressiveness suppresses grooming up the hierarchy. Quantifying the variation in food related aggressiveness between species, populations, and even individuals can help explain the absence or presence of grooming up the hierarchy. Additional interesting factors to measure, albeit more difficult to assess, are the patch qualities and interference as they tend to influence the feeding associations that form between individuals, and this will influence whether grooming is traded for tolerance. Overall, a better understanding of these factors, will help improve our understanding of grooming behaviours and social relationships in primates.

References

Achbany, Y., Fouss, F., Yen, L., Pirotte, A., and Saerens, M. Optimal Tuning of Continual Online Exploration in Reinforcement Learning. In Kollias, S., Stafylopatis, A., Duch, W., and Oja, E., editors, *Artificial Neural Networks ICANN 2006*, volume 4131 of *Lecture Notes in Computer Science*, pages 790–800. Springer Berlin Heidelberg, 2006.

Arbilly, M., Motro, U., Feldman, M. W., and Lotem, A. Co-evolution of learning complexity and social foraging strategies. *Journal of Theoretical Biology*, 267(4):573–581, 2010.

Arbilly, M., Motro, U., Feldman, M. W., and Lotem, A. Evolution of social learning when high expected payoffs are associated with high risk of failure. *Journal of The Royal Society Interface*, 8(64):1604–1615, 2011a.

Arbilly, M., Motro, U., Feldman, M. W., and Lotem, A. Recombination and the evolution of coordinated phenotypic expression in a frequency-dependent game. *Theoretical Population Biology*, 80(4):244–255, 2011b.

Aureli, F. and Schaffner, C. Relationship assessment through emotional mediation. *Behaviour*, 139(2):393–420, 2002.

Barrett, L., Henzi, S. P., Weingrill, T., Lycett, J. E., and Hill, R. A. Market forces predict grooming reciprocity in female baboons. *Proceedings of the Royal Society of London B: Biological Sciences*, 266(1420):665–670, 1999.

Barrett, L. and Henzi. Monkeys, markets and minds: biological markets and primate sociality. In Kappeler, P. and van Schaik, C., editors, *Cooperation in Primates and Humans*, pages 209–232. Springer Berlin Heidelberg, 2006.

Borgeaud, C. and Bshary, R. Wild Vervet Monkeys Trade Tolerance and Specific Coalitionary Support for Grooming in Experimentally Induced Conflicts. *Current Biology*, 25(22):3011–3016, 2015.

Borgeaud, C., Sosa, S., Bshary, R., Sueur, C., and van de Waal, E. Intergroup Variation of Social Relationships in Wild Vervet Monkeys: A Dynamic Network Approach. *Frontiers in Psychology*, 7, 2016.

Camerer, C. and Hua Ho, T. Experience-weighted Attraction Learning in Normal Form Games. *Econometrica*, 67(4):827–874, 1999.

Carne, C., Wiper, S., and Semple, S. Reciprocation and interchange of grooming, agonistic support, feeding tolerance, and aggression in semi-free-ranging Barbary macaques. *American Journal of Primatology*, 73(11):1127–1133, 2011.

Cheney, D. L., Moscovice, L. R., Heesen, M., Mundry, R., and Seyfarth, R. M. Contingent cooperation between wild female baboons. *Proceedings of the National Academy of Sciences*, 107(21):9562–9566, 2010.

Dridi, S. and Lehmann, L. On learning dynamics underlying the evolution of learning rules. *Theoretical Population Biology*, 91:20–36, 2014.

Dridi, S. and Lehmann, L. A model for the evolution of reinforcement learning in fluctuating games. *Animal Behaviour*, 104:1–28, 2015.

Dunbar, R. I. M. Functional Significance of Social Grooming in Primates. *Folia Primatologica*, 57 (3):121–131, 1991.

Dunbar, R. I. M. Modelling Primate Behavioral Ecology. *International Journal of Primatology*, V23(4):785–819, 2002.

Fairbanks, L. A. Relationships among adult females in captive vervet monkeys: Testing a model of rank-related attractiveness. *Animal Behaviour*, 28(3):853–859, 1980.

Fairbanks, L. A., Newman, T. K., Bailey, J. N., Jorgensen, M. J., Breidenthal, S. E., Ophoff, R. A., Comuzzie, A. G., Martin, L. J., and Rogers, J. Genetic contributions to social impulsivity and aggressiveness in vervet monkeys. *Biological Psychiatry*, 55(6):642–647, 2004.

Fudenberg, D. and Levine, D. K. *The theory of learning in games.* MIT Press, Cambridge, MA, 1998.

Hamblin, S. and Giraldeau, L.-A. Finding the evolutionarily stable learning rule for frequency-dependent foraging. *Animal Behaviour*, 78(6):1343–1350, 2009.

Harley, C. B. Learning the evolutionarily stable strategy. *Journal of Theoretical Biology*, 89(4): 611–633, 1981.

Henzi and Barrett, L. The value of grooming to female primates. *Primates*, 40(1):47–59, 1999.

Henzi, S. P., Barrett, L., Gaynor, D., Greeff, J., Weingrill, T., and Hill, R. A. Effect of resource competition on the long-term allocation of grooming by female baboons: evaluating Seyfarth's model. *Animal Behaviour*, 66(5):931–938, 2003.

Hirshleifer, J. Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. *Public Choice*, 63(2):101–112, 1989.

Isbell, L. A. Contest and scramble competition: patterns of female aggression and ranging behavior among primates. *Behavioral Ecology*, 2(2):143–155, 1991.

Kappeler, P. M. and van Schaik, C. P. Evolution of Primate Social Systems. *International Journal of Primatology*, V23(4):707–740, 2002.

Koenig, A. Competition for Resources and Its Behavioral Consequences Among Female Primates. *International Journal of Primatology*, V23(4):759–783, 2002.

Leinfelder, I., de Vries, H., Deleu, R., and Nelissen, M. Rank and grooming reciprocity among females in a mixed-sex group of captive hamadryas baboons. *American Journal of Primatology*, 55(1):25–42, 2001.

Lindburg. *The Rhesus Monkey in North India: An Ecological and Behavioral Study*, pages 1–106. New York: Academic Press, 1971.

Maestripieri, D. Vigilance Costs of Allogrooming in Macaque Mothers. *The American Naturalist*, 141(5):744–753, 1993.

McNamara, J. M. and Houston, A. I. Optimal foraging and learning. *Journal of Theoretical Biology*, 117(2):231–249, 1985.

O'Brien, T. G. Allogrooming behaviour among adult female wedge-capped capuchin monkeys. *Animal Behaviour*, 46(3):499–510, 1993.

Parker, G. A. and Sutherland, W. J. Ideal free distributions when individuals differ in competitive ability: phenotype-limited ideal free models. *Animal Behaviour*, 34(4):1222–1242, 1986.

Pazol, K. and Cords, M. Seasonal variation in feeding behavior, competition and female social relationships in a forest dwelling guenon, the blue monkey (Cercopithecus mitis stuhlmanni), in the Kakamega Forest, Kenya. *Behavioral Ecology and Sociobiology*, 58(6):566–577, 2005.

Pusey, A. E. and Packer, C. Dispersal and philopatry. In Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W., and Struhsaker, T. T., editors, *Primate Societies*, pages 250–266. Chicago University Press, Chicago, 1987.

Roubová, V., Konečná, M., Šmilauer, P., and Wallner, B. Whom to Groom and for What? Patterns of Grooming in Female Barbary Macaques (Macaca sylvanus). *PLoS ONE*, 10(2):e0117298+, 2015.

Rowell, T. E., Wilson, C., and Cords, M. Reciprocity and partner preference in grooming of female blue monkeys. 12(4):319–336, 1991.

Sambrook, T. D., Whiten, A., and Strum, S. C. Priority of access and grooming patterns of females in a large and a small group of olive baboons. *Animal Behaviour*, 50(6):1667–1682, 1995.

Schino, G. Grooming, competition and social rank among female primates: a meta-analysis. *Animal Behaviour*, 62(2):265–271, 2001.

Schino, G. Grooming and agonistic support: a meta-analysis of primate reciprocal altruism. *Behavioral Ecology*, 18(1):115–120, 2007.

Schino, G. and Aureli, F. Grooming reciprocation among female primates: a meta-analysis. *Biology Letters*, 4(1):9–11, 2008.

Schino, G. and Aureli, F. *Chapter 2 Reciprocal Altruism in Primates*, volume 39, pages 45–69. Elsevier, 2009.

Schino, G. and Aureli, F. Reciprocity in group-living animals: partner control versus partner choice. *Biological Reviews*, 92:665–672, 2017.

Schino, G. and Pellegrini, B. Grooming in mandrills and the time frame of reciprocal partner choice. *American Journal of Primatology*, 71(10):884–888, 2009.

Schino, G., Ventura, R., and Troisi, A. Grooming among female Japanese macaques: distinguishing between reciprocation and interchange. *Behavioral Ecology*, 14(6):887–891, 2003.

Schino, G., Di Giuseppe, F., and Visalberghi, E. The Time Frame of Partner Choice in the Grooming Reciprocation of Cebus apella. *Ethology*, 115(1):70–76, 2009.

Seyfarth, R. M. A model of social grooming among adult female monkeys. *Journal of Theoretical Biology*, 65(4):671–698, 1977.

Seyfarth, R. M. The distribution of grooming and related behaviours among adult female vervet monkeys. *Animal Behaviour*, 28(3):798–813, 1980.

Seyfarth, R. M. and Cheney, D. L. Grooming, alliances and reciprocal altruism in vervet monkeys. *Nature*, 308(5959):541–543, 1984.

Silk, J. B. Altruism Among Female Macaca Radiata: Explanations and Analysis of Patterns of Grooming and Coalition Formation. *Behaviour*, 79(2):162–188, 1982.

Sterck, E. H. M., Watts, D. P., and van Schaik, C. P. The evolution of female social relationships in nonhuman primates. *Behavioral Ecology and Sociobiology*, 41(5):291–309, 1997.

Sutherland, W. J. and Parker, G. A. The relationship between continuous input and interference models of ideal free distributions with unequal competitors. *Animal Behaviour*, 44:345–355, 1992.

Sutton, R. S. and Barto, A. G. *Reinforcement Learning.* MIT Press, Cambridge, MA, 1998.

Tanaka, I. and Takefushi, H. Elimination of External Parasites(Lice) Is the Primary Function of Grooming in Free-ranging Japanese Macaques. *Anthropological Science*, 101(2):187–193, 1993.

Thierry, B., Gauthier, C., and Peignot, P. Social grooming in Tonkean macaques (Macaca tonkeana). *International Journal of Primatology*, 11(4):357–375, 1990.

Tiddi, B., Aureli, F., Polizzi di Sorrentino, E., Janson, C. H., and Schino, G. Grooming for tolerance? Two mechanisms of exchange in wild tufted capuchin monkeys. *Behavioral Ecology*, 22(3):663–669, 2011.

van Schaik, C. P. The ecology of social relationships amongst female primates. In Standen, V. and Foley, R. A., editors, *Comparative socioecology. The behavioural ecology of humans and other mammals.*, pages 195–218. Blackwell, Oxford, 1989.

van Schaik, C. P. Social Evolution in Primates: The Role of Ecological Factors and Male Behaviour. *Proceedings of the British Academy*, 88:9–31, 1996.

Ventura, R., Majolo, B., Koyama, N. F., Hardie, S., and Schino, G. Reciprocation and interchange in wild Japanese macaques: grooming, cofeeding, and agonistic support. *American Journal of Primatology*, 68(12):1138–1149, 2006.

Wasserman, S. and Faust, K. *Social Network Analysis.* Cambridge University Press, Cambridge, 1994.

Wei, W., Qi, X.-G., Guo, S.-T., Zhao, D.-P., Zhang, P., Huang, K., and Li, B.-G. Market Powers Predict Reciprocal Grooming in Golden Snub-Nosed Monkeys (Rhinopithecus roxellana). *PLoS ONE*, 7(5):e36802+, 2012.

Wrangham, R. W. An Ecological Model of Female-Bonded Primate Groups. *Behaviour*, 75(3): 262–300, 1980.

Xia, D., Li, J., Garber, P. A., Sun, L., Zhu, Y., and Sun, B. Grooming Reciprocity in Female Tibetan Macaques Macaca Thibetana. *American Journal of Primatology*, 74(6):569–579, 2012.

Table and figures

Table 1: List of model parameters. The parameters in bold are varied to investigate its effect on the hierarchy score (h_{u}).

Parameter	**Equations**	**Meaning**
c	6	Cost of grooming
$\boldsymbol{b}$	6, 10	Benefit of being groomed
γ	6	Change in motivation when not grooming
λ	4, 8	Exploration rate
δ	5, 9, 10	Discount factor
$\boldsymbol{\beta}$	11	Regulate engage probability
ϵ	10	Impact of eviction
$\boldsymbol{d}$	1, 5	Dominance gradient
$\boldsymbol{q_0}$	3	Reference patch quality
$\boldsymbol{m}$	2	Interference
$\boldsymbol{N}$		Group size
$\boldsymbol{N_{\mathrm{P}}}$		Number of patches

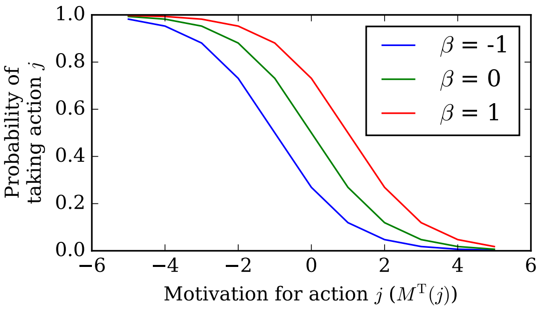


Figure 1: Probability to engage an individual as a function of the motivation (eq. 11). A higher tolerance motivation decreases the probability to engage the individual, while a higher aggressiveness increases the probability to engage.

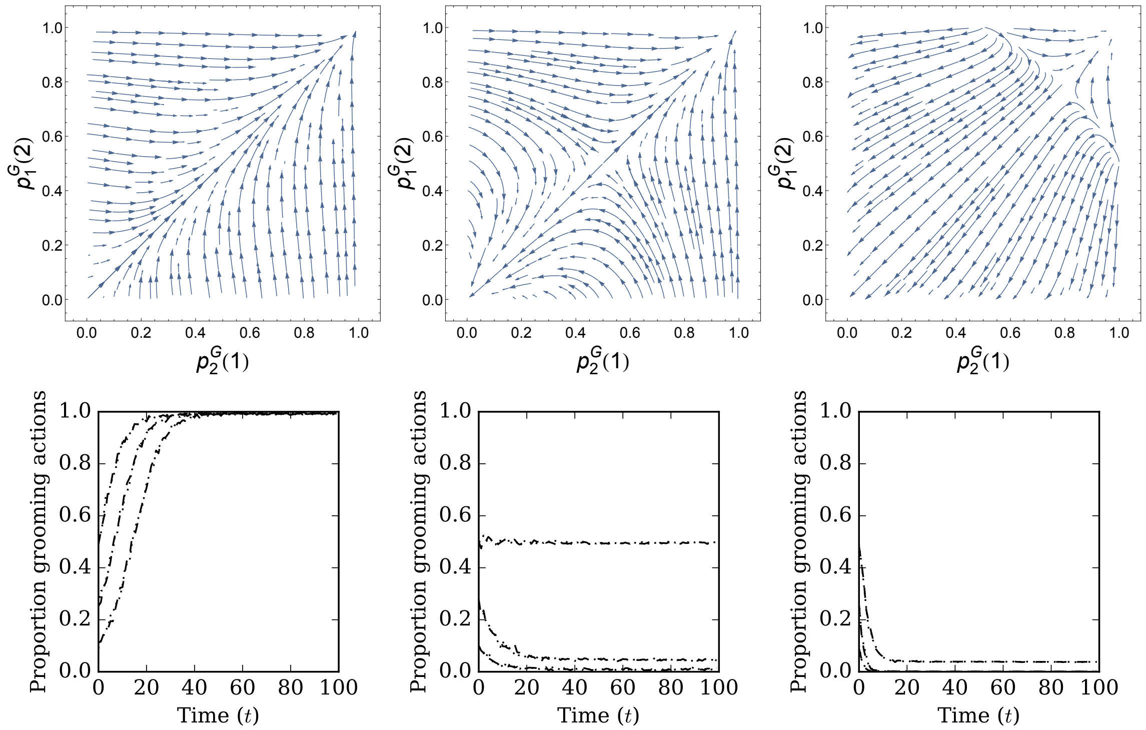


Figure 2: Solution orbits of the approximate learning dynamics given by eq. 1 of the SI (a refined version of eq. 19)and corresponding results obtained from individual-based simulations for the case: $c = 1, b = 3, \lambda = 0.25, \delta = 0.1, N = 2$. Upper panels: the solution orbits describe the direction of change of action choice and where the system will end up (either both partner groom each or they do not groom). Lower panels: individual-based simulations show the realizations of the (stochastic) model, averaged over 1000 replicates from three different initial conditions for the probability to groom at $t = 1$: $0.5, 0.25$, and 0.1. Left columns shows the results for $\gamma = 0.5$, middle column for $\gamma = 2$, and right column for $\gamma = 5$. Dashed line in the bottom row are player 1, and the dotted lines player 2. Note that for $\gamma = 2$ a single player does not groom half the time, but instead half the pairs converge on grooming only and the other half on non-grooming only.

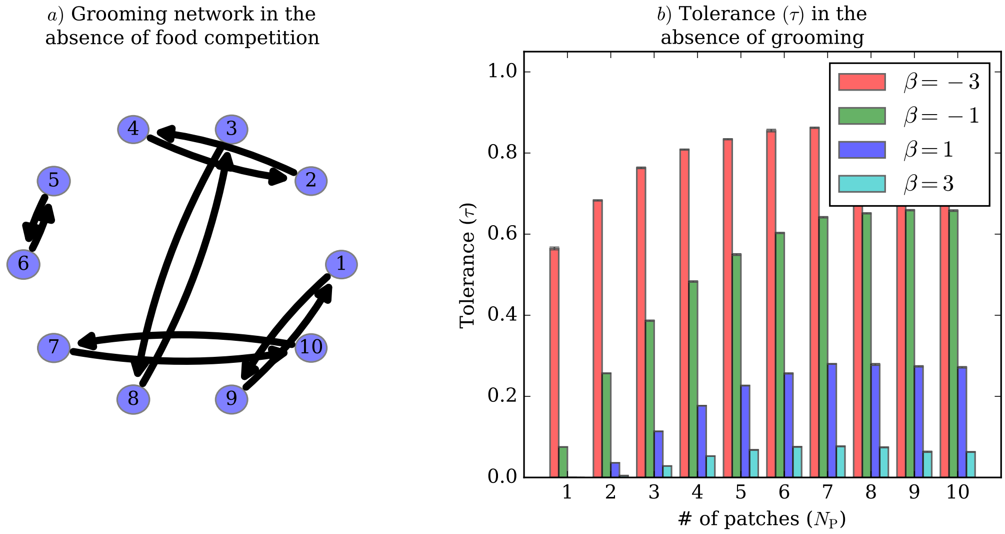
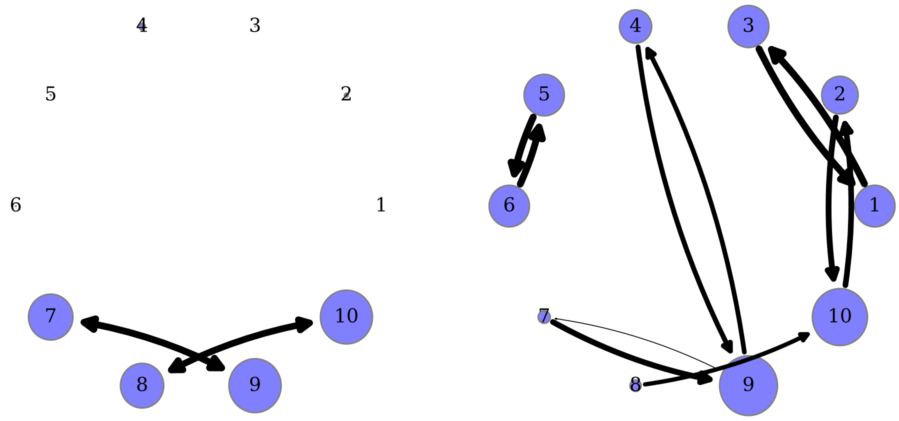
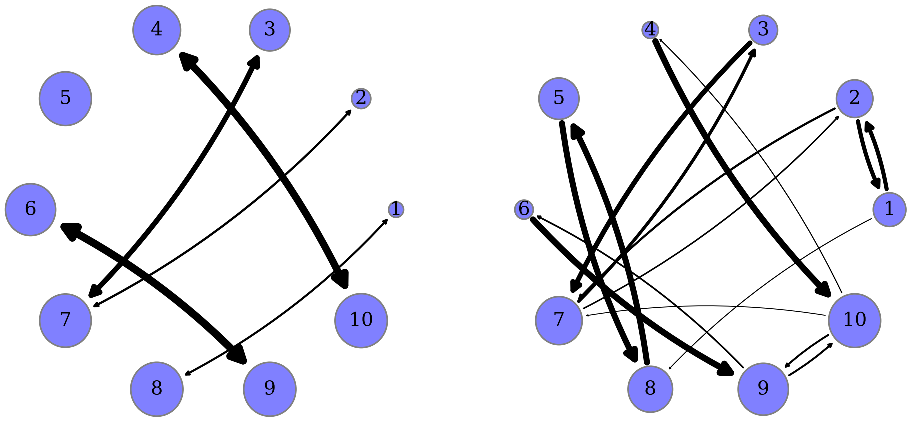


Figure 3: Panel a: typical grooming network when only the grooming stage occurs (no patch selection, aggressive interactions, and feeding). Vertex size correlates with the proportion of grooming it receives (g_i), while the edges correlate with the grooming i gives to j (g_{ij}). Individuals tend to form grooming pairs, where the members of the pair groom each other equally much. For clarity, grooming proportions below 0.10 ($g_{ij} < 0.10$) are not plotted. Panel b: tolerance (τ, mean and error bars) as a function of the number of patches, in the absence of grooming. Increasing β decreases the tolerance τ. Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \epsilon = 1, d = 2, q_0 = 6, m = 0.5, N = 10$, with patch quality given by eq. 3.



a) feeding associations

b) grooming



c) feeding associations

d) grooming

Figure 4: Feeding associations and grooming network for identical parameter values as in Fig. 3 and $\beta = 1$, but allowing for all interactions. Top row: $N_{\mathrm{P}} = 2$, bottom row: $N_{\mathrm{P}} = 5$. In the feeding association network the vertex size correlates with the frequency it feeds (f_i) while the edges correlate with the frequency i and j co-feed (a_{ij}). In the grooming network the vertex size correlates with the proportion of grooming it receives (g_i), while the edges correlate with the grooming i gives to j (g_{ij}). For clarity, associations and grooming proportions below 0.10 are not plotted (i.e., $a_{ij} < 0.10, g_{ij} < 0.10$). Note that individual 10 is the highest ranked individual. Individuals that feed together form an asymmetrical grooming relationship, while individuals that do not feed together form a symmetrical relationship (or they do not groom each other).

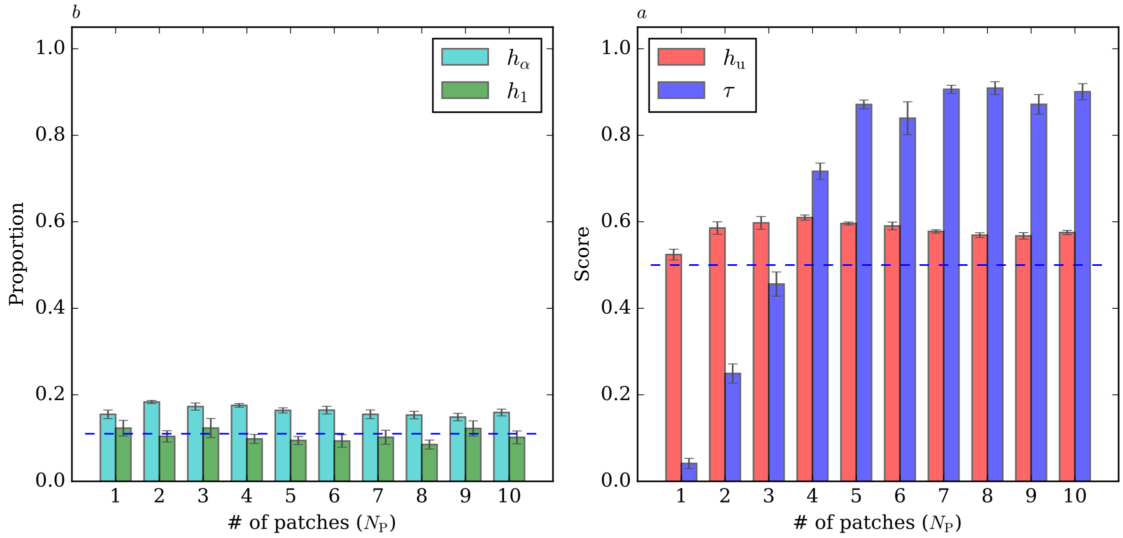


Figure 5: Mean and standard error (over 10 replicates) of the proportion of grooming directed towards the alpha (h_α) and towards the individual one rank up the hierarchy (h_1) (panel a) and the hierarchy score (h_u) and tolerance (τ) (panel b) as a function of the number of patches. Parameter values as in Fig. 4. The dashed line in panel a shows the expected proportions if grooming actions were chosen randomly. The dashed line in panel b shows $h_\mathrm{u} = 0.5$.

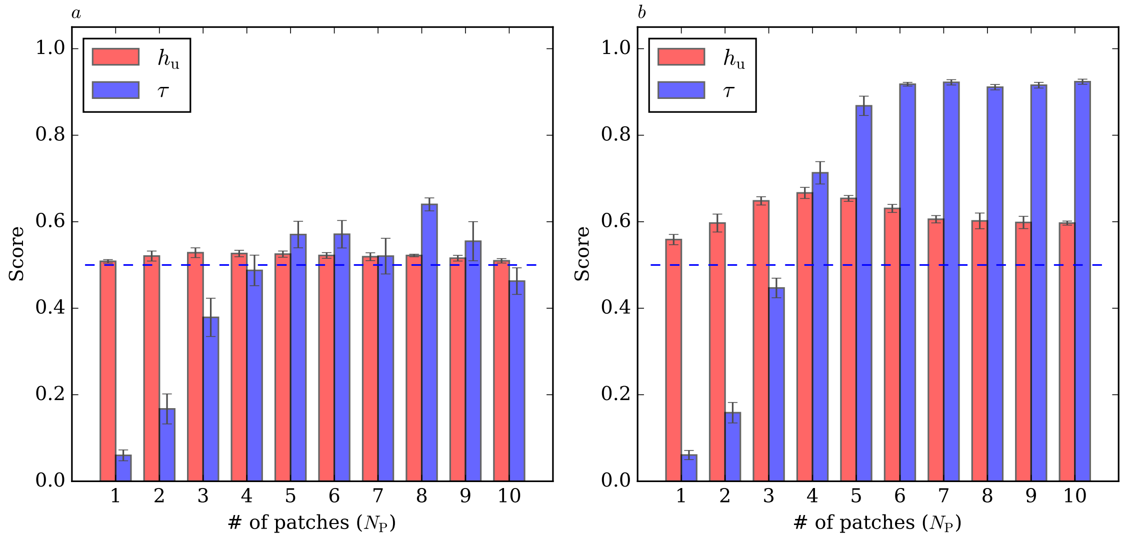


Figure 6: Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches. Parameter values as in Fig. 4. Panel a: $q_0 = 3$, panel b: $q_0 = 9$. The dashed line shows $h_\mathrm{u} = 0.5$.

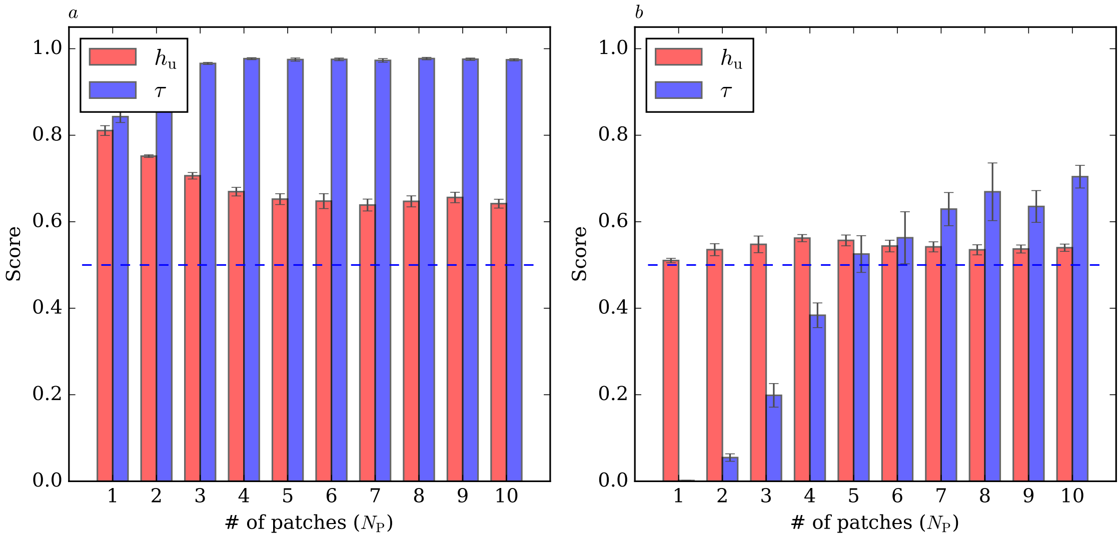


Figure 7: Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches. Parameter values as in Fig. 4. Panel a: $\beta = -3$, panel b: $\beta = 3$. The dashed line shows $h_\mathrm{u} = 0.5$.

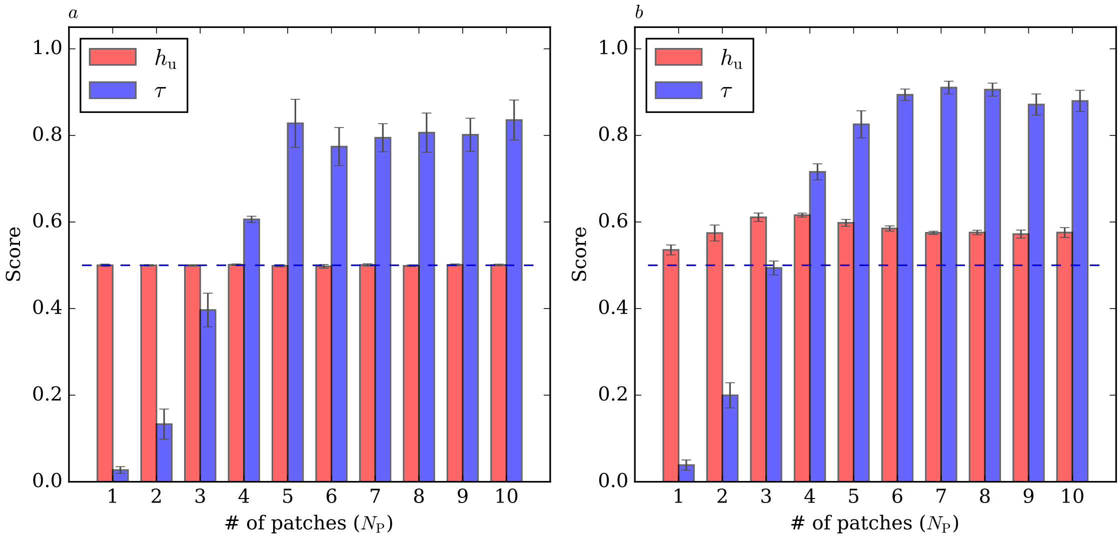


Figure 8: Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches. Parameter values as in Fig. 4. Panel a: $d = 0$, panel b: $d = 10$. The dashed line shows $h_\mathrm{u} = 0.5$.

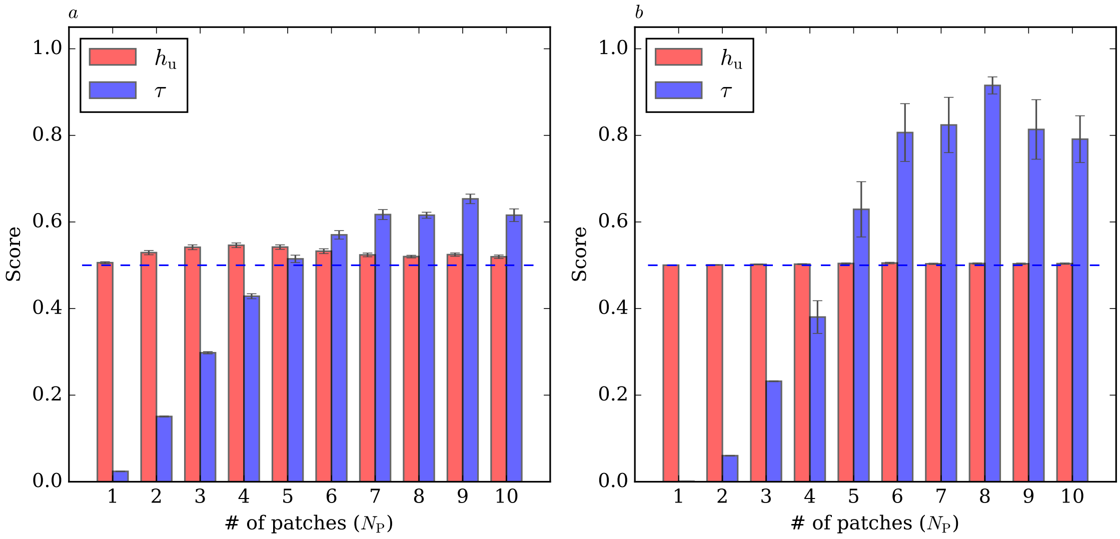


Figure 9: Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches. Parameter values as in Fig. 4. Panel a: $b = 1.5$, panel b: $b = 4.5$. The dashed line shows $h_u = 0.5$.

Supplementary material

In this supplementary material, we complement the analysis of the main text. First, we derive eq. 19. Second, we show that out choice regarding the number of time steps to reach the quasi-equilibrium is sufficient. Third, we show that our qualitative results regarding grooming up the hierarchy are robust for alternative assumptions regarding the patch quality, the number of individuals (N), and the intensity of competition modulated by the level of food wastage through interference (m). Finally, with a special version of the model we attempt to create conditions where individuals tend to groom one rank up the hierarchy, but we find that the grooming directed towards the alpha individual still outweighs the grooming one rank up the hierarchy ($h_\alpha > h_1$).

Stochastic approximation of learning dynamics

For the case $N = 2$ and assuming only grooming interactions (skipping stage 2, 3, and 4 of an interaction round), we can straightforwardly derive differential equations for the probabilities $p^{\mathrm{G}}_{1,t}(2)$ and $p^{\mathrm{G}}_{1,t}(2)$ that individual 1 grooms 2 and 2 grooms 1, respectively, by way of applying stochastic approximation as detailed in (Dridi and Lehmann, 2014, 2015). Indeed, using eqs. 5-6 of the main text (by setting $\pi^{\mathrm{F}}_{ij,t}$), we have by using eq. 11 and eq. 13 of (Dridi and Lehmann, 2014) that

$$\begin{aligned}
\frac{\mathrm{d}p^{\mathrm{G}}_{1,t}(2)}{\mathrm{d}t} &= \left(1 - p^{\mathrm{G}}_{1,t}(2)\right)\left(\delta \log\left(\frac{1 - p^{\mathrm{G}}_{1,t}(2)}{p^{\mathrm{G}}_{1,t}(2)}\right) + \lambda\left[p^{\mathrm{G}}_{2,t}(1)b - \left(p^{\mathrm{G}}_{1,t}(2)c + \left(1 - p^{\mathrm{G}}_{1,t}(2)\right)\gamma\right)\right]\right)\\
\frac{\mathrm{d}p^{\mathrm{G}}_{2,t}(1)}{\mathrm{d}t} &= \left(1 - p^{\mathrm{G}}_{2,t}(1)\right)\left(\delta \log\left(\frac{1 - p^{\mathrm{G}}_{2,t}(1)}{p^{\mathrm{G}}_{2,t}(1)}\right) + \lambda\left[p^{\mathrm{G}}_{1,t}(2)b - \left(p^{\mathrm{G}}_{1,t}(2)c + \left(1 - p^{\mathrm{G}}_{1,t}(2)\right)\gamma\right)\right]\right).
\end{aligned} \tag{1}$$

If we further assume that λ dominates δ ($\lambda \gg \delta$), then we can neglect the first terms in these equations (that account for randomness in choice) and simply write

$$\begin{aligned}
\frac{\mathrm{d}p^{\mathrm{G}}_{1,t}(2)}{\mathrm{d}t} &\sim \lambda\left(1 - p^{\mathrm{G}}_{1,t}(2)\right)\left[p^{\mathrm{G}}_{2,t}(1)b - \left(p^{\mathrm{G}}_{1,t}(2)c + \left(1 - p^{\mathrm{G}}_{1,t}(2)\right)\gamma\right)\right]\\
\frac{\mathrm{d}p^{\mathrm{G}}_{2,t}(1)}{\mathrm{d}t} &\sim \lambda\left(1 - p^{\mathrm{G}}_{2,t}(1)\right)\left[p^{\mathrm{G}}_{1,t}(2)b - \left(p^{\mathrm{G}}_{1,t}(2)c + \left(1 - p^{\mathrm{G}}_{1,t}(2)\right)\gamma\right)\right],
\end{aligned} \tag{2}$$

which is the equation presented in the main text. To compare to individual-based simulations (see Fig. 2 of the main text), we used eq. 1

Variance in the hierarchy score

In order to determine whether $T_{\mathrm{eq}} = 4000$ is a sufficient time period to reach the quasi-equilibrium, we investigated the how the standard deviation of our main statistic, the hierarchy score (h_{u}), varied over time. We find in general that the standard deviation slowly drops in roughly the first 500-2000 time steps, after which it continues to vary, but the pattern of variation remains stable (Fig. S1). Therefore, a period of $T_{\mathrm{eq}} = 4000$ is sufficient to reach the quasi-equilibrium.

Testing patch quality assumptions

We find that using different assumptions regarding patch quality does not have a qualitative effect on the outcome. We tested both the case where the patch quality is either high or low and where all patches are of equal patch quality (see section “Stage 4 - Feeding” from the main text for additional information on our assumptions regarding feeding). Despite alternative assumptions regarding quality, we continue to find that the hierarchy score (h_u) is non-monotonic in the number of patches (N_P), where the score is maximized for an intermediate number of patches, and decreases for higher or lower number of patches (Fig. S2). Additionally, for $N_\mathrm{P} = 10$ for when all patches are of equal quality we find that grooming is exactly reciprocal. Here, all individuals will maximize payoff when feeding alone on a patch, and thus there is no incentive to groom for tolerance.

Additional variables

Effect of group size

In order to determine the robustness of our results we varied the group size (N). We find we can qualitatively reproduce similar patterns for the grooming up the hierarchy score (h_u) for group sizes ranging from 5 to up to 100. However, for larger groups the aggressiveness needs to be lower (Fig. S3). For example, in order to maximize grooming up the hierarchy for $N_\mathrm{P} = 1$ with $N = 50$ we require $\beta = -6$, while for $N = 5$ we require $\beta = -3$. This is due to the extra competition that accompanies the extra individuals for $N = 50$. Therefore, larger groups need to be less aggressive in order to successfully trade grooming for tolerance. In general, however, the results presented in the main text appear to be robust for group size.

Effect of food competition due to m

Whereas the number of patches N_P determine the average number of other individuals a focal has to compete with on a food patch, the parameter m regulates the level of interference. That is, the larger m the more food is wasted when multiple individuals are feeding on the same patch. We find that increasing m results in less grooming up the hierarchy (h_u decreases, Fig. 4a, c). Reducing m has the same effect as reducing q_0, since it reduces the payoff of when individuals co-feed on a patch. If individuals gain less payoff when feeding on a patch, strong feeding associations (high a_{ij}) between individuals are less likely to occur, since individuals will vary their patch choice more. This results in low ranking individuals not grooming for tolerance. We find that increasing the average patch quality through q_0 counters the effect of increasing m (Fig. 4b, d).

Conditions favouring grooming one rank up the hierarchy

In an attempt to identify conditions where individuals are likely to groom one rank up the hierarchy, we performed several additional simulations where we (rather artificially) created conditions where we would expect such grooming to occur. We used a different equation to compute the payoff per individual on a patch here. These payoffs are shown in panel a and b of Fig. S5. As individual payoff drops sharply if more than two individuals feed on the same patch, such payoffs will “force” the group to feed in pairs on the available patches. In panel c simulations are done with the patch payoffs as in panel a. We find that the proportion of grooming directed to the alpha is still higher than one rank up the hierarchy. On closer inspection we find that the second highest ranked individual avoid feeding with the alpha here since its payoff is maximized by feeding on patch $k = 2$ and receiving extra grooming. If it would be feeding on the same patch as the alpha it would have to invest in grooming the alpha and thus, despite the higher payoff from feeding, would get a lower overall payoff. Increasing the payoff difference between patches does increase the proportion grooming that is directed one rank up the hierarchy (panel d, where payoffs are as in panel b). However, the proportion of grooming directed to the alpha still slightly outweighs the proportion of grooming one rank up the hierarchy. As such, our model cannot explain grooming of closely ranked individuals very well. Other factors not capture here may lead to such grooming to occur more frequently.

References

Dridi, S. and Lehmann, L. On learning dynamics underlying the evolution of learning rules. *Theoretical Population Biology*, 91:20–36, 2014.

Dridi, S. and Lehmann, L. A model for the evolution of reinforcement learning in fluctuating games. *Animal Behaviour*, 104:1–28, 2015.

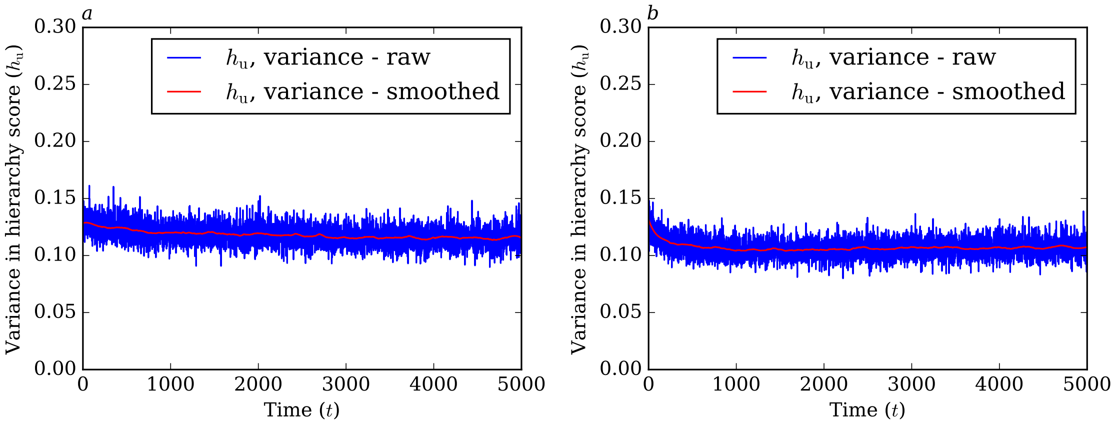


Figure S 1: The average hierarchy score (h_{u}) and its standard deviation (over 100 replicates) over time. Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \beta = 1, \epsilon = 1, d = 2, m = 0.5, N = 10$. Panel a: $N_{\mathrm{P}} = 2$, panel b: $N_{\mathrm{P}} = 5$.

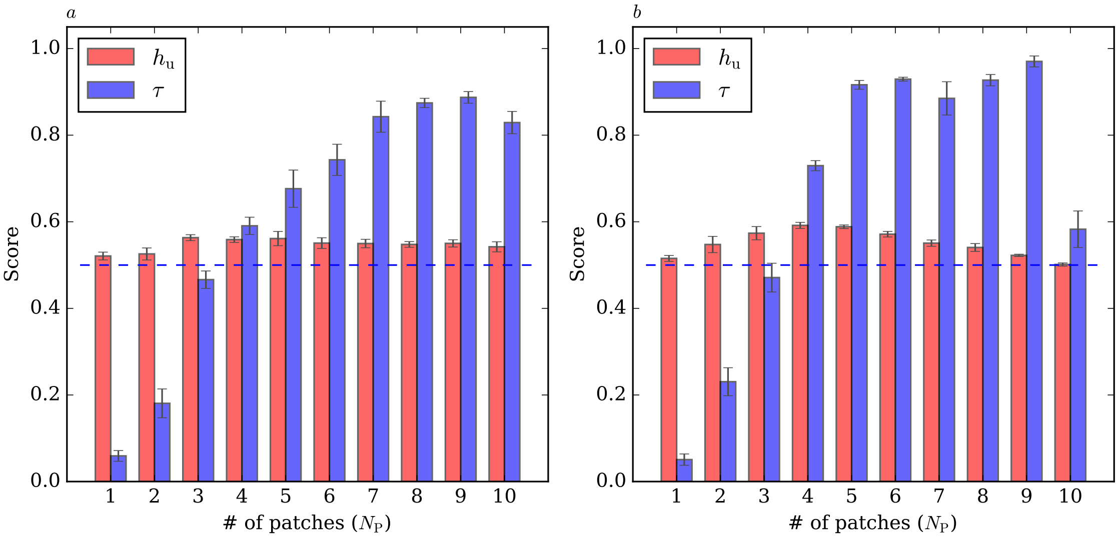


Figure S 2: Mean and standard error (over 10 replicates) of the hierarchy score (h_{u}) and tolerance (τ) as a function of the number of patches. Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \beta = 1, \epsilon = 1, d = 2, q_0 = 4, m = 0.5, N = 10$, panel a: high/low patch quality scenario, panel b: equal patch quality scenario. The dashed line shows $h_{\mathrm{u}} = 0.5$.

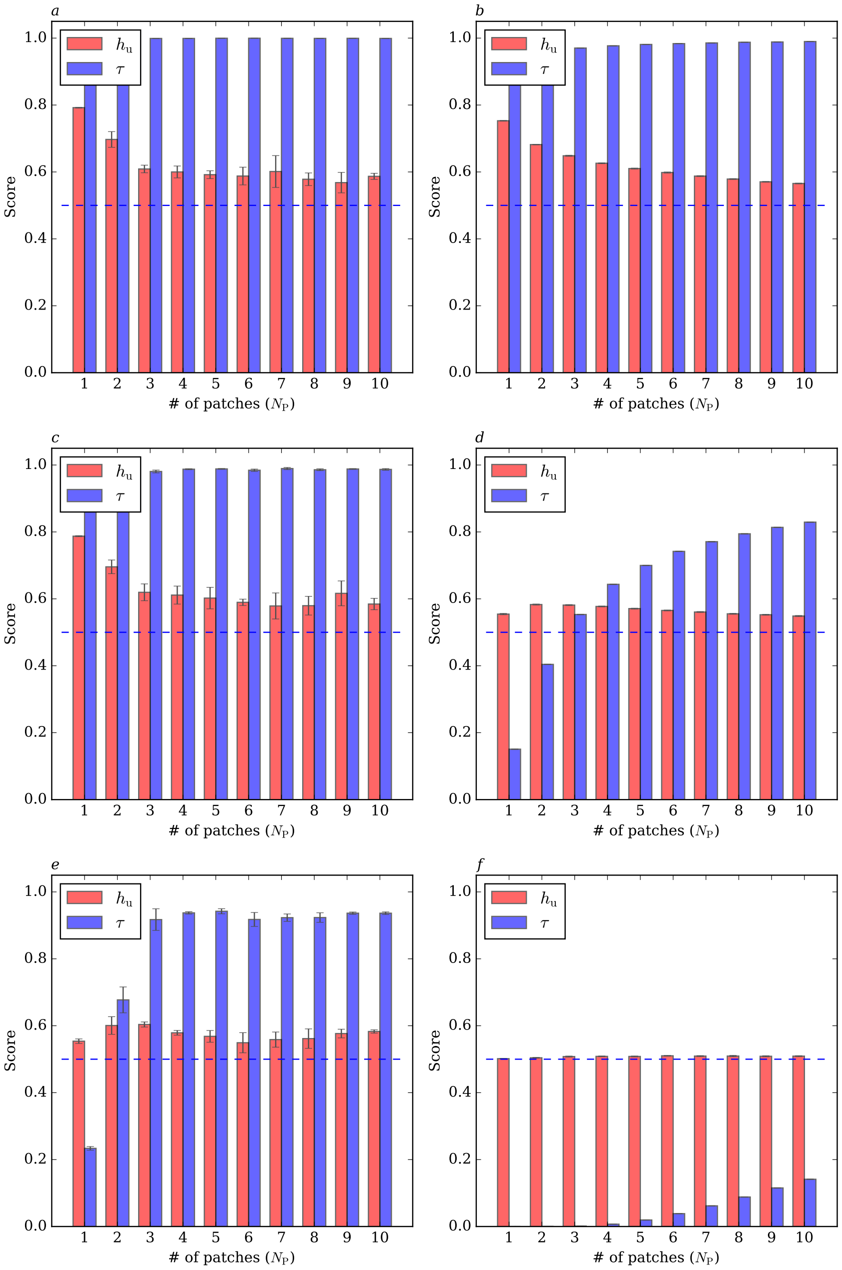


Figure S 3: Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches. Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \epsilon = 1, d = 2, q_0 = 6, m = 0.5$, with patch quality given by eq. 3. Panel a: $N = 5, \beta = -6$, panel b: $N = 50, \beta = -6$, panel c: $N = 5, \beta = -3$, panel d: $N = 50, \beta = -3$, panel e: $N = 5, \beta = 0$, panel f: $N = 50, \beta = 0$. The dashed line shows $h_\mathrm{u} = 0.5$.

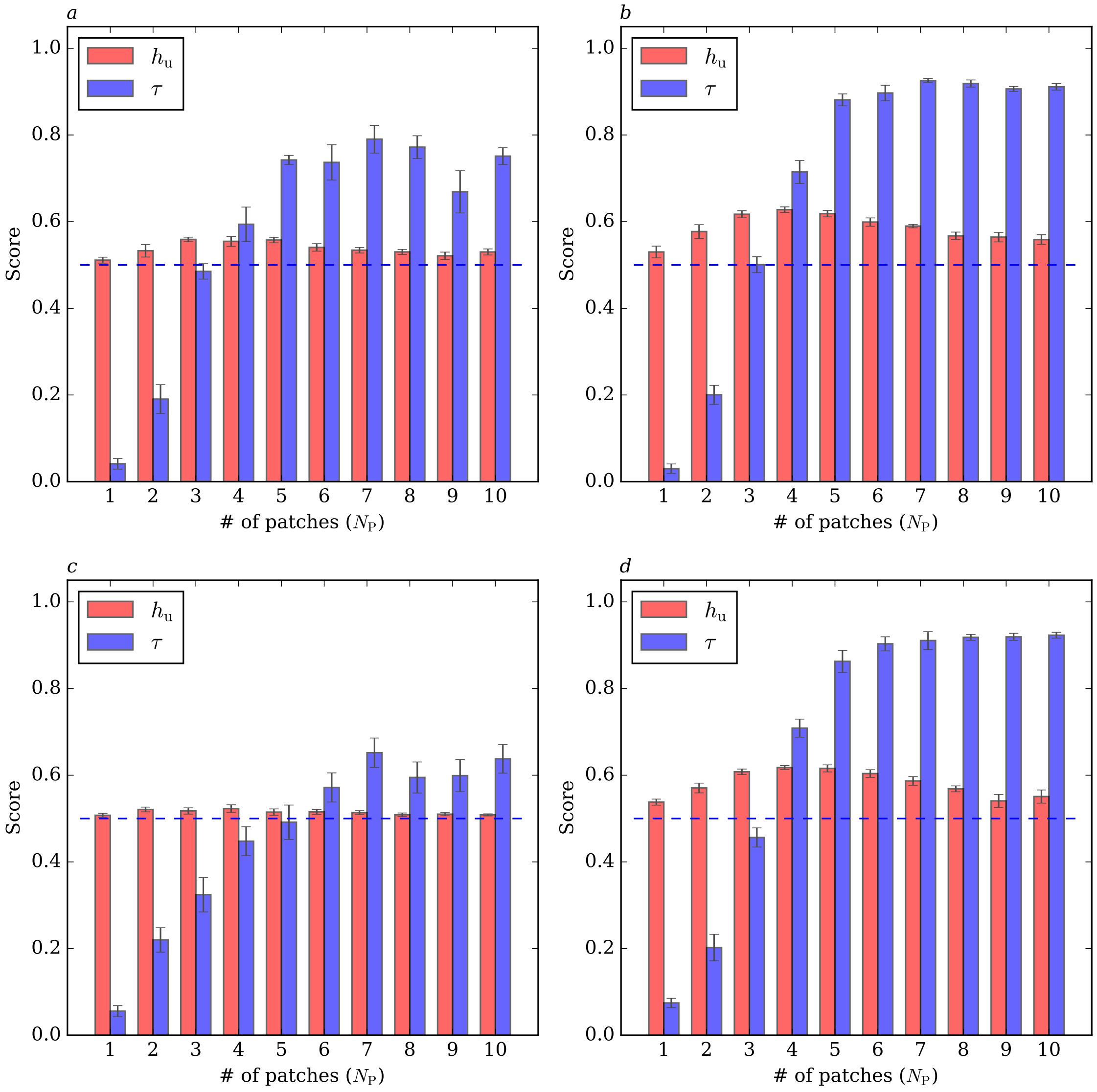


Figure S 4: Mean and standard error (over 10 replicates) of the hierarchy score (h_{u}) and tolerance (τ) as a function of the number of patches. Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \beta = 1, \epsilon = 1, d = 2$, with patch quality given by eq. 3. Panel a: $m = 1, q_0 = 6$, panel b: $m = 1, q_0 = 10$, panel c: $m = 1.5, q_0 = 6$, panel d: $m = 1.5, q_0 = 14$. Whereas increasing m makes grooming more reciprocal ($h_{\mathrm{u}} \to 0.5$), increasing q_0 cancels the effect. The dashed line shows $h_{\mathrm{u}} = 0.5$.

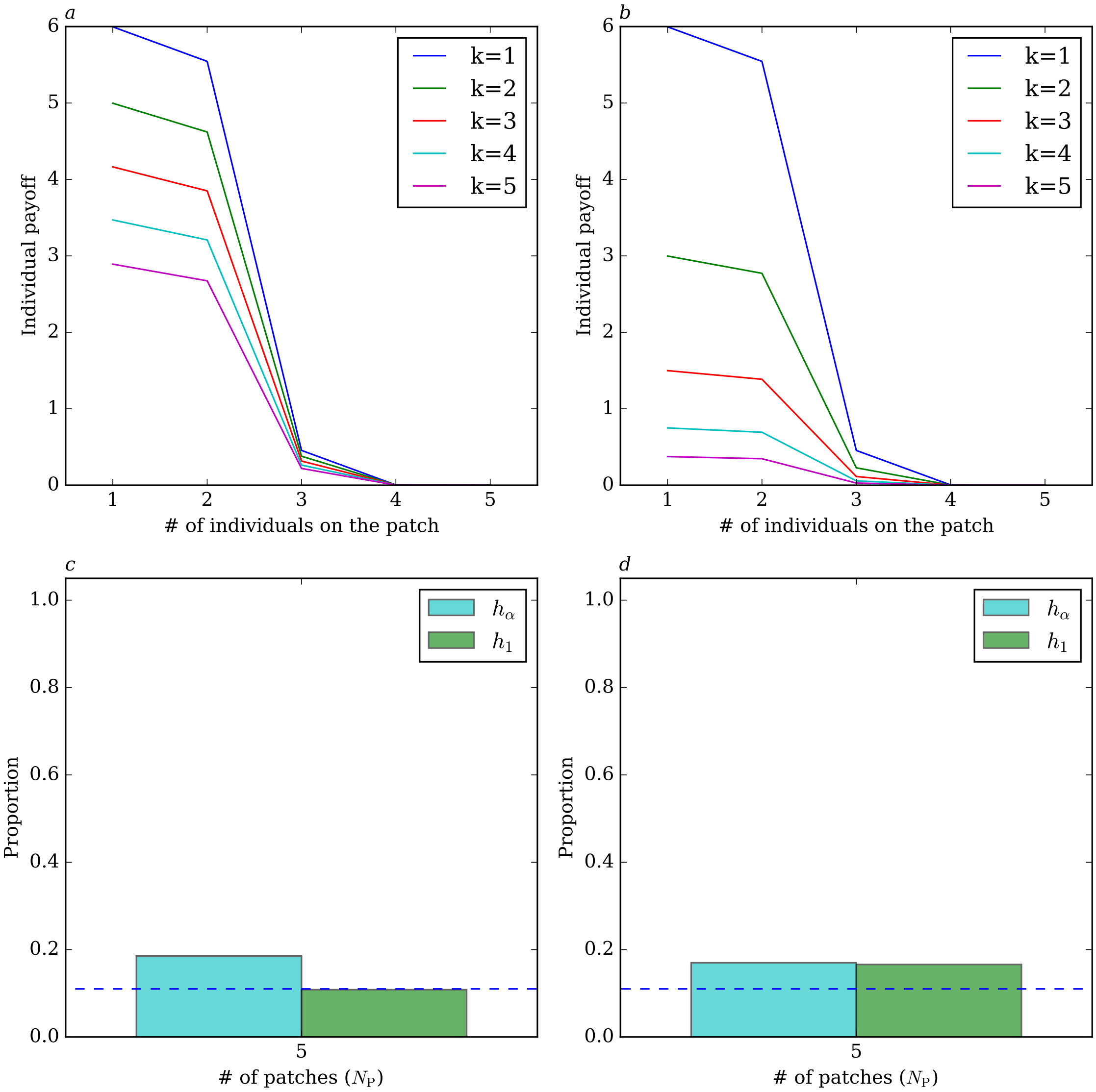


Figure S 5: Top row: payoff per individuals as a function of the number of individuals on a patch, for all five patches. Bottom row: proportion grooming alpha and proportion grooming one rank up the hierarchy (mean over 10 replicates). Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \beta = 1, \epsilon = 1, d = 2, m = 0.5, N = 10, N_\mathrm{P} = 5$.

General discussion

Social behaviour pervades all layers of life (Ben-Jacob, 2008; Poulsen and Boomsma, 2005; Schwinning and Weiner, 1998; Clutton-Brock, 2002). Understanding how organisms in varying environments navigate their behaviour in repeated social interactions is a major area of research (Maynard Smith, 1979; Axelrod and Hamilton, 1981; Riechert and Hammerstein, 1983; Camerer and Hua Ho, 1999). In this thesis I contributed to the understanding of repeated interactions by investigating two main questions. First I asked: "Under which conditions will positive reciprocity, partner switching, or punishment be dominant in a population if these partner control mechanisms coevolve?". I developed two agent-based models in the first two chapters to address this question. In chapter three of the thesis, I specifically investigated the grooming interactions among primates by addressing the question: "What are the conditions that cause grooming to be directed up the hierarchy?".

Coevolution of partner control mechanisms

The focus of the first chapter was to characterize the conditions where positive reciprocity, punishment, or partner switching is the dominant partner control mechanism to stabilize cooperation in a population of unrelated individuals. This study showed that the likelihood of partner switching emerging as the dominant partner control mechanism in a well-mixed population (meaning individuals are equally likely to interact with any other member of the population) increases with increasing number of rounds of interaction, while fewer number of rounds favoured either positive reciprocity or punishment (depending on the precise tuning of the model parameters). If interactions between individuals are, however, localized to groups then punishment is more likely to be favoured by selection compared to the well-mixed case.

There is a long list of studies investigating the conditions where one of these partner control mechanisms is favoured by selection over uncooperative strategies (e.g., Axelrod and Hamilton, 1981; Boyd and Richerson, 1988; Nakamaru and Iwasa, 2006; Powers et al., 2012; Joyce et al., 2006; Izquierdo et al., 2014). While the aim of these studies is to show that the specific mechanism can indeed favour the evolution of cooperation for the given conditions, inferences regarding the likelihood of the mechanism to be observed in natural populations cannot as easily be made. For example, demonstrating that the tit-for-tat strategy outperforms a variety of other positive reciprocity strategies in well-mixed populations with many rounds of interactions (Axelrod and Hamilton, 1981), does not necessarily mean that this strategy should be observed under natural conditions. Previous studies, as well as the first chapter of this thesis, show that under such conditions partner switching is much more likely to be the dominant partner control mechanism (Joyce et al., 2006; Izquierdo et al., 2010). Studies such as these demonstrate that the question to be asked should not be "What is the range of conditions where this particular mechanism can stabilize cooperation in a population?", but instead have a form like "Under these conditions, which mechanism is likely to evolve?". Chapter one is the first to compare the partner control mechanisms positive reciprocity, punishment, and partner switching in a single model, and as such its insights should guide future studies into developing more relevant models.

The second chapter is a continuation of the first that, for most species, increased the realism of the model by adding overlapping generations, limited migration of offspring between demes, and a cost correlating to the complexity of strategies to the model. The results of the second chapter give several valuable insights. Firstly, since the second chapter makes a direct comparison with the first, it clearly demonstrates how various assumptions can affect the competitiveness of each partner control mechanism. It was shown that partner switching benefits from regular changes in the group composition due to generational overlap as it increases the number of potential partners, while punishment is most strongly affected by the cost for strategic complexity. Secondly, the second chapter allowed to compare the competitiveness of unconditional strategies versus conditional ones. Although for both conditional and unconditional strategies conditions were identified that favoured one type over the other, conditional strategies were more likely to be favoured by selection for most of the parameter space.

Grooming up the hierarchy

In the third chapter I investigated how the intensity of food competition and other factors (such as aggressiveness) affects grooming patterns in primates: in particular, how much grooming is directed up the hierarchy. The results showed that if competition is low (i.e., there are many patches to feed on) individuals will spread out over the available patches and since they mostly do not compete over the same food source, there is little grooming up the hierarchy. If competition increases, then pairs of individuals will more often find themselves feeding on the same patch, where the lower ranked individual will groom the higher ranked one in order to be tolerated. For even higher competition, grooming up the hierarchy does not necessarily increase. Under conditions of intense food competition, multiple individuals may attempt to feed on the same patch. Here, a low ranked individual is unlikely to be tolerated by all higher ranking ones, and thus will almost always be evicted. As it will also not groom for tolerance then, overall the grooming pattern will become more reciprocal.

The model is a major step forward in primatology. Up until now, the only quantitative models to investigate grooming up the hierarchy were done by Seyfarth and Sambrook et al, the latter being an extension of Seyfarth's model (Seyfarth, 1977; Sambrook et al., 1995). However, in these models the individuals have fixed preferences on whom to groom, and the individuals would therefore not be able to adjust to a changing environment (i.e., in the absence of food competition the individuals would continue to groom up the hierarchy). By making the motivations to groom other individuals variable in the model in chapter three, allowed individuals to adjust to the state of the environment. The model replicated the main finding by Seyfarth in that grooming was found to be directed up the hierarchy under conditions where high ranked individuals can provide alternative benefits to trade for grooming (i.e., tolerance). Additionally, the absence of food competition resulted in individuals trading grooming reciprocally, a pattern that is not obtainable in Seyfarth's model.

The grooming model formalized several theories and ideas that are concerned with social relationships in primates. First, the so-called socioecological model describes how factors such as food competition determine social relationships (van Schaik, 1989; Sterck et al., 1997; Kappeler and van Schaik, 2002). As the model is verbal, no quantitative predictions can be derived from it. The grooming model developed in the third chapter, however, makes this link explicit and demonstrates how the intensity of competition (due to the number of available food patches) determines the grooming up the hierarchy. Second, one of the hypotheses on how primates make decision is by emotional bookkeeping (Aureli and Schaffner, 2002; Schino and Aureli, 2009, 2016). The idea is that a primate has partner specific emotions that are updated through the various interactions it has with others. While there is some empirical evidence supporting the hypothesis, the mechanics of it remained obscure. In the third chapter these mechanics are captured in clear and simple equations, describing precisely how the payoffs from different interactions can be translated into a single motivational state or emotion. The model shows stable grooming bonds emerging over longer periods of time as predicted by emotional bookkeeping.

Unification of both approaches

The two approaches to study repeated interactions in this thesis differ greatly from each other. The models in the first two chapters simulated the evolution of a population where individuals have different strategies. The model in the third chapter simulated the interactions in a group of primates where all individuals had a single strategy: a learning rule. While the models all had merit in their own right, increasingly valuable insights can be obtained by combining different elements of both evolutionary and learning models.

The strategies used in chapter one and two were pure (prescribing a specific action for each possible situation) and limited to a single round of memory. Such a restriction was needed in order to reduce the size of the set of strategies and keep the model tractable. However, an arguably more realistic way of modelling the strategies would be to implement learning rules, similar to those used in the grooming model. It is frequently argued that most interesting behaviours are flexible and a variety of models have investigated the evolution of learning rules (Harley, 1981; Fawcett et al., 2013; Dridi and Lehmann, 2014, 2015). Additionally, from a neuroscientific point of view learning rules appear to be more easily linked to the neuronal decision making (Dayan and Abbott, 2005; Niv, 2009). Although frequently the interest might be purely in the expressed behaviour, without considerations of the mechanisms that produce the behaviour, if one is to understand the behaviour that organisms express in repeated interactions, then eventually the underlying mechanisms need to be studied. In the first two chapters the main interest was in which form of partner control is most likely to dominate under various conditions, and in order to simplify the models, using "simple" pure strategies were used to investigate the question. However, an important next step would be to try and implement a variety of learning rules in order to approximate better the decision making process.

A second aspect of the grooming model that is worthy to incorporate into the evolutionary models is that of asymmetry. In the grooming model the individuals were asymmetric in the sense that the probability of winning a contest depended on the ranks of both individuals engaged in the contest. In nature, organisms are bound to be asymmetric in some respect. This asymmetry can be in the size of the individuals, the fighting abilities, in the ownership of a territory, or any other trait. A range of models, mostly concerned with animal contests, have described how asymmetries can affect the interactions between the individuals and determine the outcome (Maynard Smith and Parker, 1976; Selten, 1980; Hammerstein, 1981; Enquist and Leimar, 1983; Leimar and Enquist, 1984; Enquist and Leimar, 1987, 1990; Samuelson and Zhang, 1992; Binmore and Samuelson, 2001). However, studies on the evolution of cooperation and helping behaviours tend to neglect asymmetries in their models (Boyd et al., 2003; Santos et al., 2011; Shutters, 2011; García and Traulsen, 2012). However, punishment is often directed from a dominant individual towards a subordinate one (Clutton-Brock and Parker, 1995), since if a low ranked or weaker individual punishes a higher ranked or stronger one, the punishment will likely not have the desired effect, and the weaker individual is likely to refrain from punishing the other. For example, in pairs of cleaner fish it is the larger male that chases (punishes) the female and not the other way around (Raihani et al., 2010).

Asymmetries can thus potentially determine the likelihood of successful punishment: i.e., the recipient changing its behaviour. Therefore, incorporating asymmetries is an important next step in the study on the evolution of punishment, cooperation, and helping behaviours. In the models of chapter one and two individuals were assumed to be symmetrical, and thus a logical extension of those models would be to incorporate a form of asymmetry in order to investigate how this may affect the competitiveness of punishment as a partner control mechanism.

Conversely, an interesting extension of the grooming model would be to allow various traits to evolve. The main goal of this model was to investigate how food competition influences grooming patterns, and therefore the model was set up to be a proximate one. However, this means that various traits that are likely to vary between individuals were instead assumed to be constant. Aggression, for example, varies between primate species, populations, and individuals (Isbell, 1991; Fairbanks et al., 2004), and it is not unlikely that levels of aggression evolve due to the distribution and quality of food (determining the intensity of food competition) that a population experiences, since much of the aggression in primates is related to food competition (Janson and van Schaik, 1988). It would therefore be interesting to let aggression evolve, in order to determine how the intensity of food competition may result in selection of the level of aggressiveness (Vogel and Janson, 2011). Furthermore, while we assumed fixed parameters as to how motivations were updated as a result of the interactions, these may similarly evolve. For example, receiving grooming was assumed to have a fixed effect on the motivation as well as winning or losing a contest. However, winning a contest over a food source may be especially important if food sources are scarce, and selection may therefore act on such traits. While investigating the evolution of these traits in concert with the proximate effects of food competition was outside the scope of this thesis, in future research these are interesting questions to address.

Although in chapter two we investigated how interactions between related individuals may affect the evolution of partner control mechanisms, in the grooming model we made the simplifying assumption that all individuals were unrelated. Primate species are often characterized by a philopatric sex and a dispersing one, where is most cases the females are philopatric (Pusey and Packer, 1987). This obviously results in interactions between related individuals. Additionally, female philopatry often leads to the formation of a matrilineal society, such as in macaques, baboons, and vervets (Silk, 2002). Here, daughters of the matriline tend to inherit the dominance rank of the mother and related individuals regularly support one another in agonistic interactions (reviewed in Silk, 2002). Additionally, a meta-analysis showed that primates have a strong preference to groom their kin (Schino, 2001). Therefore, both grooming motivations and the outcome of agonistic interactions are likely to depend strongly on the presence of kin. In chapter three the main interest was in the effect of food competition on grooming behaviour, and therefore individuals were assumed to be unrelated. However, introducing such a structure would be an interesting extension, since other meta-analyses showed significant effects for both grooming reciprocity and grooming for rank-related benefits even when controlling for kinship (Schino, 2007; Schino and Aureli, 2008). Further models are thus required to understand how the presence of kin will affect a grooming network, when it is better to support or tolerate kin versus non-kin, and how the state of the environment influences the tendency to associate with kin or not.

Empirically informed models

The thesis originally set out to develop empirically informed models, i.e., models that are strongly based on empirical evidence. Although throughout the development of all models this was always kept in mind, it is easier said than done. The main empirical evidence on which the models in chapters one and two are based, is that in pairwise interactions in nature different partner control mechanisms may be used to stabilize cooperation. However, most models attempting to explain cooperation, tend to focus on a single control mechanism. These models therefore cannot explain the occurrence of a partner control mechanism in nature as other partner control mechanisms may outcompete the one that is tested in their model. Despite the comparison of partner control mechanisms in the current work, the models themselves still required a variety of simplifying assumptions that are "less" empirically informed, such as pure strategies. These simplifying assumptions are necessary in order to keep the model tractable. Finding the right balance is a major challenge for the development of these models.

An additional challenge was identified during the development of the grooming model. Here, much time was consumed in order to establish a set of learning rules that match primate behaviour. Understanding precisely how a primate takes its actions is a formidable task. Despite the impressive amount of primate data that has been generated, the data does not inform how an individual will change its probability to groom or engage another individual as a consequence of co-feeding with it. In other words, despite the wealth of data, translating this to decision rules is a major challenge. However, after many lengthy discussions a coherent set of equations emerged, which (as discussed in chapter three) appear to resemble the observations well.

Concluding remarks

The study on repeated social interactions is a broad topic, as exemplified by the current work. Both questions that are addressed here have explored a relatively unexplored area in their respective fields of study. Using agent-based models allowed me to characterize the conditions where specific partner control mechanisms are likely to dominate in a population engaged in pairwise interactions. Such insights are crucial in understanding the presence of cooperation and helping behaviours in nature. Additionally, I have shown how food competition and other factors of primate life influence what grooming patterns might emerge, and how varying parameters of the model can capture the variation in grooming patterns in primate populations.

References

Aureli, F. and Schaffner, C. Relationship assessment through emotional mediation. *Behaviour*, 139(2):393–420, 2002.

Axelrod, R. and Hamilton, W. D. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

Ben-Jacob, E. Social behavior of bacteria: from physics to complex organization. 65(3):315–322, 2008.

Binmore, K. and Samuelson, L. Can Mixed Strategies be Stable in Asymmetric Games? *Journal of Theoretical Biology*, 210(1):1–14, 2001.

Boyd, R. and Richerson, P. J. The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132(3):337–356, 1988.

Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, 100(6):3531–3535, 2003.

Camerer, C. and Hua Ho, T. Experience-weighted Attraction Learning in Normal Form Games. *Econometrica*, 67(4):827–874, 1999.

Clutton-Brock, T. H. and Parker, G. A. Punishment in animal societies. *Nature*, 373(6511): 209–216, 1995.

Clutton-Brock, T. Breeding Together: Kin Selection and Mutualism in Cooperative Vertebrates. *Science*, 296(5565):69–72, 2002.

Dayan, P. and Abbott, L. F. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT Press, 2005.

Dridi, S. and Lehmann, L. On learning dynamics underlying the evolution of learning rules. *Theoretical Population Biology*, 91:20–36, 2014.

Dridi, S. and Lehmann, L. A model for the evolution of reinforcement learning in fluctuating games. *Animal Behaviour*, 104:1–28, 2015.

Enquist, M. and Leimar, O. Evolution of fighting behaviour: Decision rules and assessment of relative strength. *Journal of Theoretical Biology*, 102(3):387–410, 1983.

Enquist, M. and Leimar, O. Evolution of fighting behaviour: The effect of variation in resource value. *Journal of Theoretical Biology*, 127(2):187–205, 1987.

Enquist, M. and Leimar, O. The evolution of fatal fighting. *Animal Behaviour*, 39(1):1–9, 1990.

Fairbanks, L. A., Newman, T. K., Bailey, J. N., Jorgensen, M. J., Breidenthal, S. E., Ophoff, R. A., Comuzzie, A. G., Martin, L. J., and Rogers, J. Genetic contributions to social impulsivity and aggressiveness in vervet monkeys. *Biological Psychiatry*, 55(6):642–647, 2004.

Fawcett, T. W., Hamblin, S., and Giraldeau, L.-a. Exposing the behavioral gambit: the evolution of learning and decision rules. *Behavioral Ecology*, 24(1):2–11, 2013.

García, J. and Traulsen, A. Leaving the loners alone: Evolution of cooperation in the presence of antisocial punishment. *Journal of Theoretical Biology*, 307:168–173, 2012.

Hammerstein, P. The role of asymmetries in animal contests. *Animal Behaviour*, 29(1):193–205, 1981.

Harley, C. B. Learning the evolutionarily stable strategy. *Journal of Theoretical Biology*, 89(4): 611–633, 1981.

Isbell, L. A. Contest and scramble competition: patterns of female aggression and ranging behavior among primates. *Behavioral Ecology*, 2(2):143–155, 1991.

Izquierdo, L. R., Izquierdo, S. S., and Vega-Redondo, F. Leave and let leave: A sufficient condition to explain the evolutionary emergence of cooperation. *Journal of Economic Dynamics and Control*, 46:91–113, 2014.

Izquierdo, S. S., Izquierdo, L. R., and Vega-Redondo, F. The option to leave: Conditional dissociation in the evolution of cooperation. *Journal of Theoretical Biology*, 267(1):76–84, 2010.

Janson, C. H. and van Schaik, C. P. Recognizing the Many Faces of Primate Food Competition: Methods. *Behaviour*, 105(1/2):165–186, 1988.

Joyce, D., Kennison, J., Densmore, O., Guerin, S., Barr, S., Charles, E., and Thompson, N. S. My Way or the Highway: a More Naturalistic Model of Altruism Tested in an Iterative Prisoners' Dilemma. *Journal of Artificial Societies & Social Simulation*, 9(2), 2006.

Kappeler, P. M. and van Schaik, C. P. Evolution of Primate Social Systems. *International Journal of Primatology*, V23(4):707–740, 2002.

Leimar, O. and Enquist, M. Effects of asymmetries in owner-intruder conflicts. *Journal of Theoretical Biology*, 111(3):475–491, 1984.

Maynard Smith, J. Game Theory and the Evolution of Behaviour. *Proceedings of the Royal Society of London B: Biological Sciences*, 205(1161):475–488, 1979.

Maynard Smith, J. and Parker, G. A. The logic of asymmetric contests. *Animal Behaviour*, 24(1): 159–175, 1976.

Nakamaru, M. and Iwasa, Y. The coevolution of altruism and punishment: Role of the selfish punisher. *Journal of Theoretical Biology*, 240(3):475–488, 2006.

Niv, Y. Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154, 2009.

Poulsen, M. and Boomsma, J. J. Mutualistic Fungi Control Crop Diversity in Fungus-Growing Ants. *Science*, 307(5710):741–744, 2005.

Powers, S. T., Taylor, D. J., and Bryson, J. J. Punishment can promote defection in group-structured populations. *Journal of Theoretical Biology*, 311:107–116, 2012.

Pusey, A. E. and Packer, C. Dispersal and philopatry. In Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W., and Struhsaker, T. T., editors, *Primate Societies*, pages 250–266. Chicago University Press, Chicago, 1987.

Raihani, N. J., Grutter, A. S., and Bshary, R. Punishers Benefit From Third-Party Punishment in Fish. *Science*, 327(5962):171, 2010.

Riechert, S. E. and Hammerstein, P. Game Theory in the Ecological Context. *Annual Review of Ecology and Systematics*, 14(1):377–409, 1983.

Sambrook, T. D., Whiten, A., and Strum, S. C. Priority of access and grooming patterns of females in a large and a small group of olive baboons. *Animal Behaviour*, 50(6):1667–1682, 1995.

Samuelson, L. and Zhang, J. Evolutionary stability in asymmetric games. *Journal of Economic Theory*, 57(2):363–391, 1992.

Santos, M. D., Rankin, D. J., and Wedekind, C. The evolution of punishment through reputation. *Proceedings of the Royal Society B: Biological Sciences*, 278(1704):371–377, 2011.

Schino, G. Grooming, competition and social rank among female primates: a meta-analysis. *Animal Behaviour*, 62(2):265–271, 2001.

Schino, G. Grooming and agonistic support: a meta-analysis of primate reciprocal altruism. *Behavioral Ecology*, 18(1):115–120, 2007.

Schino, G. and Aureli, F. Grooming reciprocation among female primates: a meta-analysis. *Biology Letters*, 4(1):9–11, 2008.

Schino, G. and Aureli, F. *Chapter 2 Reciprocal Altruism in Primates*, volume 39, pages 45–69. Elsevier, 2009.

Schino, G. and Aureli, F. Reciprocity in group-living animals: partner control versus partner choice. *Biol Rev*, page n/a, 2016.

Schwinning, S. and Weiner, J. Mechanisms determining the degree of size asymmetry in competition among plants. 113(4):447–455, 1998.

Selten, R. A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology*, 84(1):93–101, 1980.

Seyfarth, R. M. A model of social grooming among adult female monkeys. *J Theor Biol*, 65(4): 671–698, 1977.

Shutters, S. T. Punishment Leads to Cooperative Behavior in Structured Societies. *Evolutionary Computation*, 20(2):301–319, 2011.

Silk, J. B. Kin selection in primate groups. *International Journal of Primatology*, 23(4):849–875, 2002.

Sterck, E. H. M., Watts, D. P., and van Schaik, C. P. The evolution of female social relationships in nonhuman primates. *Behavioral Ecology and Sociobiology*, 41(5):291–309, 1997.

van Schaik, C. P. The ecology of social relationships amongst female primates. In Standen, V. and Foley, R. A., editors, *Comparative socioecology. The behavioural ecology of humans and other mammals.*, pages 195–218. Blackwell, Oxford, 1989.

Vogel, E. and Janson, C. Quantifying Primate Food Distribution and Abundance for Socioecological Studies: An Objective Consumer-centered Method. *International Journal of Primatology*, pages 1–18, 2011.

Acknowledgements

First of all, I would like to thank my supervisors, Redouan Bshary and Laurent Lehmann, for their great support throughout my PhD, and guiding my development as a researcher. The Swiss National Science Foundation funded this project, while I was hosted by the universities of Neuchâtel and Lausanne. I would also like to thank Klaus Zuberbühler and Peter Hammerstein for reviewing my thesis and interesting discussions. I thank the colleagues from the ProDoc, their workshops, and the programme in general. I thank the people from the Inkawu Vervet Project for two interesting months in the field.

I would like to thank my colleagues from Neuchâtel for the great time spent together: Mélisane Aellen, Sandra Binning, Christèle Borgeaud, Felice Di Lascio, Teresa Fagundes, Miguel Gareta, Simon Gingins, Renata Mazzei, Sharon McAfoose, Jennifer McClung, Rute Mendonça, Ana Pinto, Andrés Quinones, Dominique Roche, Albert Ros, Marc Steinegger, Zegni Tricki, and Philippe Vullioud. Similarly, I thank my colleagues from Lausanne: Piret Avila, Slimane Dridi, Erica Fumagalli, Mauricio Gonzalez-Forero, Rémy Morier-Genoud, Charles Mullon, Jorge Peña, Simon Powers, and Tadeas Priklopil. I thank the "lunch group" for all lunches and good times spent outside the work environment: Jannik Beninde, Isabelle Castro, Miguel Dos Santos, Dumas Galvez, Laura Garaud, Christian de Guttry, Mikko Lehto, Lucas Marques da Cunha, David Nusbaumer, Manuel Pompini, Ricardo Kanitz, Anshu Uppal, and Fardo Witsenburg.

I would like to thank my friends, "de bende van ellende", and both my Dutch and Roman family. In particular, I thank my parents, Teun and Lina, who have always supported my endless studying, or lack thereof. I thank all other people who in some way or another have positively contributed to my thesis. Finally, I would like to thank my significant other, Elisa. Without her unconditional support and love in both good periods and periods of stress or physical incapability I would not have gotten so far.