Explaining fine-grained properties of human cooperation : Insights from evolutionary game theory

Félix Geoffroy

▶ To cite this version:

Félix Geoffroy. Explaining fine-grained properties of human cooperation : Insights from evolutionary game theory. Psychology and behavior. Université Montpellier, 2018. English. ⟨NNT : 2018MONTG071⟩. ⟨tel-02053488⟩

HAL Id: tel-02053488

https://theses.hal.science/tel-02053488v1

Submitted on 1 Mar 2019

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En : Biologie évolutive

École doctorale : GAIA

Unité de recherche : Institut des Sciences de l'Évolution de Montpellier

Explaining fine-grained properties of human cooperation. Insights from evolutionary game theory

Présentée par Félix GEOFFROY
Le 27 novembre 2018

Sous la direction de Charlotte FAURIE
Co-dirigée par Nicolas BAUMARD
Co-encadrée par Jean-Baptiste ANDRÉ

Devant le jury composé de

Rufus JOHNSTONE, Professeur, University of Cambridge	Rapporteur
Laurent LEHMANN, Professeur, Université de Lausanne	Rapporteur
Ingela ALGER, DR, Toulouse School of Economics	Examinatrice
François ROUSSET, Université de Montpellier	Président du Jury
Nicolas BAUMARD, CR, École normale supérieure	Co-directeur de thèse
Jean-Baptiste ANDRÉ, CR, École normale supérieure	Co-encadrant

Remerciements

Je tiens en premier lieu à remercier Jean-Baptiste André pour avoir encadré ma thèse. Jean-Baptiste a su me laisser une très grande liberté dans le choix des questions de recherche, tout en m'accompagnant et en me conseillant de manière toujours pertinente. Je dois beaucoup à Jean-Baptiste, tant du point de vue intellectuel qu'humain. A chaque difficulté que je rencontrais, Jean-Baptiste avait une réponse, ou mieux, il me guidait juste suffisamment pour que j'avance par moi-même. Par sa pédagogie et sa patience, il a été pour moi un encadrant idéal.

Je remercie chaleureusement mon co-directeur de thèse, Nicolas Baumard. Nicolas a toujours été disponible pour répondre à mes questions sur le domaine d'étude de la morale qui me passionne. Les discussions avec Nicolas ont toujours été des pivots dans ma thèse, ouvrant à chaque fois de nouvelles pistes de recherche.

Je tiens à remercier Michel Raymond pour m'avoir accueilli dans l'équipe Biologie Évolutive Humaine de l'ISEM. J'ai beaucoup appris au sein de cette équipe, notamment sur des sujets dont j'ignorais jusqu'à l'existence. Pour cela je tiens à remercier tous les membres de l'équipe, ainsi que pour les nombreuses discussions passionnantes et leur bonne humeur. Merci à Valérie, Charlotte, Claire, Mélissa, Clément, Bernard et Alexandre.

Nicolas Baumard et Coralie Chevallier m'ont accueilli dans leur équipe Évolution et Cognition Sociale pendant les deux autres années de ma thèse. Ils m'ont permis de m'ouvrir aux sciences cognitives dont j'ignorais tous les tenants et les aboutissants. L'environnement au sein de l'équipe a toujours été propice à la stimulation intellectuelle, mais aussi aux rapports amicaux. Merci à Nicolas, Coralie, Pierre et Lou. Et bien sûr, merci à Hugo, avec qui j'ai partagé les hauts et les bas du quotidien de la recherche, mais toujours dans la bonne humeur !

Je souhaiterais également remercier les autres thésards du DEC avec qui j'ai passé de francs moments de camaraderie. Merci à Sophie, Héloïse, Charles et Vasilisa !

Enfin, pour leur présence et leur soutien, je remercie ma famille, ainsi que les amis avec qui j'ai partagé ces trois années enrichissantes. Merci à Romain, Anne, Louise, Camille, Paul et Victor !

Abstract

The existence of cooperation among non-kin in many species constitutes an apparent paradox for evolutionary biologists. The most commonly accepted explanation is that cooperation can be enforced by mechanisms that reward cooperators or punish cheaters. Most of the theoretical works in evolutionary game theory, however, aim only at explaining how *some* cooperation can exist at an evolutionary equilibrium, thanks to these enforcement mechanisms. Here, we aim at showing, instead, that evolutionary game theory can also explain the *fine-grained properties* of the cooperation that takes place in the living world, especially in the case of the human species. First, we address the question of the origin of enforced cooperation: How can enforced cooperation evolve from an initially non-cooperative state? Using tools from the field of machine learning, we show that enforced cooperation can evolve as a by-product of adaptation to interactions with shared interests. We also show that this process has only two possible evolutionary outcomes. Either all cooperative opportunities are enforced, which corresponds to the human cooperative syndrome, or only a very few number are, which corresponds to non-human cooperation. We also propose a variation of this model to explain why many mutualisms are exaggerated forms of cooperation with shared interests. In a second approach, we focus on one specific enforcement mechanism called partner choice. Using agent-based simulations, we show that, when individuals can freely choose their cooperative partners, the only level of effort invested into cooperation that is evolutionarily stable is the one that maximizes the social efficiency of cooperation. We then build analytical models of partner choice imported from economic matching theory. We show that the only evolutionarily stable distribution of the benefits of cooperation is both independent of bargaining power and proportional to each participant's relative contribution. Thus, partner choice explains two fine-grained properties of human cooperation, namely our preferences for the most socially efficient forms of cooperation and our concerns for fair distributions. Finally, we show that costly signalling models of cooperation can explain several properties of moral reputation, and we conclude by discussing directions for future research.

Résumé

L'existence, dans de nombreuses espèces, de comportements coopératifs entre individus non-apparentés constitue un paradoxe apparent pour la théorie de l'évolution. L'explication la plus acceptée est que les comportements coopératifs peuvent être "incités" par un mécanisme qui récompense les coopérateurs et punit les tricheurs. On parle alors de "coopération conditionnelle". La majorité des travaux en théorie des jeux évolutionnaires cherchent seulement à expliquer comment des comportements coopératifs *en général* peuvent exister à un équilibre évolutionnaire. Dans cette thèse, nous cherchons au contraire à montrer que la théorie des jeux évolutionnaires peut aussi permettre de comprendre certaines des *propriétés fines* des comportements coopératifs qu'on observe dans le vivant, en particulier dans le cas de l'espèce humaine. Tout d'abord, nous posons la question de l'origine de la coopération conditionnelle. Comment la coopération conditionnelle peut-elle évoluer à partir d'une situation initiale dans laquelle personne ne coopère ? A l'aide de méthodes empruntées à l'apprentissage automatique, nous montrons que la coopération conditionnelle peut évoluer en tant que sous-produit d'une adaptation à des interactions dans lesquelles les intérêts des participants sont alignés. Nous montrons également que ce processus évolutif ne peut aboutir qu'à deux résultats opposés. Soit toutes les opportunités de coopération sont "trouvées" par l'évolution, ce qui correspond à la prévalence des comportements coopératifs chez l'Homme, soit un nombre très réduit d'opportunités de coopération sont "trouvées", ce qui correspond aux comportements coopératifs non humains. Nous proposons également une variante de ce modèle qui permet d'expliquer pourquoi de nombreux mutualismes sont des formes exagérées de cas d'interactions basées sur des intérêts communs. Dans un second temps, nous nous concentrons sur un mécanisme particulier de coopération conditionnelle : le choix du partenaire. Nous utilisons des simulations individu-centrées, et nous montrons que si l'on peut choisir librement ses partenaires dans la coopération, alors le seul niveau d'effort investi dans la coopération qui est évolutivement stable est celui qui maximise l'efficacité sociale de la coopération. Puis, nous développons des modèles analytiques, importés de la théorie économique des appariements. Nous montrons que la seule distribution des bénéfices générés par la coopération qui est évolutivement stable ne dépend pas des rapports de force et est proportionnelle à la contribution de chacun des participants. Ainsi, la théorie du choix du partenaire explique deux propriétés fines des comportements coopératifs chez l'Homme : nos préférences pour les formes de coopération les plus socialement efficaces et notre sens de l'équité. Enfin, nous montrons que la théorie des signaux coûteux, appliquée à la coopération, peut expliquer plusieurs propriétés de la réputation morale, puis nous concluons en discutant de futures directions de recherche.

Contents

List of Figures

Chapter 1

General introduction

1.1 The evolution of cooperation

Cooperative behaviours are usually defined as actions by which an actor provides a benefit to a recipient, and which have evolved because of this benefit (West et al 2007c). Cooperative actions are ubiquitous in the natural world. They can be found from the fundamental molecular level, to complex societies (Maynard-Smith and Szathmary 1995; Dugatkin 1997; Higgs and Lehman 2014). Yet, the mere existence of cooperation has been an evolutionary puzzle for decades (West et al 2007b). Indeed, natural selection favours only behaviours that enhance the relative fitness of the actor. Thus, there is no reason for a behaviour that benefits another individual to evolve in the first place. That is even more true if the action entails a fitness cost for the actor. In this case, cooperation should even be counter-selected.

The problem of the evolution of cooperation has been a major focus in evolutionary biology. Several mechanisms have been proposed by evolutionary theoreticians to reconcile the predictions of the theory of natural selection and the observed behaviours (reviewed in Sachs et al. 2004; Lehmann and Keller 2006; West et al. 2007b; Bshary and Bergmüller 2008).

The theory of kin selection shows how natural selection can favour costly cooperative behaviours directed towards genetically related recipients (Hamilton 1963, 1964; Grafen 1984; Rousset and Billiard 2000; Rousset 2004). The core idea is that a gene can enhance its replication in two ways. It can either increase the chance of reproduction of the organism to which it belongs (direct fitness), or it can act on the reproduction of another organism which carries a copy of itself (indirect fitness). The sum of these two components of fitness is called inclusive fitness. In consequence, a behaviour that increases the fitness of a recipient can be selected even if it comes at a cost for the actor, provided that the net inclusive fitness increases, for instance if both organisms are sufficiently genetically related. This situation can be achieved via kin discrimination or limited dispersal (West et al. 2007b). Kin selection has been used to explain many instances of cooperation in nature. It is believed to play a crucial role in the major transitions in evolution (Frank 1994; Maynard-Smith and Szathmary 1995; Levin and West 2017; Cooper and West 2018), the social lives of many micro-organisms (West et al. 2007a; Strassmann et al. 2011), or the evolution of eusociality in insects (Queller and Strassmann 1998; Foster et al. 2006). Humans also exhibit specific cooperative behaviours towards their kin (Hames 1987; Anderson 2005; Bowles and Posel 2005; Krupp et al. 2008).

However, kin selection alone cannot explain all the existing forms of cooperation. Some kinds of cooperative behaviours are indeed directed towards non-kin, and even sometimes towards individuals from another species (Smith and Douglas 1987; Boucher 1988). We will refer to the later case as inter-specific cooperation, or mutualism. Under these circumstances, the only evolutionary explanation is that the cooperative behaviour actually generates a net benefit which increases the actor's direct fitness. Usually, a distinction is made between two kinds of direct benefits.

First, an individual might help another one but only because doing so directly increases her own fitness. In this case, the recipient's benefit is merely a by-product of the actor's self interested action (Connor 1995; Clutton-Brock 2009; Connor 2010; Leimar and Hammerstein 2010). One example of by-product cooperation is the phenomenon of group augmentation (Kokko et al. 2001; Bergmüller et al. 2007). Some animals may prefer to live in groups with a cooperatively breeding system. They do so because each individual benefits from a larger group size, to protect themselves more effectively against predators for example. Another example of by-product cooperation is cooperative hunting in felids, canids and chimpanzees. Here, individuals hunt in packs but only because they all benefit from the presence of each other, which increases the likelihood of the hunting success (Packer and Ruttan 1988; Scheel and Packer 1991; Boesch et al. 2006). Lastly, an actor should take care of another individual if she benefits from her existence, for example by feeding from her waste. This specific mechanism is called partner fidelity feedback and is very common in mutualisms (Bull and Rice 1991; Sachs et al. 2004; Foster and Wenseleers 2006; Weyl et al. 2010). All these examples have in common that the actor's and the recipient's interests are somewhat aligned (Roberts 2005; Leimar and Hammerstein 2010). If there is no conflict of interest, and if both the actor and the recipient benefit from a cooperative behaviour, then its evolutionary stability is guaranteed.

These situations are often modelled using the Stag Hunt Game (Skyrms 2004) or the Snowdrift Game (Doebeli et al. 2004). Some instances of human cooperation have been shown to correspond to that kind of interaction with direct benefits (Alvard 2001; Tomasello et al. 2012; Aktipis et al. 2018).

Second, and however, it is likely that the vast majority of potentially cooperative interactions among non-kin in the natural world are situations in which individuals can indeed cooperate but also have partially conflicting interests. In game theory, these interactions are called mixed-motive games and a standard tool to model them is the Prisoner’s Dilemma, which can be described using the following payoff matrix:

		Player 2	
		C	D
Player 1	C	$b-c, b-c$	$-c, b$
	D	$b, -c$	$0, 0$

with $b > c > 0$. The only evolutionarily stable strategy (Maynard-Smith and Price 1973) in the one-shot Prisoner’s Dilemma is ”Defection” (D) due to a situation of a ”perfect” conflict of interest. Generally, in the context of natural selection, and in the absence of a by-product relationship, there is no reason for the interests of two individuals to be aligned. Thus, by default, the Prisoner’s Dilemma is the most appropriate approach to consider the majority of social interactions between non-kin. Evolutionary game theory therefore predicts that cooperation should not evolve at all in this game, in spite of the fact that individuals *could* benefit together by cooperating.

Nevertheless, it has been shown that cooperation can be favoured by natural selection in the Prisoner’s Dilemma if it is enforced by a mechanism that ”rewards” cooperators, or ”punishes” defectors (West et al. 2007b). For instance, both evolutionary biologists and economists have discovered that, if the Prisoner’s Dilemma is repeated indefinitely between the same two players, then conditional cooperation strategies can be stable (Luce and Raiffa 1957; Axelrod and Hamilton 1981; Fudenberg and Maskin 1986; Mailath and Samuelson 2006). In biology, this idea has been primarily known as reciprocity (Trivers 1971). Reciprocity can be illustrated by the very simple tit-for-tat strategy which stipulates that an individual should cooperate only if her partner has cooperated in the previous round (Axelrod and Hamilton 1981). In this way, individuals who play this strategy will pay the cost of cooperation but they have the guarantee that their opponent will cooperate in return, since it is the best strategy against tit-for-tat.

Other enforcement mechanisms have been proposed. In indirect reciprocity, an individual helps another one in order to maintain her reputation so that she will be helped by third-parties in the future (Nowak and Sigmund 1998; Leimar and Hammerstein 2001). In partner choice, individuals cooperate to be chosen as cooperative partners in future interactions (Bull and Rice 1991; Noë and Hammerstein 1994, 1995). In punishment (Boyd and Richerson 1992; Boyd et al. 2003; Bowles and Gintis 2004), or in ”sanctions” (Sachs et al. 2004), cooperation is stabilized by the fear of a cost inflicted to defectors. The common idea of all enforcement mechanisms is that an individual cooperates and ”triggers” a conditional response from her partner (or third-parties) in order to benefit from cooperation in return, or to avoid future punishments.

A few examples of enforced cooperation have been described in non-human species. The ”cleaner” fish *Labroides dimidiatus*, for instance provides a service to its ”client” fish by eating their ecto-parasites. However cleaners prefer to eat their client’s mucus. If a cleaner fails to cooperate by eating the mucus instead of parasites, the client can retaliate by chasing the cleaner or leaving the interaction (Bshary 2001; Bshary and Grutter 2005). These two responses correspond to forms of sanctions and partner choice. Legume plants trade carbohydrates for nitrogen with rhizobia, their symbiotic bacteria. It has been shown that a plant can reward or punish bacteria strains according to their propensity to cooperate by providing nitrogen (Simms and Taylor 2002; Kiers et al. 2003). A similar behaviour has also been described in the plant-mycorrhizal mutualism where individuals from the two species exchange carbohydrates for phosphorus (Kiers et al. 2011). Examples of sanctions are found in insect societies (Wenseleer and Ratnieks 2006), as well as in animal societies such as meerkat groups (Clutton-Brock and Parker 1995; Young et al. 2006). Some examples of reciprocal interactions, stricto sensu, are documented in allogrooming in impalas (Hart and Hart 1992), vervet monkeys (Borgeaud and Bshary 2015), baboons (Cheney et al. 2010) and chimpanzees (Phelps et al. 2018). Other social interactions based on reciprocity have been described in bats (Wilkinson 1984; Carter and Wilkinson 2013) and reindeers (Engelhardt et al. 2015).

Humans, on the other hand, daily employ a large variety of enforcement mechanisms, including reciprocity (Trivers 1971; Dugatkin 1997; Brown et al. 2004). Thus, human cooperation cannot be fully understood using only kin selection and by-product cooperation models (Palameta and Brown 1999; Melis and Semmann 2010). For this reason, we will focus on the study of enforced cooperation, also called conditional cooperation. Although this topic has been extensively studied by game theoreticians, the current theory needs refinements before it can account for fine-grained properties of cooperative behaviours, especially in humans. In the next sections, we will present some of the limits of the current evolutionary theory of conditional cooperation and how the present work aims at tackling them.

1.2 Refining the evolutionary theory of conditional cooperation

Most of the work that has been done on the evolution of conditional cooperation aims at answering the following question: "how can the mechanism x stabilize *some* cooperation instead of no cooperation at all?" As we have seen before, several mechanisms of conditional cooperation, i.e. enforcement mechanisms, have been extensively studied in the past to answer this particular question. However, another equally important question still needs to be addressed in greater details: "which forms of cooperation are more likely to evolve?"

One of the most striking result regarding this question is the so-called "folk theorem" of iterated games (Fudenberg and Maskin 1986; Aumann and Shapley 1994). In the context of the Iterated Prisoner's Dilemma (IPD), the folk theorem states that, if the two players are arbitrarily patient, then any strategy that yields a positive payoff to both players can be an equilibrium. For instance, the tit-for-tat strategy can be an equilibrium and yields a positive payoff. Yet, other strategies can stabilize various patterns of cooperation. Intuitively, let us say that a strategy is predominant in the population and that it "punishes" the individuals who deviate from it, for example by defecting forever (as the Grim-trigger strategy). Then this strategy is optimal against itself and is therefore an equilibrium. Thus, any pattern of cooperation can be stabilized in the Iterated Prisoner's Dilemma.

The folk theorem raises a problem of multiplicity of equilibria. It predicts that many strategies, yielding many different cooperative outcomes can be stable. However, it is not possible to predict which of these equilibria are more likely to occur, a problem called the equilibrium selection problem. It has been argued that the folk theorem reduces the predictive power of the evolutionary theory of conditional cooperation (Boyd 2006). Moreover, the problem of multiplicity of equilibria has also been found in other forms of conditional cooperation, e.g. in punishment (Boyd and Richerson 1992) and in a partner choice setting with signalling (Gintis et al. 2001).

One should note that the strategy "always Defect" is also a possible equilibrium, and even an evolutionarily stable equilibrium. This raises a subsidiary question, which is the question of the origin of conditional cooperation, or the bootstrapping problem of conditional cooperation (André 2014). A rare cooperative mutant in a population of pure defectors is very unlikely to invade (Axelrod and Hamilton 1981). More generally, we lack a theory that would explain how transitions from an equilibrium to another can, or cannot, occur.

The well-studied case of iterated interactions has, therefore, yielded mixed results. On one hand, it has yielded very powerful insights on how cooperation can be stabilized by conditional behaviours. On the other hand, it cannot provide any answer to a range of interesting biological questions such as: Which forms of cooperation should evolve? Do these predictions fit the cooperative behaviours we observe in nature? Can we predict in which situation, which equilibrium is more likely to be reached? Will conditional cooperation evolve at all? Will efficient forms of cooperation evolve? To sum up, the evolutionary theory of conditional cooperation lacks models which could explain fine-grained properties of cooperative behaviours.

Previous studies have proposed new approaches to refine the predictions of conditional cooperation models. For example, several authors have modelled cooperation as a continuous investment, rather than an all-or-nothing decision. It has been done both in the Iterated Prisoner's Dilemma (Wahl and Novak 1999a and b; Killingback and Doebeli 2002; André and Day 2007) and in a partner choice context (Sherratt and Roberts 1998; Ferriere et al. 2002; Foster and Kokko 2006; McNamara et al. 2008). By this means, one can study which levels of investment in cooperation are more likely to occur at the evolutionary equilibrium. Another line of approach to the problem of multiplicity of equilibria is group selection (Boyd and Richerson 1990), or cultural group selection (Boyd and Richerson 2009). We consider the latter approach to be beyond the scope of this manuscript. Besides, we will show that solutions to the problem of multiplicity of equilibria can be found using only "between-individual" selection, and, thus, without invoking additional assumptions such as "between-group" selection.

Despite the previous approaches described above to refine the theory of conditional cooperation, we argue that this issue should be further investigated. In the context of this thesis, we will attempt to provide several methods to address this matter. We will mainly focus on human cooperation. Since humans cooperate on a daily basis and in so many different ways, and since they have been extensively studied, including in laboratory experiments, human cooperation is a perfect study case for trying to explain fine-grained properties of cooperation.

1.3 Explaining fine-grained properties of human cooperation

There is no reason to think that human cooperation is unique in the sense that it should require a dedicated and exclusive evolutionary explanation. Rather, human cooperation differs from other instances of cooperation between non-kin by its extent and its prevalence. Throughout this manuscript, we will try to explain four specific properties of human cooperation.

The first property is the general human ability to cooperate in so many circumstances, which has been made possible by enforcement mechanisms such as reciprocity (Smith 2003; Gurven 2004; Hill 2002; Bowles and Gintis 2011). Unlike many instances of non-human cooperation, human cooperation is "general", in the sense that a cooperative action expressed by a focal individual can be paid back later by several types of cooperative actions. In other words, humans have a general cognitive system that can compute how to reward an individual and can recognize a reward as well and react accordingly. In contrast, non-human forms of conditional cooperation are restricted to a limited range of rewards and sanctions. For instance, a legume plant can reward a cooperative rhyzobia strain only by sending more chemical nutrients, or providing a better shelter. Other forms of help would not be correctly "interpreted" as rewards by the rhyzobia and, therefore, would not trigger more cooperation on its side. One can see how the question of the diversity of cooperative interactions relates to the problem of multiplicity of equilibria: several cooperative equilibria are possible and, clearly, the "human cooperative equilibrium" is very different from those of other species. The question of which equilibrium is going to be favoured by natural selection cannot be addressed by the standard tool of Iterated Prisoner's Dilemma.

The second property of human cooperation that we will try to explain is the preferences humans have regarding the amount of effort they should put into cooperation. Humans have strong preferences for the most socially efficient levels of investment into cooperation (Cronk 2007; Lange and Eggert 2015; Santamaria and Rosenbaum 2011). That is, people think they should help one another only when helping is mutually advantageous, i.e. when the marginal cost of helping is less than the marginal benefit of being helped. Again, the folk theorem states that any forms of cooperation can be stable, not necessarily the most socially efficient ones. Hence, again, the question of the amount of effort invested into cooperation cannot be addressed solely with the Iterated Prisoner's Dilemma.

The third property is often referred to as "human fairness": we have precise preferences regarding the distribution of benefits and costs among social partners. For instance, in interactions that entail a first phase of collaboration, and then a second phase of division of the benefits generated in the first phase (e.g. cooperative hunting), humans show preferences for specific divisions which are not based on the relative bargaining powers (Boehm 1993, 1997; Fehr and Schmidt 1999; Camerer 2003; Fehr and Fischbacher 2003; Dawes et al. 2007; Tricomi et al. 2010), and which are proportional to the participants' respective contribution in the cooperative venture (Marshall et al. 1999; Alvard 2002; Gurven 2004). Once more, the question of the distribution of the benefits and costs of social interactions cannot be addressed with the Iterated Prisoner's Dilemma.

Finally, the last property of human cooperation that we will try to explain is the importance of moral reputation, or cooperative reputation. Humans carefully pay attention to others' cooperative behaviours so that they can accurately choose to interact with the most cooperative partners. Correspondingly, humans cooperate in very specific ways to maintain their cooperative reputation (Sheldon et al. 2000; Haley and Fessler 2005; Sperber and Baumard 2012; Baumard et al. 2013). In the context of game theory, cooperative reputation has been studied using signalling games (Smith and Bliege Bird 2000; Gintis et al. 2001; Benabou and Tirole 2006). These games, however, also entail a similar problem of multiplicity of equilibria (Riley 2001) which cannot be addressed with the IPD. We want to understand which cooperative signals are more likely to be used at the evolutionary equilibrium, and if they correspond to the rationale of moral reputation.

1.4 Specific models for specific questions

In order to tackle the different fine-grained properties of human cooperation, we will use a specific model for each case. Since the classic framework of Iterated Prisoner's Dilemma is not suited to answer our particular questions, we will have to turn to other tools that are not so commonly used in evolutionary biology. We will show that importing some modelling tools from other fields can shed light on evolutionary processes.

In order to explain the human ability to cooperate in so many ways, one must address the problem of multiplicity of equilibria. More precisely, we want to understand why it is the case that, starting from an initially non-cooperative situation, some species have never evolved cooperation, some have evolved cooperation but only in restricted interactions, and humans have evolved a general capacity for cooperation. We, therefore, must tackle the problem of the *origin* of cooperation, rather than the problem of its *stability* alone.

The question of how conditional cooperation can evolve from an initially non-cooperative population, or, more generally, the question of the transitions between two cooperative equilibria, cannot be answered using a classic game theoretical framework. Let us briefly explain why it is the case. Suppose that there is a costly cooperative action $\mathcal{A}$ that can be enforced by another action $\mathcal{R}$. The enforcement action $\mathcal{R}$ could be to provide a reward on the condition that the action $\mathcal{A}$ has indeed been performed. Note that, if the enforcement action is to reply with the same cooperative action $\mathcal{A}$, our situation would correspond to the particular case of reciprocity. Let us assume that a population of individuals is "stuck" in a non-cooperative equilibrium where no one takes the cooperative action $\mathcal{A}$, and no one rewards it with $\mathcal{R}$. There is a strong evolutionary pressure not to take the cooperative action $\mathcal{A}$, since it is costly and it will not provide any future benefit. In contrast, the rewarding response $\mathcal{R}$ is under neutral selection since the individuals are never confronted to the action $\mathcal{A}$. The rewarding response $\mathcal{R}$ is unaffected by natural selection in the same way that a reaction norm to temperature is not shaped by natural selection to provide an adaptive response to temperatures that are never experienced in the environment. Standard game theory provides no insight to the evolution of such *latent* responses. Yet, it is a question of primary interest in conditional cooperation. Indeed, if some individuals happened to evolve, by "chance", the *latent* capacity $\mathcal{R}$ to provide a reward, natural selection would now favour the evolution of the cooperative action $\mathcal{A}$. This mechanism could thus provide an explanation for the problem of equilibrium selection.

Most game theoretical models are based on a high degree of abstraction regarding the mechanistic underpinning of behaviours. By contrast, some evolutionary processes require a mechanistic approach of behaviour to be fully understood (André 2014, 2015; van den Berg and Weissing 2015; André and Nolfi 2016; Bernard et al. 2016). In our case, in order to study the transition from a cooperative equilibrium to another, we have to model how an individual would react to a situation, here a form of cooperation, that he has never encountered in its evolutionary past. Surprisingly, another branch in the study of behaviours faces a similar question: machine learning. Hence, we will import concepts from machine learning, as well as modelling methods, into the evolutionary study of cooperation. Note that other studies have drawn fruitful parallels between evolution and machine learning (Watson et al. 2014; Watson and Szathmáry 2016; Kouvaris et al. 2017).

Another angle of approach for tackling the problem of multiplicity of equilibria is partner choice. Partner choice is one of the many mechanisms of enforcement that can stabilize conditional cooperation. Yet, partner choice has a unique feature which makes it a relevant tool for investigating the question of equilibrium selection: the "outside options". Intuitively, one possible explanation for why models of reciprocity based on the IPD framework allow so many cooperative equilibria is that, when an individual is confronted with a defector, her best strategy is to defect as well, resulting in a payoff of zero. Thus, any cooperative equilibrium which provides a payoff greater than or equal to zero is a possible equilibrium. However, in many biological systems, defecting is not the only option that an individual can choose when she interacts with a defector. Many individuals would rather leave the current interaction and search for another, more cooperative partner. This mechanism is called partner choice or partner switching (Bull and Rice 1991; Noë and Hammerstein 1994, 1995), and the population of individuals is refereed to as a biological market (Noë and Hammerstein 1994, 1995). Now, one can see that the possibility to interact with another partner, i.e. the "outside option", generates a novel constraint which restricts the range of possible cooperative equilibria. Strategies which provide an individual with a payoff smaller than her outside option cannot be stable.

Thus, partner choice sheds light on two of the human cooperative properties that we want to study, namely the level of investment in cooperation and the distribution of the benefit of cooperation. If many cooperative individuals are present in the biological market, an individual should refuse to interact with a partner who invests too little into cooperation (McNamara et al. 2008), or who proposes a very "unfair" division of the benefit (André and Baumard 2011a, 2011b; Debove et al 2015a, 2015b, 2017).

Nonetheless, studying the evolution of cooperation in a biological market is a complex task, mostly because it entails an element of circularity. Indeed, the cooperative behaviours that can evolve are constrained by the individuals' outside options, but the outside options also depend on the distribution of cooperative behaviours in the population. Fortunately, economists have developed a branch of market models that we can import into evolutionary biology and use to study the evolution of cooperation in biological markets: matching theory (Chade et al. 2017).

It should be noted that, even though our models were primarily designed to study the evolution of human cooperation, the results may provide insights into other instances of conditional cooperation. Human cooperation is a good case study for conditional cooperation, but there is no reason why our models could not be applied to non-human cooperation. Our first model, which involves machine learning concepts, concerns every form of conditional cooperation, thus, it can be used to explain the evolution of non-human cooperation behaviours as well. Similarly, matching theory can be used to explain important features of non-human biological markets (Hammerstein and Noë 2016).

This manuscript is composed of three parts. Chapters 2, 4 and 5 are presented in the form of articles and start with a short section providing context and a summary of the results. Part I is dedicated to the evolution of conditional cooperation by generalization. Chapter 2 presents a model for the evolutionary origin of conditional cooperation inspired by machine learning. Chapter 3 briefly presents an extension of the general model to cases where conditional cooperation evolves by exaggeration. In Part II, we study how partner choice can help refining the predictions concerning two fine-grained properties of human cooperation. Chapter 4 presents classic models of evolutionary game theory such as adaptive dynamics to study the evolution of the investment put into cooperation. Chapter 5 introduces models from matching theory in the study of partner choice and tackles the division of the benefits of cooperation. Finally, in Part III, we first review models from the costly signalling theory and discuss how they could explain the evolution of some features of human moral reputation, despite a multiple equilibria problem (Chapter 6). We conclude with the contributions and the limits of the present work, and with further research directions in Chapter 7.

Part I

The evolution of conditional cooperation

Chapter 2

Generalization and the bootstrapping problem of cooperation

2.1 Objectives and summary

In this Chapter, we investigate the bootstrapping problem of conditional cooperation. Any form of conditional cooperation, such as reciprocity, partner choice, or punishment, entails two distinct traits: (i) the ability to cooperate conditionally upon a reward (which can take the form of a cooperation in return for instance, or an absence of punishment), and (ii) the ability to reward cooperation. Although the question of the evolutionary *stability* of conditional cooperation has been extensively studied in previous works, less attention has been given to the question of its *origin*. Because the simultaneous existence of two traits is required for conditional cooperation to be stable, its evolution raises a bootstrapping problem (Gardner and West 2004; André 2014, Barta 2016). At the end of this chapter, we will also discuss the fact that this bootstrapping problem can be seen as a particular case of the more general question of the multiplicity of equilibria in repeated games (Fudenberg and Maskin 1986; Aumann and Shapley 1994; Boyd 2006).

Here, we aim to propose a plausible solution to the bootstrapping problem. In a nutshell, the scenario is the following. Let us suppose that, cooperation is initially absent. No reward is thus ever given in exchange for a cooperation. The *ability* to reward cooperation, however, can be present but never actually expressed. That is, *latent* rewards can exist. If this is the case, then cooperation can be favoured by selection as it triggers the expression of these latent rewards, transforming them into actual rewards. Hence, if the *ability* to reward cooperation evolves for an independent reason, i.e. as a by-product of another adaptation, then conditional cooperation can eventually evolve. The question, therefore, is: Under what conditions can a latent ability to reward cooperation evolve in a species?

In this chapter, we build a model to study this scenario. The model entails three crucial assumptions. (1) During their life, individuals are engaged in several social interactions, that is, several games, but have a single cognitive mechanism that is used to take decisions in all these games. This assumption is key to studying the influence of adaptation in one game on behaviour in other games. (2) Second, we use a simple principal-agent model in which an agent can invest in a game conditionally upon a reward generated by the principal. This way, our model is sufficiently general to be applied to any form of conditional cooperation. (3) Lastly, we assume that there are games in which the principal is selected to provide a reward. Biologically, such games corresponds to *interdependent* forms of cooperation, i.e. games where the agent and the principal share common interests (Leimar and Hammerstein 2010).

Using machine learning tools, we study the capacity of the principal to reward cooperative games as a result of being selected to reward only interdependent games. That is, we want to understand under what conditions the principal can be selected to reward the agent in a limited set of games and *generalize* this ability to reward other games as well, thereby allowing the evolution of conditional cooperation in spite of the bootstrapping problem. As predicted by machine learning theory, we show that the principal generalizes only if the number of interdependent games is large enough. In the opposite case, the principal is never able to reward cooperation. Our model can explain the gap between human cooperation which is very general, occurring in a wide diversity of contexts, and cooperation in most other species which remains limited to specific kinds of behaviours or exchanges.

The rest of this chapter is a paper in preparation.

Generalization and the evolution of cooperation

Abstract

Scholars from game theory and evolutionary biology have long understood that cooperation can be stabilized if it is enforced by a reward mechanism. Any form of costly cooperation can be favoured by natural selection if individuals have, at the same time, both the ability to reward cooperative investments and the ability to monitor rewards and invest conditionally upon them. However, these two abilities need each other to be adaptive: the capacity cooperate conditionally upon rewards is useless when no rewards are provided, and vice-versa. Therefore, it is difficult to think of a way for these two abilities to have evolved from an initial population of individuals that neither reward, nor cooperate. That is,there is a chicken-and-egg, or bootstrapping, problem in the evolution of conditional cooperation. The only solution to this problem is that one side (cooperation or reward) first evolves for another reason and is then, ”recycled” into a conditional cooperation function. In this paper, we present such a scenario were individuals ”generalize” the ability to reward costly cooperative investment from previously existing interdependent forms of cooperative interactions. To do this, we combine a very simple and general principal-agent model with a machine learning framework. We show that our scenario can account for an evolutionary path towards cooperative equilibria. Our model can also explain both the fact that conditional cooperation is limited to specific forms in most taxa and the fact that, on the contrary, humans are able to cooperate in so many different ways, a situation often called the ”human cooperative syndrome”.

2.2 Introduction

As a general rule, natural selection does not favour strategies that support the common good. Except for genetically related organisms, any action that entails a fitness cost is counter-selected, no matter how beneficial it is for other individuals (Hamilton 1964; Williams 1966; West et al. 2007b). Game theoreticians often illustrate this issue with the prisoner's dilemma, or the tragedy of the commons, where cooperation cannot evolve because individuals have no incentives to aim for the common good (Hardin 1968). Owing to this gap between individual and social interests, natural selection results in a "cooperation load", a set of cooperation opportunities that are missed by evolution.

There is, however, a solution that can, at least in principle, make any costly behaviour adaptive as long as it produces a net social benefit: it has to be compensated with a reward. This can be understood with a simple principal-agent model from the theory of incentives (Laffont and Martimort 2002). A first individual, called the principal, commits herself to providing a second individual, called the agent, with a conditional reward, in order to incentivize him to perform an action that he would otherwise have not carried out. This is adaptive for the principal if the benefit she gets from the agent's action is greater than the reward she has to provide. In principle, it is possible to incentivize in this way any cooperative action that comes at a cost c and generates a benefit b, as long as $b > c$, that is as long as the action has a net social benefit.

Rewards being costly, however, an additional mechanism is needed to guarantee that the principal fulfils her promise. In the theory of incentives, rewards are usually assumed to be part of a contract enforced by institutions, but non-cooperative game theory shows that they can also be enforced endogenously if interactions are repeated (Axelrod and Hamilton 1981; Fudenberg and Maskin 1986; Mailath and Samuelson 2006). In order to attract cooperative investments, a principal must build up and maintain the reputation of being someone with whom cooperation pays off. To that aim, she must deliver on her promise and truly reward cooperation in all instances. Hence, cooperating and rewarding are both evolutionarily stable provided agents assess the rewarding tendency, that is to say the "reputation", of principals and decide to cooperate accordingly.

In principle, therefore, any form of cooperation can be evolutionarily stable as long as it is expressed conditionally on the presence of a future reward. This lies at the heart of all models of cooperation with non-kin: direct reciprocity (Axelrod and Hamilton 1981), indirect reciprocity (Nowak and Sigmund 2005), punishment (Boyd and Richerson 1992), partner choice (Noë and Hammerstein 1994, 1995). The problem, however, is that, whereas rational agents will always find the cooperative solution as long as it is mutually profitable, owing to a chicken-and-egg problem there is no guarantee that biological evolution will also find it (André 2014). Cooperation can only be adaptive provided at least some partners reward it. But rewarding others' cooperation can only be under selection if a significant fraction of them already cooperate conditionally. That is, neither cooperation nor rewards is favoured in the absence of the other. The non-cooperative situation, where agents do not cooperate and principals do not reward, is thus an evolutionary trap from which natural selection cannot escape. This should potentially prevent cooperation from evolving in every form of conditional cooperation such as reciprocity, reputation-based partner choice, or punishment (Gardner and West 2004; McNamara et al. 2008; André 2014; Barta 2016). This raises the question: How can cooperation emerge in spite of this bootstrapping problem?

Another function that is notoriously subject to the same type of bootstrapping problem is communication. Two complementary traits must be present for communication to occur: the ability to produce a signal and the ability to respond to the signal, neither of them being adaptive without the other. For this reason, evolutionary biologists understand that communication can only evolve if one side (the signal or the response) preexists for another reason, and is then "recycled" into a communicative function (Krebs and Dawkins 1984; Scott-Phillips et al. 2012). The same must necessarily hold in the case of conditional cooperation. Its evolution necessarily requires that one side (cooperation or reward) preexists for an independent reason and is then "recycled" into a conditional cooperation function (André 2015). In this paper, we aim to understand this recycling process, delineate its conditions of occurrence, and describe its consequences.

At first, for evolutionary biologists, the recycling story seems to be no more than a case by case proximate narrative, with no general, ultimate principle. But this is mistaken. Recycling is using a biological function in a context that differs from the one this function has initially been selected for. It turns out that the performance of biological organisms outside the range of situations for which they have been selected can actually be understood in a principled manner. In an other field, machine learning, this understanding is even the fundamental underlying topic. The object of machine learning is to understand and make use of the fact that "behavioural devices" can sometimes perform well outside the range of contexts they have originally been made for, a possibility called *generalization*.

Because natural selection is an adaptive process, like learning, the concept of generalization can readily be extended to biological evolution (Arak and Enquist 1993; Enquist and Arak 1993; Johnstone 1994; Enquist and Johnstone 1997; Ghirlanda and Enquist 2003; Watson et al. 2016; Watson et al. 2014; Watson and Szathmáry 2016; Kouvaris et al. 2017). In this paper, we will show that conditional cooperation can evolve for the same reason that machines generalize. More precisely, conditional cooperation evolves if the fact to reward others' cooperation becomes a general ability that extends beyond the very set of contexts for which it has initially been selected. We will understand that the conditions necessary for such evolutionary generalization to take place provide key insights on the nature of conditional cooperation in the wild, its limits in extant species, and its phylogenetic distribution.

2.3 Games of life

Evolutionary game theoreticians generally study the evolution of social behaviours in a single game, isolated from all other facets of individuals' life. This entails the implicit assumption that individuals evolve separate rules of decision adapted to each and every game. This, however, does not reflect the way individuals really make decisions. An individual's life is a succession of situations (some of which are games others are not) and, although each situation entails a specific problem, the individual must eventually make all decisions with the same cognitive device. The brain is divided into various circuits specially evolved to deal with different types of situations (e.g. feeding, mating, cooperating, see Barrett and Kurzban 2006), but all these circuits must be activated conditionally upon circumstances. That is, ultimately, it is the same brain that must, in a way or another, measure the parameters of every situation and eventually make a decision. Hence, the evolution of social behaviour should not be modelled as a series of independent evolutionary processes with a set of independent games, but rather as a single evolutionary process by which the decision machinery of individuals adapts to an entire set of games (Samuelson 2001; Bednar and Page 2007; Bednar et al 2012; Mengel 2012).

This premise has two kinds of implications. First, adaptation to one game can interfere with adaptation to other games so that perfectness cannot be achieved in all games owing to cognitive constraints or trade-offs (Samuelson 2001; Mengel 2012). Here, we will not consider this aspect of the multi-game problem. We will focus on the second implication which is more general and unavoidable. Adaptation to one set of circumstances that the organism regularly encounters is likely to shape the way the organism *would* respond to circumstances that it never encounters. That is, selection in existing situations shapes the organism's latent responses: decisions it would make if it had to but are selectively neutral because it never actually has to (Arak and Enquist 1993). Latent responses are never expressed but nevertheless play a key role in social evolution, as individuals are under selection to make use of others' latent responses, thereby revealing them, when this can be beneficial. For instance, in mate choice, latent preferences can be exploited via supernormal-stimuli, such as males displaying exaggerated traits to attract females more effectively (Ryan 1990; Ryan and Keddy-Hector 1992; Ghirlanda and Enquist 2003). We will see that this phenomenon can play a major role on the evolution of conditional cooperation.

Our multi-game model is as follows. Individuals are confronted, across their life, to a given set of situations, chosen in a random order, which we all call games. Each game involves two individuals: a principal and an agent. These two individuals are characterized by a decision mechanism (a behavioural "controller") that evolves by natural selection (see section *Generalization creates latent rewards* for more details on the nature of the controller). For simplicity, we assume that the principal and the agent are drawn from two separate populations (but the model would be almost identical otherwise). In each game, the agent first decides whether he wants to play or decline the game. If the agent decides to play, he pays a personal cost c while the principal gets a benefit b, and the principal then has the opportunity to decide on a quantitative reward r that she can transfer to the agent. There are four types of games depending on the values of b and c.

In **Selfish** games, the agent personally benefits from investing (that is, she pays a negative cost $c < 0$) and the principal may either benefit or be unaffected by the game ($b \geq 0$; we do not consider games that would hurt the principal, that is where $b < 0$). Selfish games comprise, in particular, the very large amount of situations in which the agent simply acts in a self-serving way without involving the principal ($b = 0$). For instance, the agent's solitary foraging activities belong to the category of selfish games (note that, strictly speaking, these situations should not be called "games" as they need not involve any actual interaction).

In **Wasteful** games, the agent pays a cost for investing (that is, $c > 0$) and her investment either does not benefit the principal ($b = 0$) or benefits her insufficiently to compensate for the cost ($b < c$). Hence, there cannot be any mutually beneficial reward in wasteful games. Would the principal provide a sufficient reward to compensate for the agent's cost, she would be worst off than if the agent had not invested at all. These games correspond to the overwhelming majority of situations in which the agents should simply not invest because it is useless, costly for herself, and not beneficial enough for others.

In **Cooperative** games, the agent pays a cost $c > 0$ for investing (that is, for cooperating), and her investment generates a benefit $b > c$ for the principal. Even though the agent has no "immediate" benefit in cooperating, a mutually beneficial agreement can be found if the principal rewards with $r \in]c, b[$. Cooperative games thus represent all the situations where cooperation has a net social benefit but raises an evolutionary bootstrapping problem. Throughout this paper, we will speak of "cooperation" (in a slightly unorthodox manner, see West et al. 2007c) to qualify specifically the fact that an agent invests in a cooperative game in exchange for a reward. We aim to understand how cooperation, in this sense, can evolve.

Finally, there is a fourth category of games, intermediate between cooperative and selfish, that we call "**Interdependent**" games. Interdependent games comprise situations in which the agent performs an investment that causally interacts with the principal. As a result, the eventual net benefit of the investment, hence the precise nature of the game, varies as a function of the quality of the principal. In the literature, interdependent games are often referred to as "by-product cooperation" (Clutton-Brock 2009; Connor 2010; Leimar and Hammerstein 2010), or cooperation through interdependence (Roberts 2005). This kind of cooperation occurs, for instance, when an ant colony invests in protecting an acacia tree where they have a shelter. The eventual benefit of the investment, for the ant colony, then depends on the acacia's quality (Heil et al. 2009). If the acacia is of high quality, the ants benefit from protecting it because they preserve a good shelter (in this case the interaction is a Selfish game). But if the acacia is of low quality, then the ants may be losing their time, especially if better acacias are available (in this case, the interaction is either a Cooperative or a Wasteful game). Another example is a collective action. The benefit of investing into a collective action (e.g. a collective hunt) with a partner eventually depends on the quality of this partner.

In interdependent games, the eventual benefit of the investment for the agent can be seen as an "automatic" reward, r_a, offered by the principal because it is a causal response of the principal to the investment made by the agent, the benefit of which (the value of r_a) depends on the principal's quality. It is not, however, an "active" reward, in the sense that the principal does not produce r_a on purpose to reward the agent, but for an independent reason. Accordingly, this "automatic" reward has no cost for the principal. Acacias for instance need not purposefully be good shelters. Some acacias just happen to have properties that make them worthy (nor not) of being protected by ants (in reality acacias actively invest in being good partners in their interaction with ants, but this is a secondary consequence of social evolution that we will understand later).

Formally, interdependent games are characterized by a positive cost $c > 0$ for the agent, an automatic reward r_a that is sometimes sufficient to compensate for the cost ($r_a > c$) and sometimes insufficient ($r_a < c$), and a benefit b for the principal. We assume that this benefit is always sufficient to compensate for the agent's net cost, that is $b > c - r_a$. Consequently, if the automatic reward is insufficient ($r_a < c$), an active reward r can always be found that would make the interaction mutually beneficial.

Overall, therefore, every game is characterized by three payoff parameters: the cost $c \in [-1, 1]$ for the agent, the benefit $b \in [0, 1]$ for the principal, and the automatic reward $r_a \in [0, 1]$ (that is, the part of the agent's benefit that depends upon the principal's quality). In addition to these three parameters, every game is also characterized by a vector $\mathbf{s}$ of arbitrary parameters called "spurious" features. These represent the vast majority of properties of the situation, that an individual can measure and respond to even though they are not informative about the game's payoff.

Each game proceeds as follows. The agent is informed of the parameters of the situation that are independent of the principal (b, c, and $\mathbf{s}$), and decides on a course of action. He can either (1) decline the game, (2) play the game unconditionally, without evaluating the principal's quality, or (3) "pay to see" the principal's quality before making a final decision. That is, before proceeding with the actual game, the agent pays a fixed "monitoring cost" to measure the overall reward ($r_a + r$) that he would eventually get, would he decide to play the game. Biologically speaking, this is meant to capture all the diverse manners in which one can obtain information on others. For instance, the agent could collect information by observing some aspects of the principal's phenotype that are correlated with her quality, by playing the game once to "test" the principal, by finding out about the principal's past behaviour (that is, evaluating her "reputation"), etc. Irrespective of the precise way in which this occurs, the important point is that nothing comes for free. Obtaining, computing, and remembering information about the principal entails a cost m.

If the agent has decided to decline (1), the game ends with no payoffs and the next game is randomly chosen. Otherwise, the principal measures all the game parameters (including his personal "automatic reward" r_a) and decides on the active reward r that she offers to the agent in response to his investment. If the agent has decided to play the game unconditionally (2), the game proceeds. On the other hand, if the agent has chosen to "pay to see" before making a decision (3), he is then informed of the principal's quality ($r_a + r$), and then decides whether he wants to actually play the game or decline.

Eventually, if the agent decides to play the game (conditionally or not), the payoffs are given to the players: the principal receives $b - r$, and the agent receives $r_a + r - c$ or $r_a + r - c - m$ depending on whether he payed the monitoring cost or not.

Latent rewards are pivotal in the evolution of cooperation

Initially, the principal is assumed to never reward any game ($r = 0$ in all cases). The agent is under selection (i) to invest unconditionally in all Selfish games because they are always profitable, (ii) to invest conditionally in all Interdependent games because they are sometimes profitable and sometimes not depending on the principal's quality (we assume that the monitoring cost is low enough, see *Methods*),that is the agent should invest in an Interdependent game only when the principal's quality is such that $r_a + r > c$, and (iii) to decline unconditionally all Wasteful and all Cooperative games because they are never profitable anyway in the absence of active reward. Cooperation is, therefore, absent.

Assuming the agent is perfectly shaped by natural selection and behaves optimally in all games, the principal is, in turn, under selection to behave optimally in Selfish and in Interdependent games. In Selfish games, she is under selection to give a reward $r = 0$. Any strictly positive reward would be a loss because the agent invests unconditionally in these games. In Interdependent games, the principal is under selection to give also $r = 0$ when the automatic reward happens to be sufficient to recoup the agent's cost ($r_a > c$) but a positive reward $r = c - r_a + \epsilon$ when $r_a < c$, that is the principal should always compensate for the agent's net loss in order to incentivize his investment. On the other hand, the principal's propensity to reward Wasteful and Cooperative games is not under selection because the agent never invests in these games anyway. Hence, the reward that the principal would offer in these games, if the agent did play them, are *latent* rewards.

It turns out that the evolution, or non-evolution, of cooperation depends critically on the nature of latent rewards. To understand, consider a given Cooperative game. Let us first assume that the latent reward that the principal would return in this game, if the agent did play it, is $r < c$. In this case, would the agent invest, he would lose fitness units. He is therefore under selection to keep on declining this game. Hence, cooperation does not evolve, that is the absence of cooperation is evolutionarily stable. Let us now assume, on the contrary, that the latent reward is $r > c$. In this case, would the agent play the game, he would gain rather than lose fitness units. Put differently, the latent reward constitutes a yet undiscovered benefit that the agent is under selection to collect. In this case, on the contrary, cooperation evolves.

2.4 Generalization creates latent rewards

The problem is that evolutionary game theory cannot directly tell us anything about latent rewards. They have no evolutionary ”weight”. Yet, as we will now see, their evolution can be understood in a principled manner provided one takes into account the fact that selection on actual rewards indirectly affects the evolution of latent ones. In supervised machine learning, a machine (e.g. an artificial neural network, or any other decision-making device), is trained to map a given set of (sensory) inputs, called the training set, to specific outputs. But the aim of machine learning is not only to produce machines that can produce correct answers in contexts they have already encountered. The whole point of this technique, its ”raison d'être” , comes, precisely, from the fact that machines are able to *generalize* beyond their training set. That is, machines can also produce correct decisions in situations they have never been trained to deal with. That is, situations that belong to their ”test set” but not their training set. And this ability to generalize is not a contingent, lucky, outcome of learning. On the contrary, generalization obeys a systematic logic, and occurs in specific circumstances that can be understood.

Being two adaptive processes, learning and evolution by natural selection, have very similar properties. Generalization, in particular, does have an evolutionary counterpart even though it is largely overlooked by biologists (Számadó and Szathmáry 2006). The set of situations for which a population is actually under selection corresponds the population's training set, that is the situations for which the population received feedback from the environment via natural selection. Conversely, the set of situations for which a population is not under selection, because they do not occur frequently enough to constitute a selective pressure, corresponds to the population's test set. Any response on the test set is an unintended property of the way organisms generalize beyond their training set.

In our model, the principal's training set consists of the set of games that the agent does play, conditionally or not. Throughout her life, the principal is confronted to these games and must actually decide on a reward. Hence, she receives feedback from natural selection. The principal's test set, on the other hand, consists of the set of games that the agent does not actually play. The principal never actually has to decide on a reward in these games. Hence, she never receives any feedback from natural selection. Put differently, the principal's test set consists of all the games in which her reward is only latent. The evolution, or non-evolution, of cooperation depends on the properties of these latent rewards. That is, it depends on the ability of the principal to generalize.

To study this process, we modeled the principal as an artificial neural network (ANN, see *Methods* for details). In each game, she receives as a sensory input a vector containing all the parameters of the game (b, c, r_a, and $\mathbf{s}$) and her output is a quantitative reward r, that she would give to and agent who would play this game. Previous works have already studied the evolution of social behaviours with ANNs (Arak and Enquist 1993; Enquist and Arak 1993; Johnstone 1994; André and Nolfi 2016; Debove et al. 2017). Here, we chose to use the most widespread optimization technique in supervised machine learning, called back-propagation (le Cun 1988; Rojas 1996). In addition to being strictly equivalent to the way adaptive dynamics formalizes the effect of natural selection (see *Methods*, Hofbauer and Sigmund 1990; Geritz et al. 1998), back-propagation is also a simpler and faster method than most genetic algorithms.

At first, we only let the principal evolve, keeping the agent's behavior as an exogenous constant. Namely, we assume that the agent invests unconditionally in all Selfish games, conditionally in all Interdependent games, and declines all Wasteful and Cooperative games. Figure 2.1 plots the evolution of the principal's behaviour over time. In equilibrium, the principal provides the optimal reward for every game in her training set, that is she never rewards Selfish games and always rewards just the right amount in Interdependent games (Figure 2.1: solid lines, see also Figure 2.4). What matters more to us, however, is that, as a side effect of this adaptation to games the principal actually has to play, she also rewards the agent in a fraction of Cooperative games even though she never actually encounters these situations (Figure 2.1 dashed line). That is, she generalizes.

Note that Wasteful games are also part of the principal's test set, however, they are of little interest here (still, see Figure 2.4).

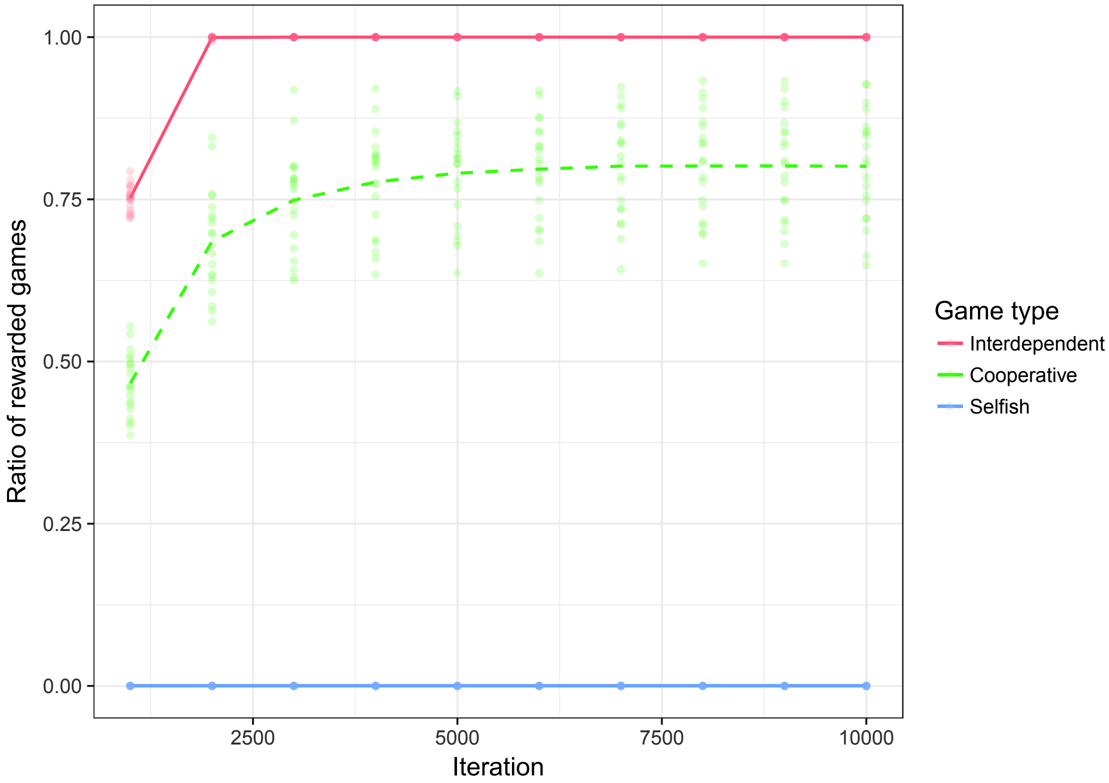


Figure 2.1: **Adaptation and generalization through evolutionary time**. The y-axis is the ratio of games of a certain type for which the principal provides a reward sufficient to motivate the agent investment. The two solid lines represent the principal behaviour on the training set: in red the ratio for Interdependent games and in blue the ratio for Selfish games. The dashed green line is the ratio for the test set (Cooperative games). There are 100 Interdependent games and 100 spurious features. Each points corresponds to a simulation, each line is the mean over 20 simulations. See *Methods* for the other parameters.

Figure 2.2 then shows how generalization depends upon circumstances. In particular, the principal almost never generalizes when only a few Interdependent games are present in her training set. That is, she rewards optimally the few Interdependent games she encounters, but in no other circumstances. In contrast, the principal generalizes more when many Interdependent games are present in her training set. That is, she does not only reward optimally the agent in Interdependent games, she is also able to calculate in all circumstances the right reward she should offer, including in circumstances she has never been selected for. The extent of generalization also depends upon the number of spurious features. The more spurious features, the largest number of Interdependent games is necessary for generalization to take place (which is a standard machine learning result; see *Discussion*).

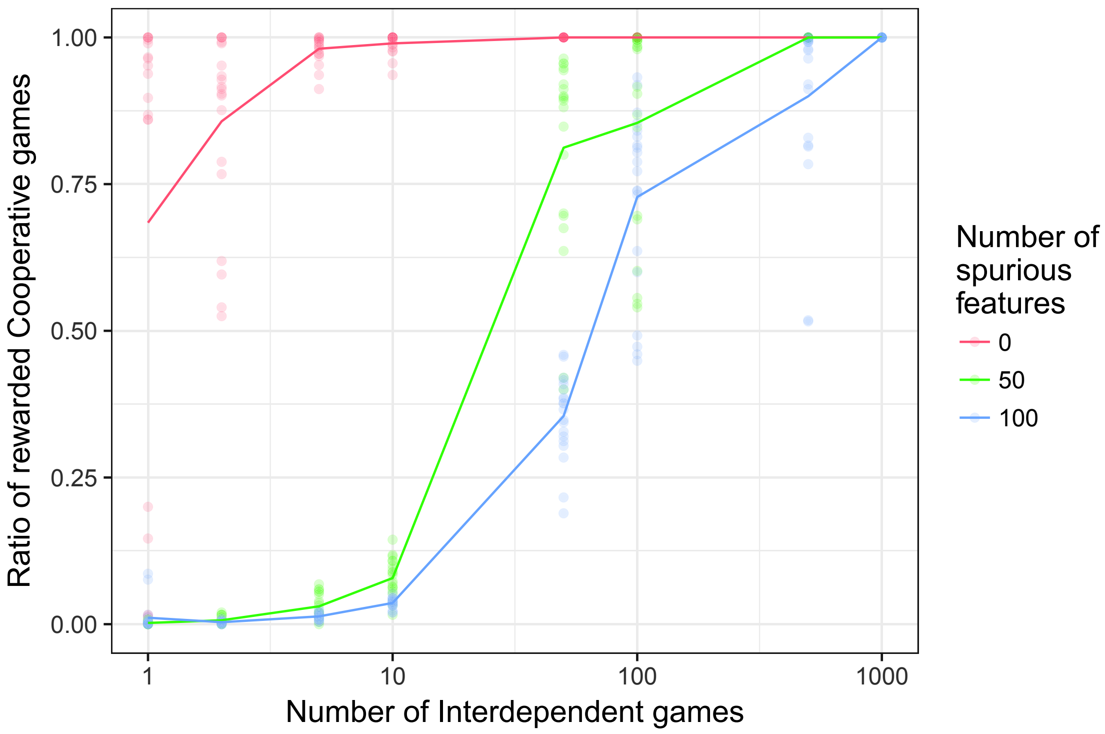


Figure 2.2: **Generalization as a function of the number of Interdependent games**. The y-axis is the ratio of Cooperative games (test set) for which the principal provides a reward that would be sufficient to motivate the agent to invest. The x-axis is the number of Interdependent games she is confronted to in her training set. Each point is a simulation result. The solid line is the mean over the 20 simulations per combination of simulation parameters. See *Methods* for the parameters.

2.5 Latent rewards foster even more generalization

Now, in addition to the evolution of the principal, we also consider the fact that the agent is shaped by natural selection to collect all the potential benefits that are present in the latent rewards of the principal. To do so, we simply assume that the agent is optimized by natural selection. That is, once a principal is fixed with a given propensity to reward, we assume that the agent evolves to take advantage of these rewards (see *Methods*). If the principal happens to possess a latent reward $r > c - r_a$ for a given game, then the agent always invests in this game because he makes a profit. Henceforth, this game now becomes part of the principal's training set, i.e. her reward is no more latent. Put differently, the principal's training set increases as more and more latent rewards are discovered by the agent, becoming actual rewards.

Figure 2.3 shows the outcome of this co-evolution (see also *Methods* for details). In comparison with the model where only the principal was allowed to evolve, we find an even stronger dichotomy between two evolutionary regimes. When the number of Interdependent games is below a threshold, generalization is almost entirely absent. Only Interdependent and Selfish games are played. When the number of Interdependent games exceeds this threshold, generalization is almost complete. Every Interdependent and Selfish game, as well as almost every Cooperative game is played.

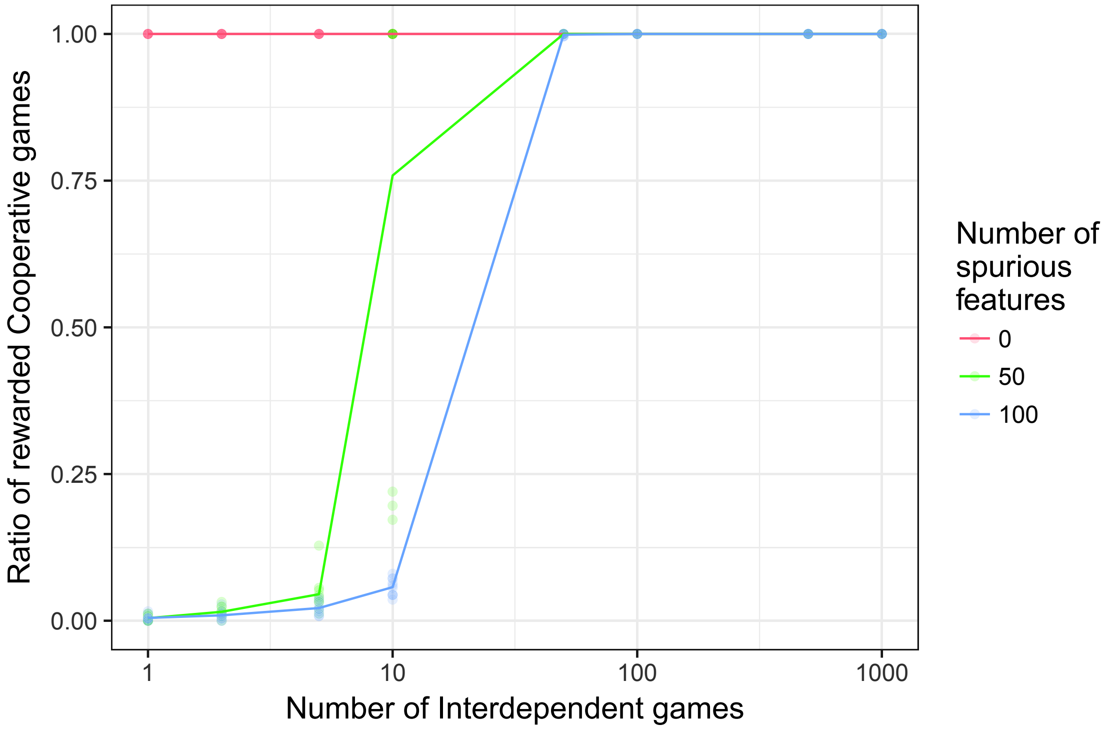


Figure 2.3: **Generalization and Cooperation co-evolve**. The y-axis is the ratio of other distinct Cooperative games (test set) for which the principal provides a reward that would be sufficient to motivate the agent to invest. The x-axis is the number of Interdependent games she is confronted to in her training set. Each point is a simulation result. The solid line is the mean over the 20 simulations per combination of simulation parameters. See *Methods* for the parameters.

2.6 Discussion

By incorporating ideas from the field of machine learning into a simple principal-agent model, we are able to propose a plausible evolutionary solution for the bootstrapping problem of conditional cooperation. We postulate that the ability to reward a cooperative investment can evolve as a by-product of the ability to provide a reward in other interdependent interactions (by-product cooperation). Using a machine learning framework, we show that, indeed, principals generalize their rewarding decisions. In other words, they are able to reward cooperative investments even though they have not been selected to do so (Figure 2.1 and 2.2). Yet, the extent to which generalization occurs crucially depends on the number of different Interdependent games present in the individuals' ecology (Figure 2.2).

The reason why a small number of Interdependent games hinder generalization is the same as in the well-known problem of over-fitting with imbalanced data in the field of machine learning. Consider a simple classification task where an ANN is trained to classify several examples, i.e. vectors of inputs, as ”positive” or ”negative”. The training set of examples is said to be imbalanced when the number of positive examples is much larger than the number of negative ones, or vice-versa. Several works in the machine learning literature have shown that imbalance of the training set has a dramatic effect on the generalization performance (Provost 2000; Weiss and Provost 2001; Japkowicz and Stephen 2002; Mazurowski et al. 2008; He and Garcia 2009). More precisely, when the number of positive examples in the training set is very small, the ANN typically learns to classify them accurately, but too specifically, reducing its classification performance for other positive examples in the test set. This situation is called over-fitting.

In our model, the principal is selected to perform a regression task and not a classification one. However, the imbalanced data problem is also present. When the number of Interdependent games is much lower than the number of Selfish games, the ANN is optimized to provide no reward at all ($r = 0$) for most of its training set examples and to provide a reward $r > c - r_a$ for only a few examples. We then expect the ANN to over-fit, i.e. to learn to recognize Interdependent games by using the spurious features which are highly specific, and not the more general payoff features b, c and r_a. When the number of Interdependent games is larger, the training set becomes less imbalanced and the ANN is less likely to over-fit. It is, therefore, likely to generalize its rewarding ability on Cooperative games (Figure 2.2). If the number of Interdependent games is high enough, generalization can be complete, in the sense that, the principal would be able to reward every possible Cooperative games even though she has never encountered them.

In a second approach, we have allowed the agent to take advantage of the principal's "latent rewards", namely its capacity to reward cooperative investments even though they never actually occur. In this case, when the principal has a latent reward in a given Cooperative game, the agent is selected to invest in this game. As a result, the training set of the principal now contains an additional game for which she is now actually selected to provide a reward. The training set is now more balanced, thus, the principal becomes less likely to over-fit: generalization begets generalization. In the end, we observe only two very distinct syndromes: complete over-fitting without generalization, versus complete generalization (Figure 2.3).

We argue that the two very distinct syndromes that our model highlights correspond to the two major observations concerning costly cooperative interactions in the living world. Many instances of conditional cooperation are observed across taxa, ranging from mutualisms to intra-specific interactions (Hart and Hart 1992; Simms and Taylor 2002; Cook and Rasplus 2003; Kiers et al. 2003; Heil et al. 2009; Archetti et al. 2011; Gomes and Boesch 2011; Kiers et al. 2011). However, they are often limited to specific and stereotyped forms. Rhizobia and legume roots only exchange chemical constituents (Simms and Taylor 2002). Impalas reward grooming by grooming in return (Hart and Hart 1992). At the other end of the spectrum, humans have a capacity to cooperate in a large variety of interactions. Our model reconciles these apparently two contradictory lines of observations by showing that the only evolutionarily stable equilibrium entails either very few reward instances at all, or a complete ability to reward cooperative investments. For the majority of species, evolution has not got rid of the "cooperative load": the reward solution has not been "discovered", or only partially. This is a consequence, ultimately, of the bootstraping problem of cooperation : the evolution of cooperation necessary entails to overcome the bootstrapping problem. As a consequence, cooperation can only evolve by the mean of generalization. Hence, by definition, only two kinds of cooperative "syndromes" are possible: no cooperation (or only in a few specific circumstances), and "complete" cooperation, with a general capacity to enforce any forms of helping.

Moreover, the non-human forms of conditional cooperation seem to result from an "exaggeration" of a previously existing form of by-product cooperation. For instance, concerning the ant-acacia mutualism, it is possible that, at first, acacias provided a benefit to ants as a by-product, then ants would be selected to take care of the acacia conditionally to these "passive rewards". Acacia would then exploit the ants' "latent preferences" by actively providing larger rewards, triggering more help from the ants. Yet, one would never observe an acacia proposing a new type of reward different from nectar or shelter, precisely because ants have no reason to have evolved a "latent preference" for it. Indeed, this is exactly what our model predicts: if too few examples of by-product cooperation are initially present, generalization is unlikely to occur because principals only pay attention to low-level, "spurious", features in order to take their rewarding decisions.

On the other hand, if principals generalize, they actually pay attention to the agent's genuine investment cost c and not anymore to the spurious features. Once this is the case, principals evolve a very broad capacity to reward agents in any possible cooperative situation. This result may provide an answer to the well-know puzzle of "folk theorem" and equilibrium selection in repeated games Conditionality can stabilized many forms of cooperation without specifying which cooperative equilibrium will actually evolve (Fudenberg and Maskin 1986; Aumann and Shapley 1994). A solution to this problem has been proposed in humans by the means of group selection between populations that exhibit different cooperative equilibria (Boyd and Richerson 1992; Boyd 2006). Here, we show that generalization is the only way for evolution to actually reach some cooperative equilibria. It follows that only two situations are possible: either no cooperative equilibrium is achieved, or the full range of cooperative interactions are enforced because generalization is complete.

According to our results, human enforced cooperation evolved as a generalization of abilities that initially evolved to deal with *interdependent* activities (see also Tomasello et al. 2012). This entails that, in a way or another, the "human cooperative syndrome" is a consequence of the fact that we were engaged in more interdependent activities than other species. That is, we were very often in situations in which (i) one's behaviour had a positive, by-product, effect on con-specifics and (ii) the eventual benefit of this behaviour depended upon others' quality. In practice, this probably corresponds essentially to situations of "collective action", that is situations in which several individuals act for a common good that they reach through a behavioural interaction. Our result, therefore, then raises the new question: Why would these interactions be more frequent in the Human ecology than in the ecology of other species? There are several possible answers that come to mind but, eventually, this question should be studied with care in order not to confound proximate and ultimate explanations. Hence, we refrain from answering it here.

2.7 Methods

Initialization. We run simulations coded into Python. In each simulation, a set of games is generated whose payoff parameters and spurious features are randomly chosen. 500 games of each types are generated, except for Interdependent games whose number vary between 1 and 1000 depending on the simulation. A principal, modelled as an Artificial Neural Network is generated with a single hidden layer of 5 neurons and with randomly chosen synaptic weights. For each vector of inputs (payoff parameters and spurious features of a single game), the ANN's output is the principal's reward value r for this game.

Pre-selection phase. It can happen that, by chance, a randomly generated ANN provides a large enough reward to the agent in some Cooperative games ($r > c$). This would be an artefact of our initialization procedure. To avoid these, we first train the ANN to provide a null reward in every game.

Selection phase. We then train the ANN to provide the optimal reward in each game of the training set as explained in section *Latent rewards are pivotal in the evolution of cooperation*. Note that the ANN is not trained in the games in which the agent never invests.

Gradient descent. There is a formal equivalence between natural selection and simple forms of learning (Watson and Szathmáry 2016). For instance, many machine learning methods are based on gradient descent. This principle is also used in a class of models which have been extensively used to describe the evolutionary process: adaptive dynamics (Hofbauer and Sigmund 1990; Geritz et al. 1998). In the case of natural selection, the organisms' traits gradually evolve towards maximisation of fitness. In the case of gradient descent, the ANN's synaptic weights are gradually modified according to the gradient of the error function which quantifies how accurate the ANN is. Both mechanisms lead to decision machines which provide the accurate output for every set of inputs that they have encountered.
In our simulations, we use a simple back-propagation algorithm, i.e. gradient descent applied to ANNs. More precisely, we use the mini-batch gradient descent with a batch size of 50 and a learning rate $\alpha = 0.2$. That is, for each iteration, we compute the ANN's responses for 50, randomly chosen, games (i.e. vectors of inputs). We then compute the mean squared error over the 50 games. For each game, the error is the difference between the reward that is produced by the ANN and the adaptive reward that should evolve by natural selection (see section *The principal's evolutionarily stable strategy*). Lastly, we updated each synaptic weight of the ANN in the direction of the gradient towards error minimization (le Cun 1988; Rojas 1996), in proportion of the gradient intensity and of the learning rate α.

Test set. At the end of the selection phase, we test the ability of the principal to reward in games that she has never encountered (test set). Typically, we generate 500 other Cooperative games and record the output of the ANN for each of these games (Figure 2.2 and 2.3).

The agent's evolutionarily stable strategy

The agent must decide, for every game, whether he wants to (1) decline, (2) play unconditionally or, (3) play conditionally on the principal's quality. His evolutionarily stable strategy (ESS) in this setting can be understood as follows.

Monitoring the principal in a given game can only be worth it if the sign of the agent's net payoff in this game is uncertain (that is, if $r_a + r - c$ is sometimes positive and sometimes negative in the same game, depending on the principal's quality). Otherwise, it is always better for the agent to spare the monitoring cost, and play or decline the game unconditionally. If the sign of $r_a + r - c$ is uncertain, on the other hand, the agent's optimal strategy is less straightforward. In the sake of simplicity, throughout this paper, we aim to choose parameters such that monitoring is always worth it in this case.

Formally, consider a given game with an investment cost c and a variable return on investment $x = r + r_a$, where x is drawn from a continuous density distribution $\phi(\cdot)$. It is straightforward to show that, in this game, investing conditionally is better than declining iff $m < \int_c^{+\infty} (x - c)\phi(x)dx$, and better than investing unconditionally iff $m < \int_0^c (c - x)\phi(x)dx$. That is, investing conditionally is worth it if the monitoring cost m is lower than both the expected benefit and the expected cost of the game. In practice, throughout the paper, we assume that the monitoring cost is always sufficiently low for all Interdependent games to satisfy this condition. And we also assume that, once the principal rewards a given Cooperative game, the value of this reward remains uncertain (due to some unspecified variability) such that monitoring always remains adaptive.

In consequence, the evolutionarily stable strategy (ESS) of the agent is the following.

1. In all games where $r + r_a - c$ is always negative, the agent must decline unconditionally.
2. In all games where $r + r_a - c$ is always positive, the agent must play unconditionally.
3. In all games where the sign of $r + r_a - c$ is uncertain, the agent must pay m to monitor the principal's quality and then play the game iff $r + r_a - c > 0$

Note that all the agents' decisions when $r + r_a - c = 0$ are neutral.

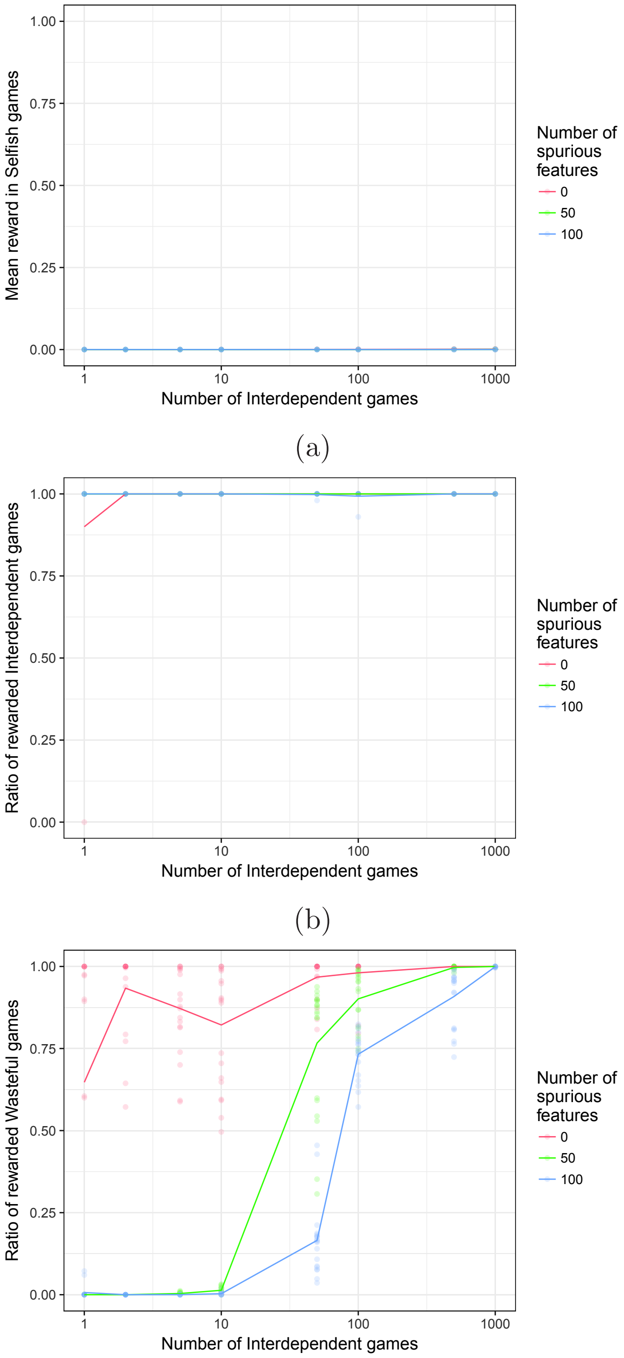
In our analysis, we assume that the agent is well optimized by natural selection. That is, when a given principal is fixed as the resident, the agent always plays the above optimal strategy.

The principal's evolutionarily stable strategy

The principal should actively produce a reward only when it is worth it, that is only when the agent takes her reward into account to make a decision. Accordingly, the evolutionarily stable strategy of the principal is the following.

1. In all games where the agent will, anyway, decide to play unconditionally, the principal should offer no reward at all.
2. In all games where the agent will play conditionally on the sign of $r + r_a - c$, the principal should offer a reward $r = c - r_a + \epsilon$, where $\epsilon > 0$ is just sufficient to incentivize the agent to play the game.
3. In all games where the agent will, anyway, decline unconditionally, the principal's reward is neutral because it never actually occurs. It constitutes a "latent" facet of the principal's strategy.

2.8 Supplementary Figures

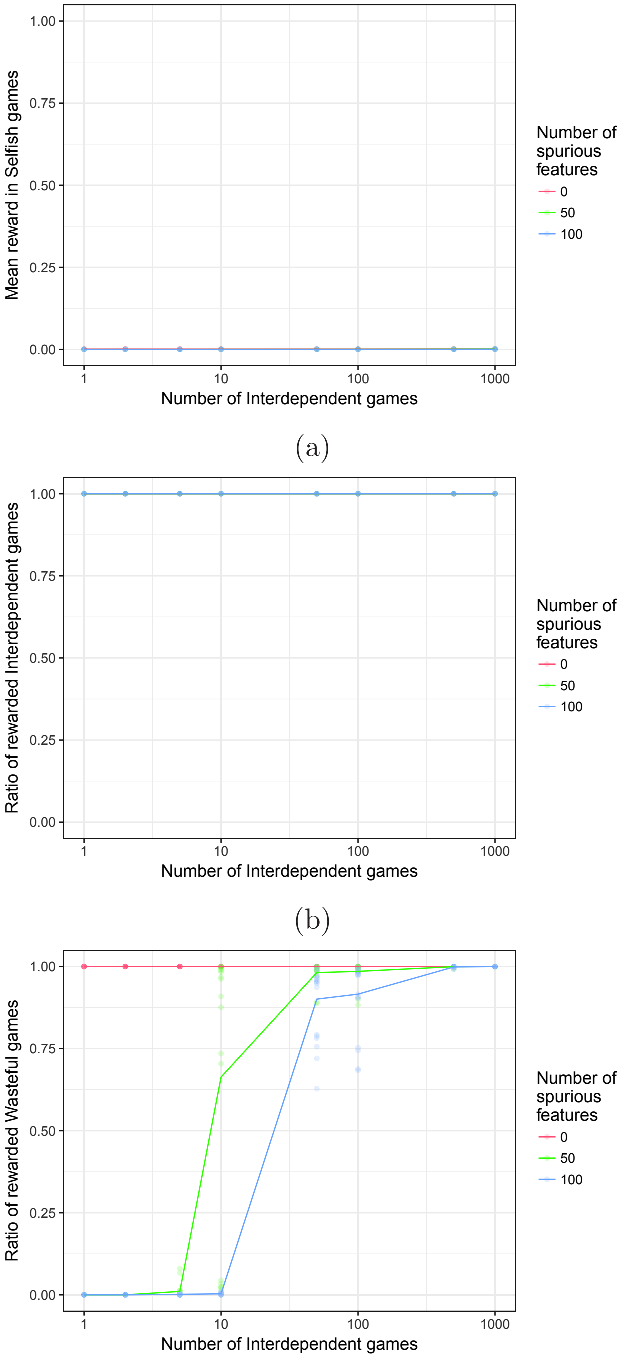


(a)

(b)

(c)

Figure 2.4: **Additional results: performance on the training set and generalization in the test set.** Cooperative games are never played. The x-axis is the number of Interdependent games she is confronted to in her training set. The y-axis is (a) the mean reward that the principal provides for Selfish games, or the ratio of (b) Interdependent games or (c) Wasteful games for which the principal provides a reward that is sufficient to motivate the agent to invest. Each point is a simulation result. The solid line is the mean over the 20 simulations per combination of simulation parameters. See *Methods* for the parameters.



(a)

(b)

(c)

Figure 2.5: **Additional results with co-evolution: performance on the training set and generalization in the test set.** Cooperative games are played. The x-axis is the number of Interdependent games she is confronted to in her training set. The y-axis is (a) the mean reward that the principal provides for Selfish games, or the ratio of (b) Interdependent games or (c) Wasteful games for which the principal provides a reward that is sufficient to motivate the agent to invest. Each point is a simulation result. The solid line is the mean over the 20 simulations per combination of simulation parameters. See *Methods* for the parameters.

Chapter 3

The evolution of cooperation by exaggeration

This chapter is not an article, but rather a work in progress.

3.1 Introduction

In the previous chapter, we have proposed a model to address the problem of the *origin* of conditional cooperation. Using a multi-game approach, we have shown how conditional cooperation can evolve by generalization from interdependent forms of cooperation.

However, in our previous model, we have discretized each cooperative interaction in a distinct game. This assumption cannot account for the fact that some cooperative interactions are more similar to each other than they are to other interactions. A trivial example is the fact that ”investing $x + \epsilon$ units of nectar” is more similar to ”investing x units of nectar” than it is to ”providing a shelter”. As a consequence, when an organism is selected to produce the adaptive reward for a set of situations, it will more likely generalize its behaviour to similar situations than to very different ones. Here, we are referring to ”similar” opportunities not in the sense of yielding similar payoffs, but rather in the more proximate sense of being detected by the same sensory inputs.

One example of generalization to similar forms of cooperation is the phenomenon of *exaggeration*. If individuals are selected to produce a behaviour for a given range of values of a sensory input, they are most likely to generalize their behaviour to other values of the same sensory input, even though they have never encountered them. This phenomenon is well-known in sexual selection, for instance, where females who are selected to choose larger males generalize their preferences to ”exaggerated” values of size (Ryan 1990). Males can then "exploit" the latent preferences, which can lead to a runaway of exaggeration (Ryan and Keddy-Hector 1992; Arak and Enquist 1993).

Similarly, this phenomenon could be at work in cooperative interactions as well. Many instances of conditional cooperation in mutualisms seem to be exaggerated forms of pre-existing by-product cooperation. For instance, in the ant-acacia mutualism, if ants get a benefice from nectar production, and if they protect the tree more or less efficiently, depending on the nectar quantity, then acacias might incentivize the ants by providing even more nectar. This type of interaction has sometimes been called pseudo-reciprocity (Leimar and Connor 2003). Many examples of trade might be the result of exaggeration from an initial state of an "accidental market" (Sachs and Hollowell 2012; Morris and Schniter 2018), i.e. a situation of commensalism, or by-product cooperation. For example, if two bacteria colonies benefit from each others' wastes, they might evolve a conditional mechanism to provide a benefit to each other to increase their waste depending on the other's condition.

In this chapter, we describe a continuous version of the previous principal-agent model meant to capture the role of quantitative exaggeration in the evolution of cooperation in non-human species in which "across-the-board" generalization did not take place.

3.2 Methods

A continuous principal-agent model

We use a continuous principal-agent model. We assume that agents and principals belong to two distinct populations, respectively of size N_A and N_P. Principals are defined by a costly quality level q. Every time an agent meets a principal, the agent can benefit from the principal's quality by making a costly investment x. In return, she receives a benefit of size $q \cdot x$ and the principal receives a benefit as a by-product of the agent's investment. For each interaction, $P_A(x)$ and $P_P(x)$ are respectively the agent's and the principal's payoff:

$$P_A(x) = qx - cx^2$$
$$P_P(q) = bx - q^2.$$

By default, principals are selected to have the minimal possible quality $q = 0$.

We suppose that agents have the capacity to obtain information about the principal's quality q. Assuming that information is perfect, for each level of quality q, the agent's optimal investment $x^*(q)$ is straightforward:

$$\frac{\partial P_A}{\partial x}(x^*) = 0 \iff x^*(q) = \frac{q}{2c}. \tag{1}$$

The higher the quality, the larger the optimal investment.

If we suppose that agents behave optimally for any value of q, principals are now under selection to increase their quality, in order to attract the agents' investments. Under the condition that an agent always invests $x^*(q) = q/2c$, the optimal quality level q^{**} for a principal is:

$$\frac{\partial P_P}{\partial q}(q^{**}) = 0 \iff q^{**} = \frac{b}{4c}. \tag{2}$$

Therefore, under the unrealistic assumption that agents are "Darwinian demons", the evolution of conditional cooperation by exaggeration is possible: the principals are selected to take advantage of the agents' plasticity in order to always make them invest more, up to the point for which the quality becomes too costly. This process corresponds to the evolution of cooperation by exaggeration.

Modelling the agent's reaction norm

However, natural selection is unlikely to produce "Darwinian demons" that behave optimally for any values, including ones that they have never been selected to deal with. If all principals have an initial value of $q = 0$, then the agent's "reaction norm" should be $x^*(0) = 0$, but is neutral for every value $q > 0$. This neutral, or "latent", portion of the agents' reaction norm nonetheless can have a dramatic impact on the evolution of the principal's quality. Indeed, for a given value of the quality q, and a given reaction norm $x(q)$, the selection gradient for quality is positive if

$$\frac{\partial P_P}{\partial q}(q) > 0 \iff \frac{\partial x}{\partial q}(q) > \frac{2q}{b}. \tag{3}$$

In other words, if the derivative of the reaction norm $x(q)$ is large enough, it is adaptive for a principal to slightly increase her quality, so that she can benefit from a slightly larger investment.

The evolution of the rewarding capacity of the principal, hence, relies on the "hidden" shape of the agent's reaction norm. Standard game theory, however, cannot make predictions about such neutral aspects of behaviours. We will thus use concepts from machine learning to answer our problem.

First, we assume that, for a reason or another, the principal's quality is noisy. That is, principals are defined by an evolving trait which is the average value $\bar{q}$ of quality.

At birth, and randomly, some principals express a slightly higher or slightly lower value of quality q in a given range $[\bar{q} - \sigma, \bar{q} + \sigma]$ (with the additional condition $0 \leq q \leq 1$). In this case, agents are selected to provide the optimal investment for every value $q \in [q_{min}, q_{max}]$. We are interested in the following question: how do agents generalize outside this range? More precisely, we want to investigate how selection on a given range of sensory inputs impacts the response to other input values to which individuals have not been selected to respond. In machine learning jargon, as in the previous chapter, we want to understand how optimization on a "training set" of inputs can impact the generalization on another, distinct "test set" of inputs.

The agents' reaction norm is modelled using an Artificial Neural Network with q as input, one hidden layer of 5 neurons, and which provides the investment decision $x(q)$ as output. The synaptic weights of the Artificial Neural Network are the agent's evolving traits.

We need to model the co-evolution between the agent's reaction norm $x(q)$ and the principal's mean quality $\bar{q}$. For this aim, we use two different simulation methods.

First, we use a classic Moran process with non-overlapping generations. For every generation, all the agents interact with all the principals sequentially and, for each interaction, an agent makes an investment decision depending on the quality q of the principal she is interacting with. Each individual's fitness is proportional to the sum of the payoffs in every encounter she has made during her life. The mean quality $\bar{s}$ and each synaptic weight can mutate with a probability μ.

Second, we use a machine learning technique. We update the synaptic weights using a back-propagation algorithm. A single agent, modelled as an Artificial Neural Network, interacts with all the principals sequentially and, each time, makes an investment decision. For each synaptic weight, the gradient in the direction of the reduction of the error, or the augmentation of fitness, is derived using back-propagation (le Cun 1988; Rojas 1996). The weights are then updated by a small increase in the direction of the selection gradient. As we have seen in the previous chapter, back-propagation is very similar to adaptive dynamics (Hofbauer and Sigmund 1990; Geritz et al. 1998). In parallel, we model the principal as in the previous case: an evolving population which is submitted to selection and mutation a the end of every generation. This "hybrid technique" might seem rather strange, however, we will see that it yields quantitatively similar results than the classic Moran process.

The two methods will allow us to investigate both (i) the agents' capacity to generalize to situations that they have never been selected to deal with, and (ii) if generalization can lead to the exaggeration of the principal's quality.

We start every simulation with a "pre-selection" phase, in which we arbitrarily select the agents (or train the single agent in the case of back-propagation) to provide an investment $x(r) = 0 \quad \forall r$. In the "selection" phase, we then run long enough simulations for an equilibrium to be reached.

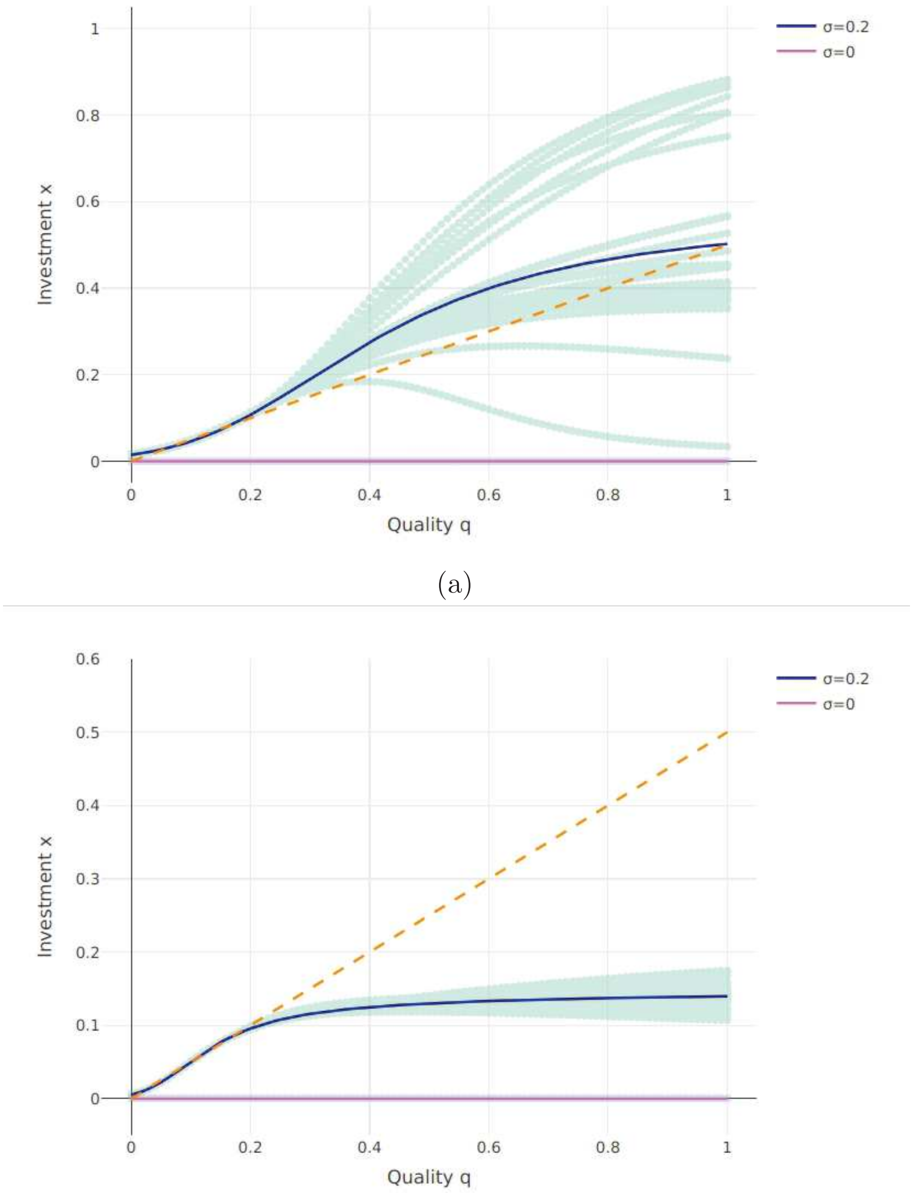
3.3 Results

Investigating the agent's generalization

We first present simulations in which the principal population does not evolve. Namely, the mean quality trait $\bar{q}$ of the principal is fixed at $\bar{q} = 0$. Thus, considering the phenotypic noise of principals, agents only encounter principals with quality $q \in [0, \sigma]$. Agents (or the single agent in the case of back-propagation) are under selection to make the adaptive investment $x^*(q) = \frac{q}{2c}$ only for $q \in [0, \sigma]$. Outside this range, the investment decision is neutral, or latent.

Figure 3.1 shows the agent's reaction norm at the equilibrium for both optimization techniques, and for two values of σ. For $\sigma = 0$, there is no variability of quality among principals: the quality is always zero. We therefore observe the reaction norm shaped during the pre-selection phase. As expected, with both optimization techniques, the resulting reaction is flat ($x(r) = 0 \quad \forall r$). In contrast, for $\sigma = 0.2$, the agent's reaction norm at the equilibrium is an increasing function of q. More precisely, for the range $q \in [0, 0.2]$, i.e. the training set, the reaction norm is very close to the theoretical optimal investment $x^*(q) = \frac{q}{2c}$ (Figure 3.1, orange dashed line). We also check that an equilibrium is reached, by following the evolution of the "error" agents make in their investments (Figure 3.2).

As expected, outside of the training set, namely on the "test" set $q \in [\sigma, 1]$, the reaction is very different from the optimal investment strategy (Figure 3.1). Moreover, the results are highly variable across simulations. These results are not surprising, since there is no selection for a particular investment for this range of qualities. However, at the inferior boundary of the test set ($q = \sigma + \epsilon$), the reaction norm is very similar to the reaction norm on the training set. By continuity, agents at least partially generalize their investment strategy for qualities that are close to the training set: they invest more when the quality is higher (equation (1)). Consequently, a principal with a quality slightly higher than σ will benefit from a larger investment and will be favoured by natural selection (equation (3)).



(a)

(b)

Figure 3.1: **The agent reaction norm on the training set and on the test set at the equilibrium.** **(a)** The reaction norm evolves by natural selection. Each point represents the mean investment value over all agents in a single simulation. The solid line is the mean investment value over 30 simulations. **(b)** The reaction norm "evolves" via back-propagation. Each point represents the investment for a simulation. The solid line is the mean investment 30 simulations. The dashed orange line is the optimal reaction norm $x^*(q) = \frac{q}{2c}$. Parameters are $c = 1$; $N_P = 100$; $\mu = 0.001$; and **(a)** $N_A = 100$

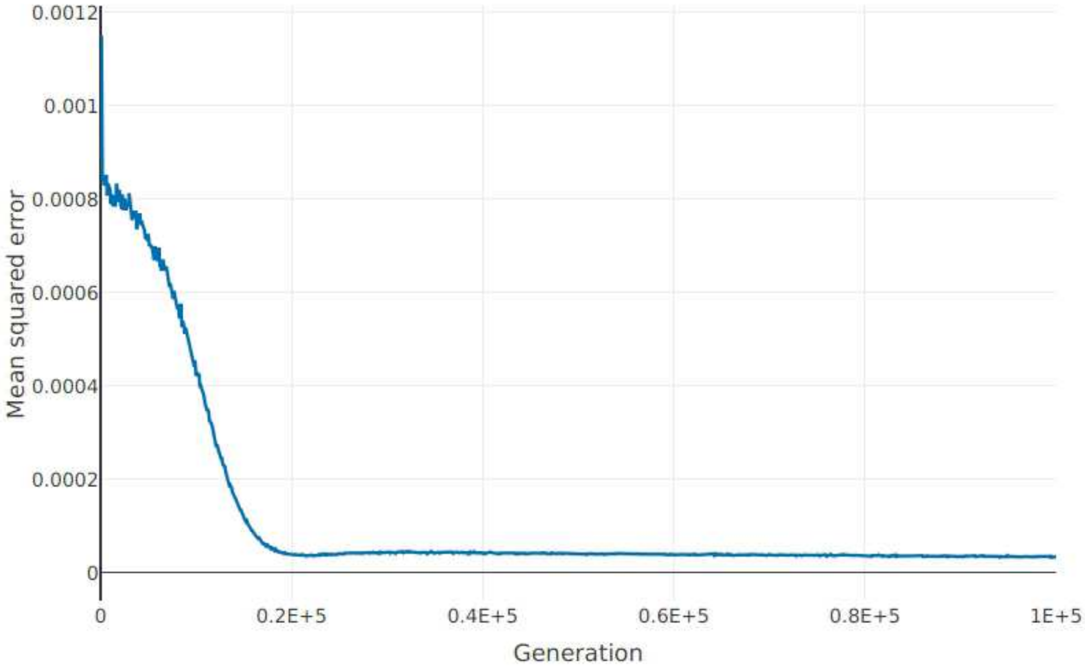


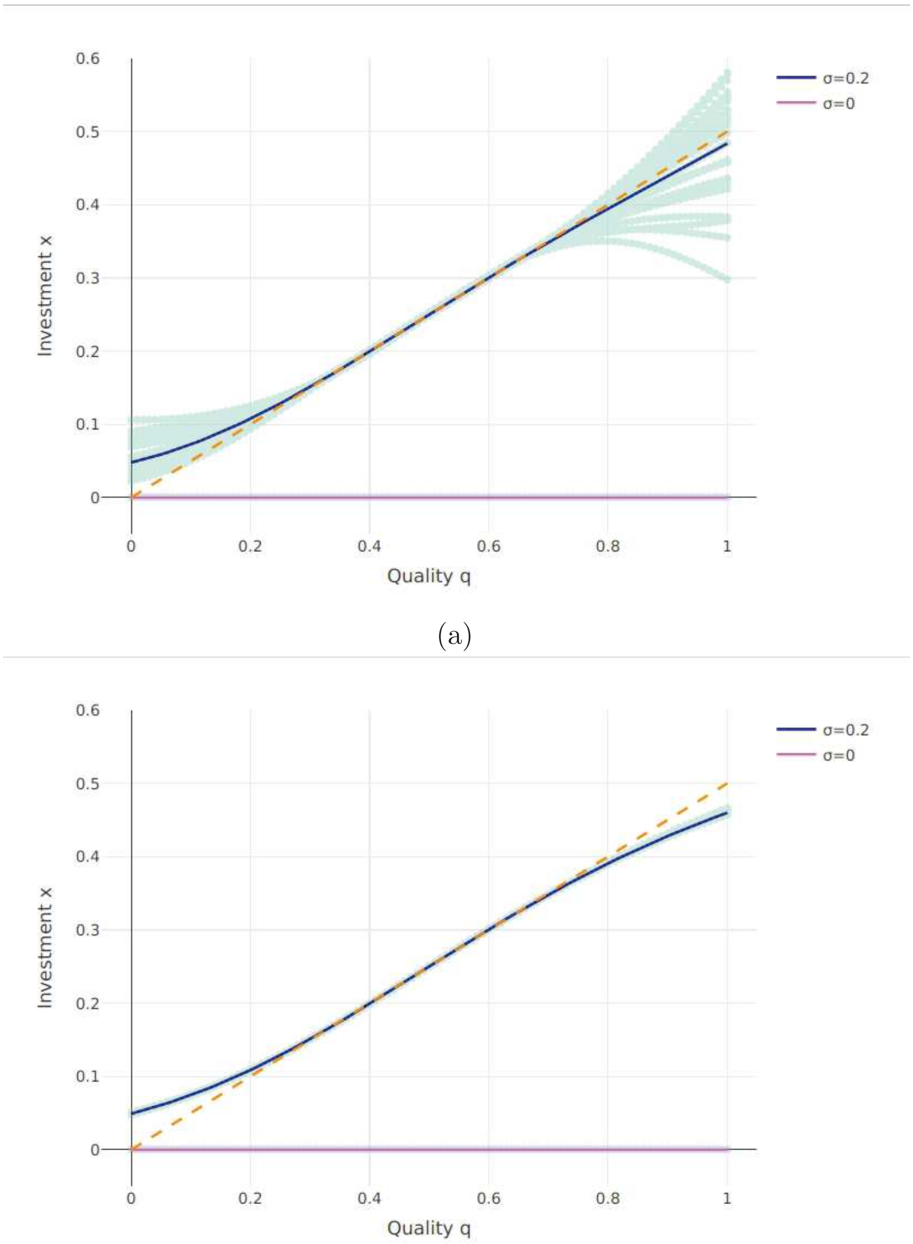
Figure 3.2: **Evolution of the agent's error on the training set.** The mean squared error is shown for the back-propagation case. For each principal with quality q, the error is the difference between the output of the Artificial Neural Network (the agent's investment) and the optimal investment $x^*(q) = \frac{q}{2c}$. At every generation, the mean squared error is computed over all principals in the population. Parameters are the same as in Figure 3.1.

Exaggeration of quality

We now allow the principal's mean quality trait $\bar{q}$ to co-evolve with the agent's reaction norm. In this situation, the ”training set” of the agent is no longer fixed since it is determined by the qualities of the principals population. Put differently, there is an evolutionary feedback between the two populations: if the mean quality changes, agents are under selection to provide the optimal investment on a new range, thus, they might generalize even more, which would impact the evolution of the mean quality in return.

The results of this particular co-evolutionary process are shown in Figure 3.3 for the agents, and in Figure 3.4 for the principals. The co-evolutionary process goes as follow. At a given time, all the principals' qualities are in a given range $[\bar{q} - \sigma, \bar{q} + \sigma]$. Agents are selected to provide the optimal investment only for $q \in [\bar{q} - \sigma, \bar{q} + \sigma]$. As we have seen, if agents respond optimally, a mutant principal with a slightly higher quality is favoured, as long as equation (3) holds. Agents are now under selection to behave optimally on a new range. This process goes on until equation (3) stops being true. Indeed, Figure 3.4 shows that principals evolve always higher quality values until the dynamics stabilizes around $\bar{s} = 1/2$, which is precisely what we would have predicted in the case of optimal individuals (equation (2)).

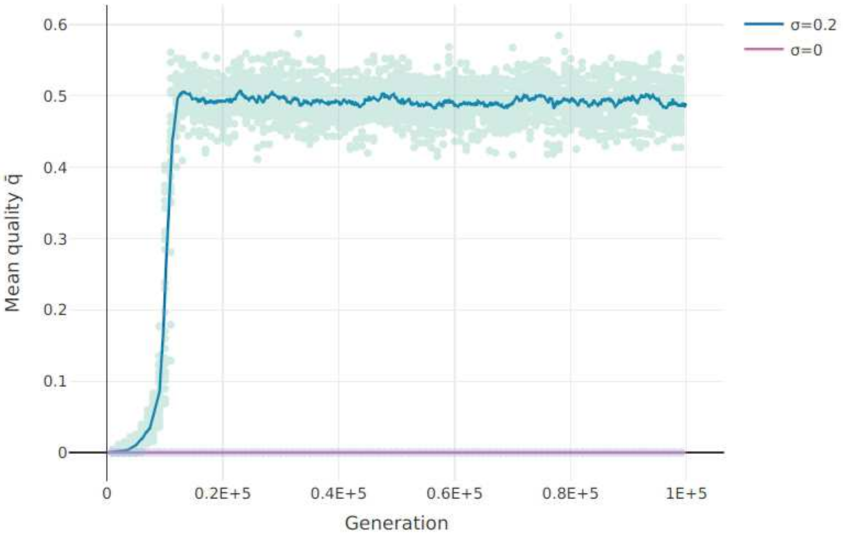
In the absence of variability in the principals population ($\sigma = 0$), on the other hand, neither the investment reaction norm, nor the mean quality evolve. They remain at their initial values (Figures 3.3 and 3.4).



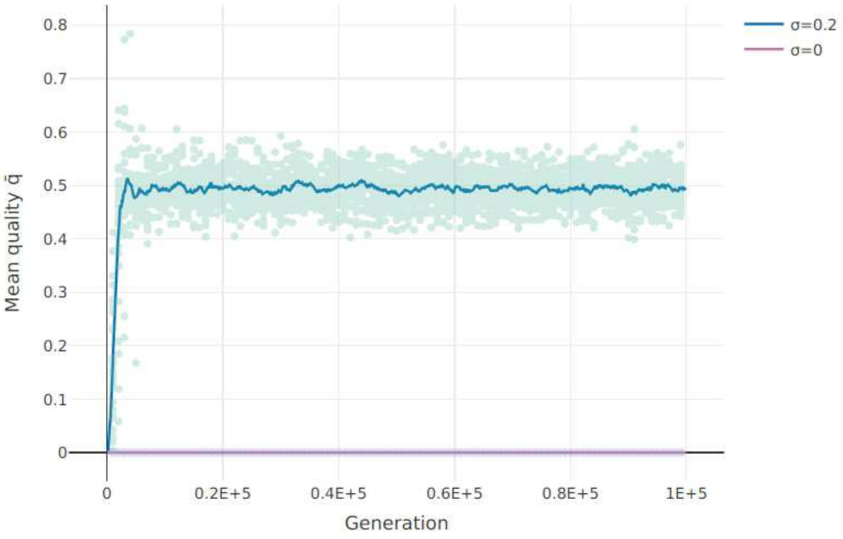
(a)

(b)

Figure 3.3: **The agent reaction norm when co-evolution is allowed. (a)** The reaction norm evolves by natural selection. Each point represents the mean investment value over all agents in a single simulation. The solid line is the mean investment over 30 simulations. **(b)** The reaction norm "evolves" via back-propagation. Each point represents the investment for a simulation. The solid line is the mean investment over 30 simulations. The dashed orange line is the optimal reaction norm $x^*(q) = \frac{q}{2c}$. Parameters are the same as in Figure 3.1.



(a)



(b)

Figure 3.4: **Evolution of the principal's mean quality $\bar{q}$.** Each point represents the average $\bar{q}$ over all principals in a single simulation. The solid line is the average over 30 simulations. **(a)** The reaction norm evolves by natural selection. **(a)** The reaction norm ”evolves” via back-propagation. Parameters are the same as in Figure 3.1.

3.4 Discussion

We have built a continuous principal-agent model to study the evolution of cooperation by exaggeration. We have supposed that, initially, principals benefit from the agents' investments as a by-product. Yet, agents are under selection to make the optimal investment in a given range of qualities: the higher the quality, the higher the investment (equation (1)). By continuity, agents are very likely to generalize their investment strategy to quality values that they have never been selected to deal with, but that are close to the "training set" (Figure 3.1).

In consequence, principals are now under selection to take advantage of the agents' "latent" preferences. The mean quality level thus evolves towards higher values (Figure 3.4). A co-evolutionary process thus takes place: the evolution of the range of existing qualities shapes the evolution of new latent preferences in the agent (Figure 3.2), which, again, generate a selective pressure for higher levels of quality. This exaggeration process goes on until the principals stop benefiting from increasing their quality, i.e. when the cost of quality becomes higher than the by-product benefit of investment (as predicted by "Darwinian demons", equation (2)).

Hence, the agents' latent preferences are part of the principals' adaptive landscape. In other words, principals "manipulate" the agents' plasticity to increase the by-product benefits they derive from their investments. In the end, from a pure situation of by-product cooperation, exaggeration can take place and leads to a situation of enforced cooperation, in which principals actively "reward" the agents' investments. Our model can be viewed as a continuous version of pseudo-reciprocity (Connor 1986; Leimar and Connor 2003; Leimar and Hammerstein 2010).

This model of "continuous generalization" completes our previous, more general, model of generalization (Chapter 2). Our previous model, indeed, could not account for many instances of generalization observed in non-human conditional cooperation, especially in mutualisms. It seemed to make the prediction that "no generalization" and "full generalization" are the only two possible outcomes of evolution. Our exaggeration model can account for the possibility of generalization over a single, continuous, dimension.

Part II

Insights from partner choice theory

Chapter 4

Competitive altruism and the runaway of cooperation

4.1 Objectives and summary

In this part, we focus on one specific type of conditional cooperation called partner choice. Partner choice occurs when individuals can refuse to interact with some partners and interact with others instead. It is a very general mechanism and is not restricted to cooperation. Other social interactions are based on partner choice as well, such as mate choice in the context of sexual selection (Noë 2017). In this Chapter, we will focus on partner choice in the context of cooperation, whereas, in Chapter 5, we extend our analysis to biological markets in general. A biological market is a population of individuals who can exchange benefits and choose their partners. Many forms of human cooperation seem to satisfy both criteria (Barclay 2013).

Partner choice can enforce cooperation in a very simple an intuitive way. If individuals can freely choose their partners, and if a variability of cooperative types co-exist in the population, then, natural selection will favour ”choosy” individuals who preferentially interact with more generous partners in cooperative interactions (Bull and Rice 1991; Noë and Hammerstein 1994, 1995). In return, the most cooperative types will increase in frequency, since they are more likely to be chosen, and therefore to receive the benefit of cooperation. This process is also called competitive altruism (Roberts 1998) and leads to a gradual runaway of cooperation towards higher values.

In this Chapter, we review the existing models of partner choice and argue that the literature lacks an extensive framework to address one specific question: at which level of cooperation does the runaway stop? Intuitively, it should stop when the cost of increasing the level of cooperation becomes larger than the benefit of attracting partners. Nevertheless, biological markets are complex, and deriving these values is not an easy task. The question of the end point of the runaway is all the more interesting that, when cooperation yields diminishing returns, the level of cooperation determines the social efficiency of the interaction.

We propose two models to study the runaway of the level of investment in cooperation. In both models, we analyse the joint evolution of the individuals' choosiness and the individuals' noisy level of investment in cooperation. Our first model is a classic adaptive dynamics model. It predicts that, under perfect partner choice, the runaway stops at very high levels of investment, where the social efficiency of cooperation is close to zero. We show that this result is due to a particular assumption, which prevents the individuals to behave conditionally upon their own level of cooperation. In a second model, we relax this assumption and, by using agent-based simulations, we show that, under perfect partner choice, the runaway stops at the level that maximizes social efficiency. We therefore argue, first, that partner choice can explain a fundamental property of human cooperation, which is the preferences for mutually beneficial forms of interactions, and, second, that modelling partner choice is not a trivial task, but that, fortunately, an extensive literature in economics is suited to do so. We will follow this lead in Chapter 5.

The rest of this chapter comes from a paper peer-reviewed and recommended by *Peer Community in Evolutionary Biology*:

Geoffroy, F., Baumard, N., & André, J.-B. (2019). Why cooperation is not running away. bioRxiv, ver. 5 peer-reviewed and recommended by *PCI Evol Biol*. doi: 10.1101/316117

Why cooperation is not running away

Abstract

A growing number of experimental and theoretical studies show the importance of partner choice as a mechanism to promote the evolution of cooperation, especially in humans. In this paper, we focus on the question of the precise quantitative level of cooperation that should evolve under this mechanism. When individuals compete to be chosen by others, their level of investment in cooperation evolves towards higher values, a process called competitive altruism, or runaway cooperation. Using a classic adaptive dynamics model, we first show that, when the cost of changing partner is low, this runaway process can lead to a profitless escalation of cooperation. In the extreme, when partner choice is entirely frictionless, cooperation even increases up to a level where its cost entirely cancels out its benefit. That is, at evolutionary equilibrium, individuals gain the same payoff than if they had not cooperated at all. Second, importing models from matching theory in economics we, however, show that, when individuals can plastically modulate their choosiness in function of their own cooperation level, partner choice stops being a runaway competition to outbid others, and becomes a competition to form the most optimal partnerships. In this case, when the cost of changing partner tends toward zero partner choice leads to the evolution of the socially optimum level of cooperation. This last result could explain the observation that human cooperation seems to be often constrained by considerations of social efficiency.

4.2 Introduction

Cooperation among non-kin constitutes a puzzle for evolutionary biologists, and a large body of theoretical models, inspired by game theory, have been developed to solve it. The most commonly accepted explanation is that cooperation can be enforced if it triggers a conditional response on the part of others (West et al. 2007b). Several enforcement mechanisms have been proposed: direct reciprocity (Trivers 1971; Axelrod and Hamilton 1981; Lehmann and Keller 2006), indirect reciprocity (Nowak and Sigmund 1998, 2005; Leimar and Hammerstein 2001), punishment (Boyd and Richerson 1992; Boyd et al. 2003; Bowles and Gintis 2004) and partner choice (Bull and Rice 1991; Noë and Hammerstein 1994, 1995; Sachs et al. 2004). A growing number of experimental studies support the idea that, among this set of mechanisms, partner choice is likely to be particularly influential in nature, both in inter-specific and in intra-specific interactions (Bshary and Schäffer 2002; Simms and Taylor 2002; Kiers et al. 2003, 2011; Fruteau et al. 2009; Schino and Aureli 2009; Hammerstein and Noë 2016). Besides, partner choice is also believed to play a major role in human cooperation, where friendships and coalitions are common (Barclay 2013, 2016; Baumard et al. 2013; and see Discussion).

The key idea of partner choice models is that, when one happens to be paired with a defecting partner, one has the option to seek for another, more cooperative, partner present in the "biological market" and interact with her instead of the defector. This possibility allows cooperators to preferentially interact with each other, and, consequently, prevents any invasion by free-riders (Eshel and Cavalli-Sforza 1982; Bull and Rice 1991; Noë and Hammerstein 1994, 1995; Ferriere et al. 2002; Bergstrom 2003; Aktipis 2004, 2011; Sachs et al. 2004; Fu et al. 2008; Barclay 2011).

So far, the primary objective of most partner choice models has been to explain how *some* cooperation can exist at all in an evolutionary equilibrium. On this ground, models have reached a clear answer: partner choice can trigger the evolution of cooperation. In this paper, however, we are interested in another issue that models generally consider with less scrutiny: that of understanding the quantitative *level* of cooperation that should evolve under partner choice.

This analysis is crucial because the quantitative level of cooperation determines the "social efficiency", also called the Pareto efficiency, of interactions. Cooperating too little is inefficient because individuals miss some opportunities to generate social benefits. But cooperation, as any investment, is likely to have diminishing returns (Altmann 1979; Weigel 1981; Killingback and Doebeli 2002). As a result, there is a "socially optimal" amount of cooperation, an intermediate level where the sum of the helper and helpee's payoff is maximized. Cooperating more than this amount is hence also inefficient, because it increases more the cost of cooperation than it raises its benefit. In the extreme, there is even a "wasteful" threshold beyond which the overall cost of cooperation becomes larger than its benefit. If two partners cooperate more than this threshold, the net benefit of their interaction is negative, that is they are both worst off than if they had not cooperated at all.

Prima facie, partner choice appears to be a unidirectional pressure acting on the evolution of cooperation, unlikely to generate an intermediate equilibrium. Competition to be chosen by others, called "competitive altruism" (Roberts 1998; Hardy and Van Vugt 2006; Nesse 2009), should lead to a runaway of cooperation, as it does in sexual selection (West-Eberhard 1983). In principle, this runaway should proceed up to the point where the cost of investing into cooperation cancels out the benefit of finding a partner (West-Eberhard 1979; Fisher 1999 p.152) that is up to the "wasteful" threshold where cooperation becomes fruitless. Is competitive altruism, however, balanced by opposite forces, leading to an evolutionary stabilization of cooperation below this threshold? Is this level socially optimal, or does partner choice lead to the investment into counterproductive forms of cooperation to out-compete others as it does in sexual selection?

In the theoretical literature on partner choice, relatively little attention has been given to these questions. First of all, a large proportion of models consider cooperation as an all-or-nothing decision and thus cannot study its quantitative level (Eshel and Cavalli-Sforza 1982; Bergstrom 2003; Aktipis 2004; Fu et al. 2008; Chen et al. 2009; Aktipis 2011; Suzuki and Kimura 2011; Sibly and Curnow 2012; Campennì and Schino 2014; Izquierdo et al. 2014; Chen et al. 2016; Wubs et al. 2016). Second, some models consider cooperation as a quantitative trait but do not entail diminishing returns, and are thus ill-suited to study the social efficiency of cooperative interactions (Sherratt and Roberts 1998; Foster and Kokko 2006; Nesse 2009; Song and Feldman 2013). Third, still other models consider cooperation as a quantitative trait with diminishing returns, but they only focus on one side of the problem –the evolution of cooperation– considering the other side –the strategy employed by individuals to choose their partner– as an exogenous parameter (Wilson and Dugatkin 1997; Ferriere et al. 2002; Barclay 2011; Wild and Cojocaru 2016).

To our knowledge, only one existing model studies the joint evolution of cooperation and partner choice in a quantitative setting with diminishing returns (McNamara et al. 2008). However, McNamara et al. (2008) make two key assumptions that turn out to have important consequences: (i) they assume that variability in the amount of cooperation is maintained owing to a very large genetic mutation rate on this trait, which prevents natural selection to act efficiently, and (ii) they restrict the set of possible strategies to choose one's partner in such a way that individuals can never do so in an optimal manner.

In this paper, we build a model inspired by McNamara et al. (2008), in which a quantitative level of cooperation expressed by individuals jointly evolves with a quantitative level of choosiness regarding others' cooperation, while relaxing these two assumptions. First, we observe that competition to be chosen as a partner leads to a joint rise of both cooperation and choosiness up to a level that depends on the efficiency of partner choice that is, in particular, on the cost of changing partner. The more efficient is partner choice, the higher cooperation is at evolutionary stability. Moreover, when the cost of changing partner is low, cooperation can rise beyond its socially optimal level. In fact, in the limit where partner choice is entirely frictionless (i.e. the cost of changing partner is zero), cooperation and choosiness rise up to the "wasteful threshold" where the cost of cooperation entirely cancels out its benefit. Individuals gain the same payoff than if they had not cooperated at all. Hence, at first sight, our analyses show that partner choice generates no systematic trend toward the socially optimal level of cooperation.

However, we then import tools from the economics literature and assume that individuals can plastically modulate their choosiness in function of their own cooperation level. This plasticity allows every individual to behave optimally on the biological market, which did not occur in the first model. In this second approach, we show that assortative matching emerges. That is, more cooperative individuals are also choosier and thus interact with more cooperative partners. As a consequence of this assortment, and provided that partner choice is efficient enough, cooperation evolves to the socially optimal level, where the mutual efficiency of cooperation is maximised.

4.3 Methods

Partner choice framework

We model partner choice in an infinite size population using Debove et al. (2015a)'s framework. Solitary individuals randomly encounter each other in pairs at a fixed rate β. In each encounter, the two players decide whether they accept one another as a partner (see below how this decision is made). If one of the two individuals (or both) refuses the interaction, the two individuals immediately split and move back to the solitary pool. If both individuals accept each other, on the other hand, the interaction takes place and lasts for an exponentially distributed duration with stopping rate τ, after which the two individuals move back to the solitary pool again.

The ratio β/τ thus characterizes the "fluidity" of the biological market. If β is high and τ is low, individuals meet each other frequently and interact for a long time. In such an almost frictionless market, partner choice is almost cost-free so they should be choosy about their partner's investment in cooperation. Conversely, if β/τ is low, individuals rarely meet potential partners and interact for a short time. In such a market, on the contrary, individuals should accept any partner.

Regarding the encounter rate, here we assume that β is a fixed constant independent of the density of available partners, an assumption called "linear search" that captures a situation in which already paired individuals do not hinder the encounters of solitary individuals (Diamond and Maskin 1979). In the Supplementary Information, however, using simulations we also analyse the model under the assumption that β increases linearly with the proportion of solitary individuals in the population, an assumption called "quadratic search" that corresponds to a situation in which already matched individuals interfere with the encounters of solitary individuals (and that is also equivalent to the classic mass-action kinetics used in mathematical epidemiology). In the paper, we only describe the results obtained under linear search. The results obtained under quadratic search are qualitatively similar (see the Supplementary Information).

Regarding the nature of the social interaction, we consider a quantitative version of the prisoner's dilemma in continuous time. Each individual i is genetically characterized by two traits: her cooperation level x_i, and her choosiness y_i. Cooperation level x_i represents the quantitative amount of effort that an individual i is willing to invest into cooperation. Choosiness y_i represents the minimal cooperation level that an individual i is willing to accept in a partner, i.e. every potential partner j with cooperation $x_j \geq y_i$ will be accepted, whereas every potential partner with $x_j < y_i$ will be rejected. Once an interaction is accepted by both players, at every instant of the interaction, each player invests her effort x_i (see below for the payoff function), and the interaction lasts in expectation for $1/\tau$ units of time, where τ is the stopping rate of the interaction.

When they are solitary, individuals gain a payoff normalized to zero per unit of time. When involved into an interaction, they gain a social payoff that depends on both partners' cooperation level. The cooperative interaction is a continuous prisoner's dilemma: making an investment brings benefits to the partner but comes at a cost to the provider. As stated in the introduction, we make the additional assumption that cooperation has diminishing returns (Altmann 1979; Weigel 1981; Killingback and Doebeli 2002). This induces the existence of an intermediate level of cooperation at which the sum of the partners' gains is maximized, the so-called "social optimum". An individual i paired with j gains the following social payoff $\Pi(x_i, x_j)$ per unit of time:

$$\Pi(x_i, x_j) = x_j - cx_i^2$$

Hence, the expected payoff of an individual i paired with j is

$$\frac{x_j - cx_i^2}{\tau}$$

where τ is the stopping rate of the interaction. The socially optimal level of cooperation is $\widehat{x} = 1/2c$. Beyond this level, the net benefit of cooperation decreases. Eventually, the interaction becomes entirely profitless, or even costly, if individuals invest more than the "wasteful threshold" $x = 1/c$. We allow both cooperation and choosiness to take any positive real value.

Previous studies demonstrated that the existence of some variability among individuals is necessary to stabilize conditional cooperation (Ferriere et al. 2002; Foster and Kokko 2006; McNamara et al. 2008; McNamara and Leimar 2010; Song and Feldman 2013). If every possible partner is equally cooperative, then there is no need to be choosy with regard to the quality of one's partner, and choosiness cannot be evolutionarily stable. In order to capture the effect of variability in the simplest possible way, we assume that individuals do not perfectly control their investment into cooperation (as in Song and Feldman 2013 and André, 2015 for instance). An individual's actual cooperation level x_i is a random variable which follows a truncated to zero normal distribution around the individual's gene value $\bar{x}_i$, with standard deviation σ. In what follows, we call cooperation level the genetically encoded cooperation level that individuals aim for, and "phenotypic cooperation" the actual level of cooperation that they express after phenotypic noise. For the sake of simplicity, here, we assume that an individual's cooperation level is randomized at every encounter. In the Supplementary Information, however, we also consider the alternative assumption where phenotypic noise occurs only once at birth (see also section *Hard-Wired choosiness*).

We are interested in the joint evolution of cooperation, and choosiness by natural selection. We undertake and compare the consequences of two distinct assumptions. In a first approach, we assume that both cooperation and choosiness are hard-wired traits, that is each individual is characterized by a single level of cooperation $\bar{x}$ and a single choosiness y, both expressed unconditionally. In a second approach, we still assume that cooperation is a hard-wired trait, but we consider that choosiness is a reaction norm by which individuals respond to their own phenotypic cooperation.

Hard-Wired choosiness

Here, we assume that each individual is genetically characterized by two traits: his level of cooperation $\bar{x}$ and his choosiness y and we are interested in the evolution of these two traits by natural selection. For this, we need to derive the fecundity of a rare mutant m playing strategy $(\bar{x}_m, y_m)$ in a resident population r playing strategy $(\bar{x}_r, y_r)$. The mutant fecundity is proportional to her cumulative lifetime payoff G_m, which can be written as (see SI for a detailed analysis of the model):

$$G_m = \frac{\bar{\Pi}_m \alpha_m \beta}{\alpha_m \beta + \tau}$$

with α_m the mean probability for an encounter between the mutant and a resident to be mutually accepted, and $\bar{\Pi}_m$ the mutant mean social payoff (see Table 1 for a list of the parameters of the model). This expression is similar to the classical sequential encounter model of optimal diet (Schoener 1971).

The evolutionary trajectory of the two traits (choosiness and cooperation) can be studied from the analysis of the selection gradient on each trait:

$$\begin{cases} \left.\frac{\partial G_m}{\partial \bar{x}_m}\right|_{\substack{\bar{x}_m=\bar{x}_r \\ y_m=y_r}} \\ \left.\frac{\partial G_m}{\partial y_m}\right|_{\substack{\bar{x}_m=\bar{x}_r \\ y_m=y_r}} \end{cases}$$

We could not derive an analytical expression of the evolutionarily stable strategy. However, we numerically computed the selection gradient on each trait, in order to study the evolutionary trajectories.

Table 4.1: Parameters of the model

Parameter	Definition
$\bar{x}_i$	Cooperation level of individual i (mean value before applying noise)
y_i	Choosiness of individual i
σ	Standard deviation of the phenotypic cooperation distribution
β	Encounter rate
τ	Split rate
$\Pi(x_i, x_j)$	Social payoff of an individual i matched with a partner j
c	Cost of cooperation
α_i	Mean probability for an individual i to interact when she encounters a resident
$\bar{\Pi}_i$	Mean social payoff for an individual i interacting with a resident
G_i	Cumulative lifetime payoff of an individual i

Plastic choosiness

Because cooperation is subject to phenotypic noise (i.e. one does not perfectly control one's own level of cooperation), it could make sense, at least in principle, for individuals to adapt plastically their degree of choosiness to the actual phenotypic cooperation that they happen to express. For instance, it could make sense for those individuals who happen to be phenotypically more generous to be also choosier, and vice versa. In our second model, we aim to explore the consequences of this possibility. To do so, we assume that choosiness is not a hard-wired trait, but a plastic decision that individuals take in function of their own phenotypic cooperation. An individual's "choosiness strategy" is thus defined as a reaction norm rather than a single value.

Our aim in this second model is to study the joint evolution of cooperation $\bar{x}$ on one hand, and of the "choosiness strategy" $y(x)$, defined as the shape of a reaction norm, on the other hand. One facet of this problem, therefore, consists in seeking for the equilibrium choosiness strategy in a situation where both one's own quality (one's phenotypic cooperation level) and the quality of one's prospective partners vary. Matching theory, a branch of micro-economics, provides tools to resolve this problem. Here we briefly explain this approach, and show how it applies to our problem.

In a first category of approaches, called matching models, changing partner is assumed to be entirely cost-free (Gale and Shapley 1962; Becker 1973). That is to say, agents have an infinite amount of time available to find each other. In this setting, theory shows that there is a unique equilibrium choosiness strategy: an individual with phenotypic cooperation x should only accept to interact with individuals with at least the same phenotypic cooperation level x, i.e. the equilibrium reaction norm is the identity function. This equilibrium strategy leads to a strictly positive assortative matching in which individuals are paired with likes.

The second category of approaches, called search and matching models, accounts for frictions in the matching process, i.e. incorporates an explicit cost for changing partner (Chade et al. 2017). These models actually correspond exactly to our own partner choice framework. Individuals randomly encounter each other at a given rate and, when an individual refuses an interaction, she has to wait for some time before encountering a new partner. Unfortunately, the equilibrium choosiness reaction norm $y^*(x)$ cannot be analytically derived in these models. However, Smith (2006) has shown that a mathematical property of the social payoff function $\Pi(x_i, x_j)$ allows predicting the shape of this reaction norm. If the social payoff function $\Pi(x_i, x_j)$ is strictly log-supermodular, then $y^*(x)$ is strictly increasing with x. If this is the case, the more an individual invests into cooperation, the choosier she should be. This equilibrium is called a weakly positive assortative matching. Log-supermodularity is defined as the following: $\Pi(x_i, x_j)$ is strictly log-supermodular only if $\Pi(x_i, x_j)\Pi(x_k, x_l) > \Pi(x_i, x_l)\Pi(x_k, x_j)$ for any investments $x_i > x_k$ and $x_j > x_l$.

Matching and search and matching models are, however, only interested in characterizing the equilibrium choosiness strategy of individuals, assuming a given, fixed, distribution of cooperation levels. As a result, matching models can offer an insight into the evolution of choosiness, but not into the joint evolution of choosiness and cooperation. To study this joint evolution in the case where choosiness is a reaction norm, and not a single value, we developed individual-based simulations.

Individual-based simulations

In addition to our analytical models, we run individual-based simulations coded into Python. We simulate the joint evolution of cooperation and choosiness in a Wright–Fisher population of N individuals, with the same lifespan L and non-overlapping generations. Mutations occur at rate μ and mutant genes are drawn from a normal distribution around the parent's gene value, with standard deviation σ_{mut}. Large effect mutations are implemented with probability μ_l. They do not alter the equilibrium result and they allow to speed up the joint evolution process. We run long enough simulations for both choosiness and cooperation to stabilize. In contrast with previous papers (Sherratt and Roberts 1998; Foster and Kokko 2006; McNamara and Leimar 2010), here we consider a continuous rather than discrete trait space, because Sherratt and Roberts (1998) have shown than too much discretization can produce undesirable consequences when studying a joint evolution process. In the Supplementary Information, we also present additional simulations based on a Moran process with overlapping generations, where the lifespan of individuals is determined by a constant mortality rate (see also section *Hard-Wired choosiness* and McNamara et al. 2008).

We run simulations both under the assumption that choosiness is hard-wired, and under the assumption that it is a reaction norm. In the second case, we test two types of reaction norms. First, we consider polynomial functions, the coefficients of which evolve by natural selection. Second, we consider step functions with evolving coefficients coding for the value of choosiness for each interval of cooperation. In the initial generation, all reaction norms are set to a constant zero function, so that individuals are never choosy at initiation.

4.4 Results

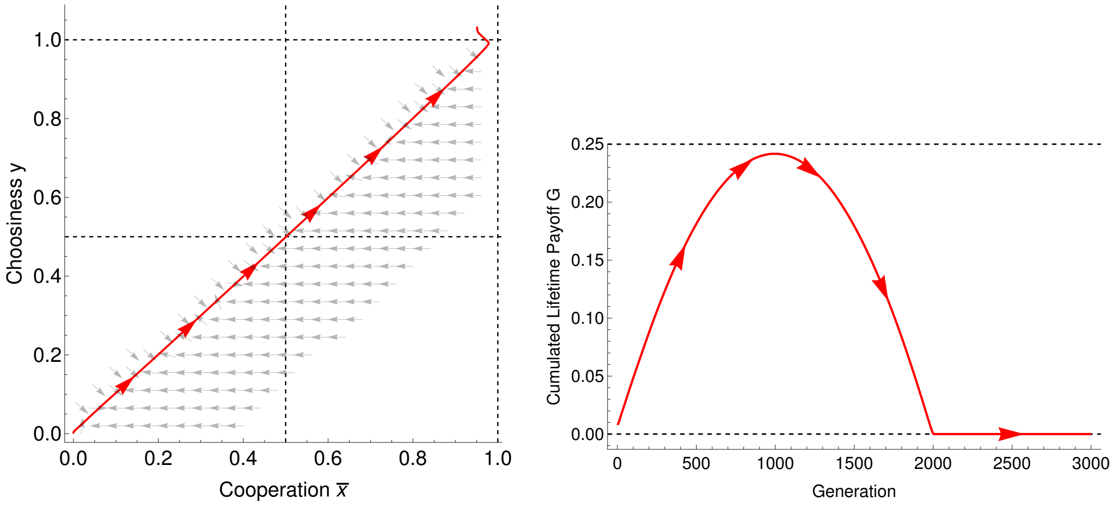
Hard-wired choosiness

Without variability in cooperation ($\sigma = 0$), there is no selective pressure to be choosier and, therefore, to be more cooperative. The only Nash equilibrium is $(\bar{x}, y) = (0, 0)$, see SI for a demonstration.

When phenotypic cooperation is variable, however, the evolutionarily stable strategy cannot be formally derived. We therefore study the joint evolutionary dynamics of cooperation and choosiness by plotting numerically the selection gradients acting on both traits. In Figure 4.1, we show the evolutionary dynamics of cooperation, choosiness, and average payoff, in a case where partner choice is very effective. When starting from an initially selfish population, cooperation and choosiness jointly rise above zero (Fig. 4.1a). At first, this leads to an increase of the net social payoff (Fig. 4.1b) because cooperation is efficient (that is, the marginal benefit of increasing cooperation for the helpee is larger than its marginal cost for the helper). At some point, however, cooperation reaches the socially optimal level where the net payoff of individuals is maximized. Beyond this level, the marginal cost of increasing cooperation is larger than the marginal benefit, but the evolutionary runaway of cooperation and choosiness does not stop. Cooperation keeps on rising toward higher values, thereby decreasing the net payoff (Fig. 4.1b). Eventually, cooperation and choosiness stabilize when cooperation is so high, and therefore so inefficient, that its cost entirely cancels out its benefit (the so-called "wasteful threshold"). That is, at ESS, individuals gain the same payoff than if they had not cooperated at all.

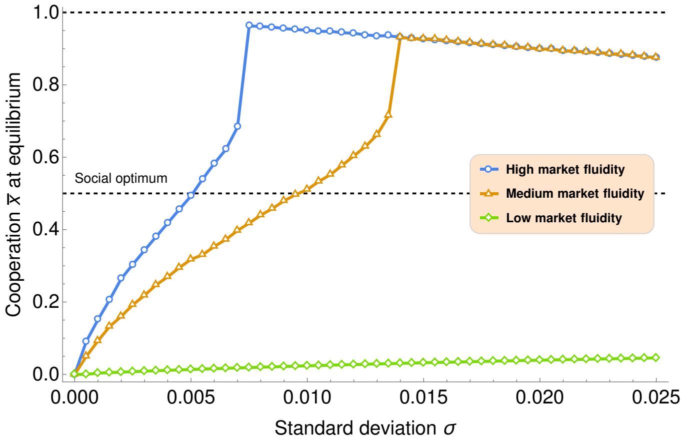
This runaway process, however, only occurs if partner choice is very efficient. If partner choice has more frictions, the rise of cooperation and choosiness halts at an intermediate level between 0 and the wasteful threshold. In Figure 4.2, we plot the level of cooperation (Fig. 4.2a), the level of choosiness (Fig. 4.2b) and the average payoff (Fig. 4.2c) reached at evolutionary stability, in function of the efficiency of partner choice (that is, in function of the parameter β controlling the fluidity of the social market and the parameter σ controlling the extent of phenotypic variability). As partner choice becomes more efficient, the evolutionarily stable cooperation and choosiness monotonously rise from zero up to the wasteful threshold (Fig. 4.2a, b). Accordingly, the net payoff obtained by individuals at evolutionary stability varies with the efficiency of partner choice in a non-monotonous way. Increasing the efficiency of partner choice has first a positive and then a negative effect on payoff (Fig. 4.2c). In the extreme, when partner choice is frictionless, cooperation and choosiness increase up to the "wasteful threshold" $x = 1/c$ at which cooperation is entirely profitless (as was shown in Fig 4.1). Note that, in this case, choosiness is even slightly larger than the "wasteful threshold" at equilibrium because, due to phenotypic variability, some individuals cooperate beyond $x = 1/c$ which makes it adaptive to request higher values of cooperation. In fact, when phenotypic variability is too high (large σ), individuals are so choosy at evolutionary equilibrium that the equilibrium level of cooperation is reduced (Fig. 4.2a). These results have been confirmed in individual-based simulations (see SI).

The runaway process can be understood intuitively. In any population, some individuals cooperate more than average, in particular owing to phenotypic variability. As a result, if partner choice is sufficiently fluid, it is adaptive to accept only these hyper-generous partners. Hence, choosiness increases by natural selection beyond the average cooperation level. In turn, this favours individuals who cooperate more than average, i.e. the mean level of cooperation increases by natural selection, etc. The extent to which this process goes on depends, however, on the efficiency of partner choice owing to the existence of a trade-off between the cost and benefit of choosiness. The runaway process stops at the point where the expected benefit of finding a better partner is not worth the risk of remaining alone.

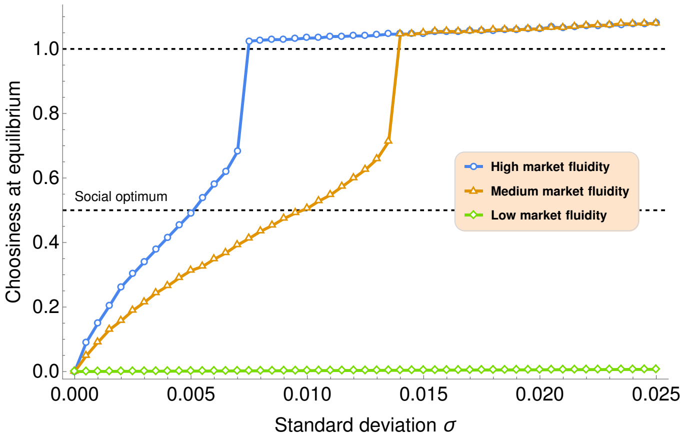


(a) (b)

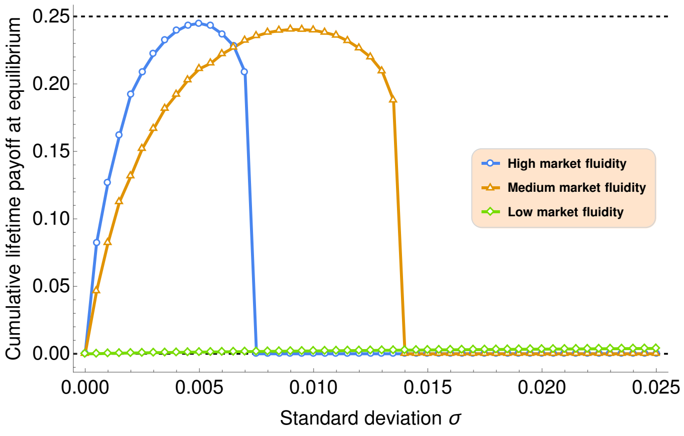
Figure 4.1: **Analytical and numerical results of the adaptive dynamics model. (a)** The grey arrows show the vector field of the selection gradient on both cooperation and choosiness. The red arrows show an evolutionary trajectory starting from an initial selfish population $(\bar{x}, y) = (0, 0)$. **(b)** The red arrow shows the corresponding evolution of the cumulative lifetime payoff G for a resident individual. Parameters are $c = 1$; $\sigma = 0.025$; $\beta = 1$; $\tau = 0.01$. The socially optimal solution is $\widehat{x} = 1/2$ and the interaction becomes profitless if both individuals invest $x = 1$.



(a)



(b)



(c)

Figure 4.2: **Analytical results of the adaptive dynamics model for a range of parameters.** Equilibrium values are shown for **(a)** cooperation, **(b)** choosiness and **(c)** cumulative lifetime payoff as a function of the encounter rate β to manipulate the market fluidity, and for three values of the standard deviation $\sigma = 0.0001; 0.01; 0.02$ respectively for low, medium and high phenotypic variability. Other parameters are the same as in Fig 4.1

In our model so far, the cost and benefit of switching partner are only determined by two parameters (the market fluidity, β/τ, and the amount of phenotypic variability, σ). Under more realistic biological assumptions, however, the cost of rejecting a partner should also depend on other parameters. For instance, one could model mortality as a stochastic process. The risk of dying while searching for a new partner would then constitute a supplementary cost of choosiness (McNamara et al. 2008). In the Supplementary Information, we develop a model based on a Moran process where individuals are subject to a constant mortality rate. As expected, ceteris paribus, the runaway process results in lower levels of cooperation and choosiness at evolutionary equilibrium when the mortality rate is high. Cooperation, however, still rises beyond the socially optimal level, even up to the wasteful threshold, if β is large and if the mortality rate is not too high.

Also, in our model, so far, we assume that an individual's phenotypic level of cooperation is randomized in every encounter. The distribution of cooperative types in the solitary population is thus a fixed and exogenous property. To test the robustness of our results, in the Supplementary Information, we analyse an alternative case where the phenotypic level of cooperation of an individual is randomized only once, at birth. In this case, the distribution of cooperative types in the solitary population is not an exogenous, fixed, property. More cooperative individuals are less likely to be solitary than average because they are rapidly accepted as partners (McNamara et al. 2008). Hence, the population of solitary individuals tends to be biased toward selfish phenotypes. As a result, the cost of being choosy is larger. Yet, in SI we show that the runaway process still occurs in this case, including up to the "wasteful threshold", as long as partner choice is efficient enough.

Note that Ferriere et al. (2002) and Wild and Cojocaru (2016, inspired by Barclay 2011) also showed that partner choice could, under some circumstances, drive the evolution of cooperation up to a "wasteful threshold". However, in both models, the choosiness strategy was fixed, and not necessarily optimal; it did not evolve jointly with cooperation. The present results are thus more robust and general.

Plastic choosiness

Here, an individual's choosiness is a reaction norm to her own phenotypic cooperation, and we used search and matching models (see Section *Plastic choosiness*) to derive the two following predictions regarding the evolutionarily stable reaction norm:

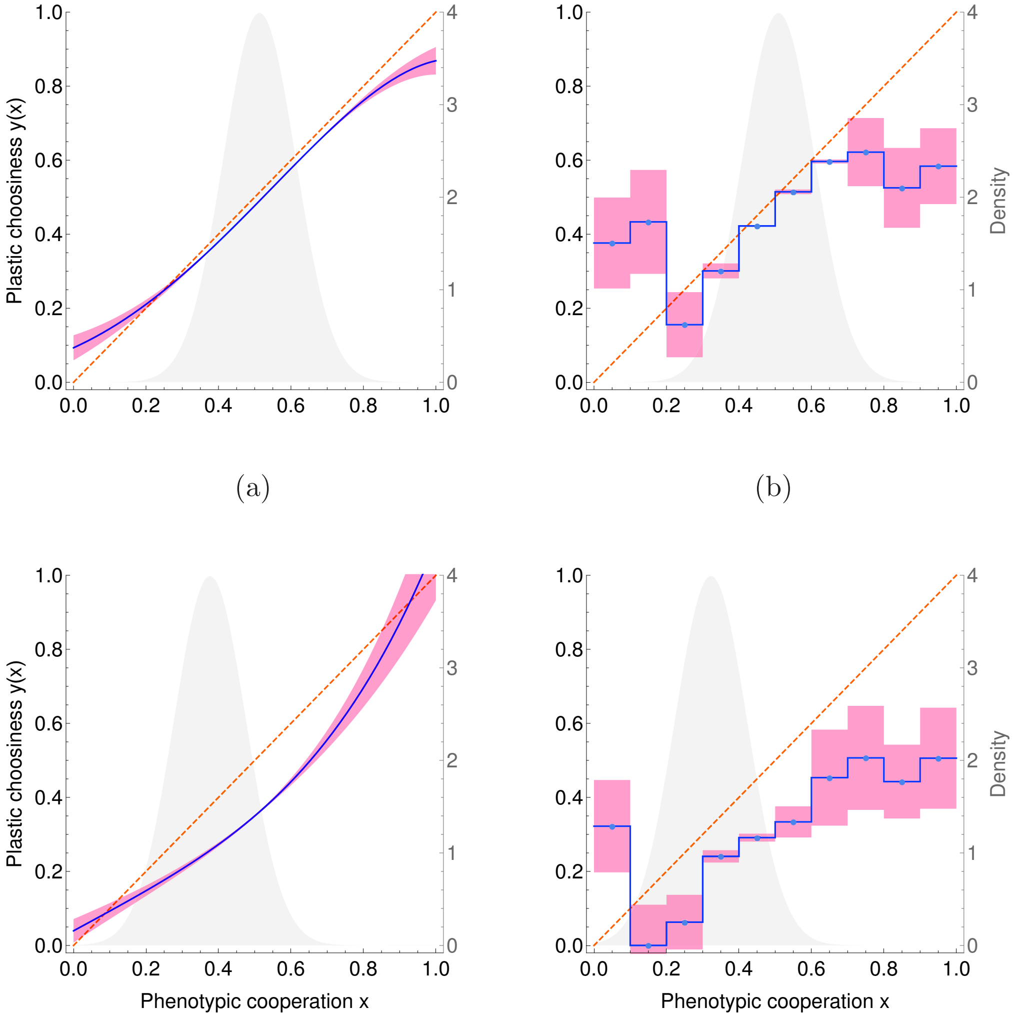
(i) If the social payoff function is strictly log-supermodular, an individual's optimal choosiness is a strictly increasing function of her own cooperation (weakly positive assortative matching).

(ii) If the market fluidity β/τ is high, the reaction norm should be close to $y^*(x) = x \quad \forall x$ (strictly positive assortative matching).

We first show that our production function Π is strictly log-supermodular. Indeed $\Pi(x_i, x_j)\Pi(x_k, x_l) > \Pi(x_i, x_l)\Pi(x_k, x_j)$ is equivalent to

$$(x_i - x_k)(x_j - x_l)(x_i + x_k) > 0$$

which is true for all $x_i > x_k \geq 0$ and $x_j > x_l$. Accordingly, search and matching models show that the optimal choosiness strategy is an increasing reaction norm, i.e. more phenotypically cooperative individuals should also be choosier, leading to a positive assortative matching at equilibrium (phenotypically generous individuals are matched with other generous individuals, and vice versa).



(a) (b)

(c) (d)

Figure 4.3: **Plastic choosiness at the equilibrium - Agent-based simulations.** The equilibrium reaction norms over 30 simulations are shown in blue, and the corresponding 99% confident intervals are shown in red with **(a-b)** high market fluidity $\beta = 1$, **(c-d)** low market fluidity $\beta = 0.01$, **(a-c)** a polynomial reaction norm, and **(b-d)** a discrete reaction norm. The orange dashed line is the optimal reaction norm for a frictionless matching market (strong form of positive assortative matching). The distribution of phenotypic cooperation at equilibrium are shown in grey. Parameters are $c = 1$; $\sigma = 0.1$; $\tau = 0.01$; $\mu = 0.001$; $\sigma_{mut} = 0.05$; $\mu_l = 0.05$; $N = 300$; $L = 500$.

Individual-based simulations confirm this result. Figure 4.3 shows the reaction norm at evolutionary equilibrium in these simulations: choosiness is strictly increasing, at least around the levels of phenotypic cooperation that are actually present at equilibrium. Outside this range, selection is very weak on the reaction norm, and we observe larger confidence intervals. As expected, when the market tends to be frictionless, the reaction norm becomes very close to the identity function, that is to a strict positive assortative matching (Fig. 4.3a and b, orange dashed line).

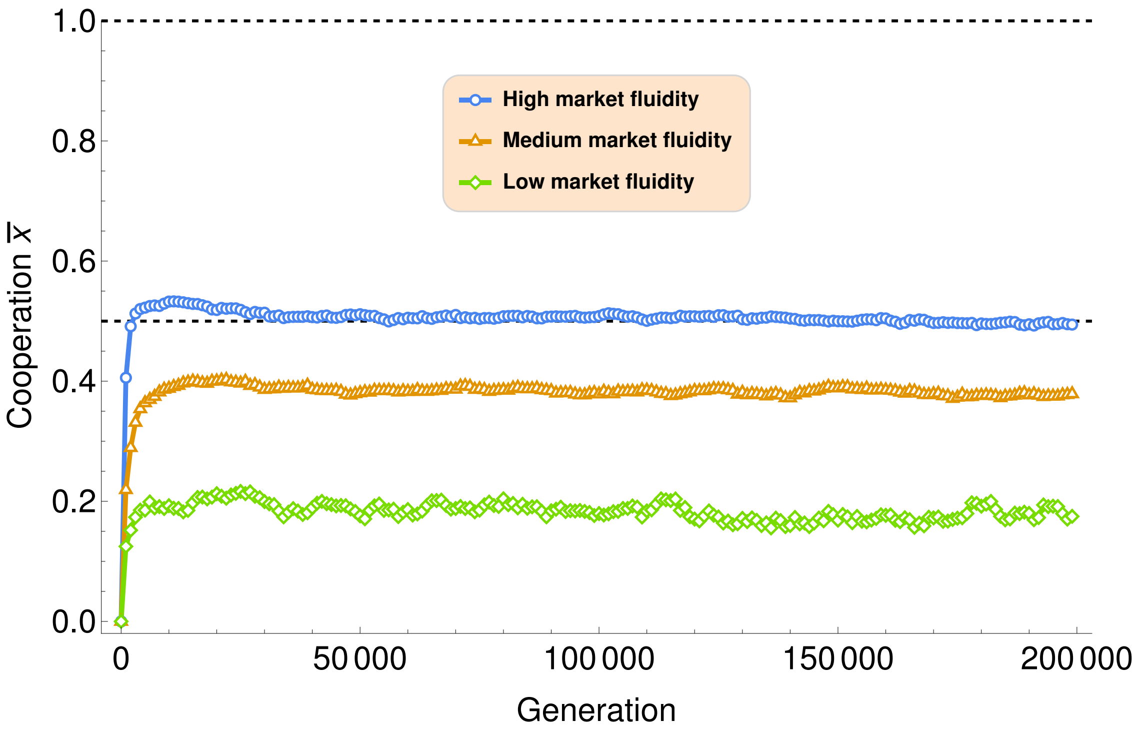


Figure 4.4: **Evolution of cooperation for a polynomial reaction norm - Agent-based simulations.** The average cooperation over 30 simulations is shown for three values for the encounter rate $\beta = 0.001; 0.01; 0.1$ respectively for low, medium and high market fluidity. Other parameters are the same as in Fig 4.3. The socially optimal solution is $\widehat{x} = 1/2$ and the interaction becomes profitless if both individuals invest $x = 1$.

Importantly, the evolution of a plastic rather than hard-wired choosiness strategy has a key consequence regarding the efficiency of cooperation at evolutionary equilibrium. In contrast with the hard-wired case, when choosiness is plastic cooperation never rises above the socially optimal level. As the efficiency of partner choice (that is, market fluidity) increases, the level of cooperation at evolutionary stability increases but, at most, it reaches the socially optimal level and never more (Fig. 4.4). In particular, when partner choice is very efficient, cooperation evolves precisely towards the socially optimal level, i.e. the level that maximizes the net total payoff of individuals ($\widehat{x} = 1/2c$).

This result can also be understood intuitively. In the first model where choosiness was hard-wired, it was adaptive to increase one's cooperation level beyond the population mean because, by doing so, an individual could switch from "being rejected by everyone", to "being accepted by everyone". The runaway process, therefore, proceeded until cooperation had no benefit at all. In contrast, in the present model where choosiness is plastic, increasing one's cooperation level is beneficial because it allows one to access better partners. Hence, this is useful only provided the benefit of accessing a higher quality partner is larger than the cost of being more cooperative. As a result, cooperation only rises up to the social optimum, where its net benefit is maximized.

4.5 Discussion

Most theoretical works on the evolution of cooperation by partner choice aim at explaining how *some* cooperation can be evolutionarily stable. They do not aim at understanding which specific quantitative *level* of cooperation should evolve. In this paper, we have raised this second question. We have considered a model where cooperation has diminishing returns, such that the most efficient level of cooperation (the level that maximises social welfare) is intermediate. We have investigated whether partner choice can account for the evolution of an efficient level of cooperation in this case. In this aim, we have modelled, both numerically and with individual-based simulations, the joint evolution of two traits: cooperation, the effort invested into helping others, and choosiness, the minimal level of cooperation that an individual is willing to accept in a partner.

In a first model, we have found that the mechanism of partner choice entails no systematic force favouring an efficient level of cooperation. On the contrary, when partner choice is effective enough, the level of cooperation increases evolutionarily toward very large values, beyond the socially optimal level. In the extreme, when partner choice is very effective, cooperation even increases up to a level where its cost entirely cancels out its benefit. That is, at evolutionary equilibrium, individuals gain the same payoff than if they had not cooperated at all.

To understand intuitively, consider a population with a given distribution of cooperation levels, with some particularly generous individuals, some particularly stingy individuals, and a given mean cooperation level. In such a population, provided that the variability of cooperation is sufficiently large and the market sufficiently fluid, it is always adaptive to accept only partners that are slightly better than average (McNamara et al. 2008). Hence, natural selection favours individuals with a choosiness always slightly larger than the average cooperation level. In turn, this choosiness selects for mutants whose cooperation level is larger than the mean, which leads to a gradual increase in cooperation. Importantly, this runaway process has no particular reason to stop when cooperation is maximally efficient. Rather, it stops when the cost of searching for more generous individuals exceeds the benefit of interacting with them (Fig. 4.2). As long as partner choice is effective (i.e. the cost of searching is low), it is always worth trying to find a better than average partner, irrespective of whether the current mean level of cooperation is below or beyond the socially optimal level. Hence, partner choice can prompt individuals to invest into counterproductive forms of cooperation to outbid others, leading to an eventually fruitless arms race.

In a second approach, in line with matching models from the economic literature, we have designed a model in which choosiness is implemented as a reaction norm to the individual's own cooperation level (see Section *Plastic choosiness*), the shape of which evolves by natural selection. In this case, both our analytical model and complementary individual-based simulations show that the evolutionarily stable reaction norm is a monotonously increasing function of cooperation (Fig. 4.3). This implies that more generous individuals are also choosier, leading to a positive assortative matching: generous individuals tend to interact with other generous individuals, and vice versa. Furthermore, if the biological market is fluid enough (i.e. if the cost of changing partner is low), this positive assortative matching becomes very close to a perfect matching in which individuals with a given level of cooperation always interact with other individuals with the exact same level (Fig. 4.3a and b).

In this case, and in sharp contrast with the model in which choosiness is a hard-wired trait, cooperation does not reach the counterproductive level where its cost cancels out its benefit when partner choice is very cheap (Fig. 4.4). More precisely, when the market is very fluid, the evolutionarily stable cooperation becomes very close to the social optimum, i.e. the amount of cooperation that maximizes the sum of the partners' payoffs. This can also be understood intuitively. Because of the strict assortment between cooperative types, individuals with a given cooperation level interact with other individuals with the exact same level. Hence, pairs of individuals become the effective units of selection, like if interactions occurred among genetic clones (Eshel and Cavalli-Sforza 1982; Wilson and Dugatkin 1997; Aktipis 2004; Akçay and Van Cleve 2012). Consequently, the socially optimal level of cooperation is favoured.

Hence, the fruitless runaway of cooperation that occurs in a model with hard-wired choosiness is a consequence of the assumption that individuals cannot optimally adapt their degree of choosiness to local circumstances. If individuals are allowed to behave optimally, which entails in the present case to adapt plastically their choosiness to their own generosity, then partner choice looks less like a competition to outbid others, and more like a competition to form efficient partnerships with others, which leads to a very different outcome regarding the net benefits of cooperation.

Previous work has shown that assortative matching favours the evolution of cooperation (Hamilton 1971; Eshel and Cavalli-Sforza 1982; Bergstrom 2003). For instance, in kin selection, assortment between relatives drives the evolution of cooperation (Hamilton 1964; Rousset 2004). To our knowledge, Wilson and Dugatkin (1997) first discussed the consequences of assortative matching for the evolution of socially efficient levels of cooperation. Alger and Weibull (2013; 2016) have studied the evolution of social preferences, rather than strategies, under assortative matching. However, both analyses did not explicitly model a partner choice strategy, let alone the evolution of this strategy, but merely assumed that assortment occurs in one way or another. In contrast, here, we have studied the joint evolution of choosiness and cooperation, showing how a positive assortative matching can emerge from a simple partner choice mechanism.

In another related work, using individual-based simulations McNamara et al. (2008) also observed a form of assortative matching in the joint evolution of cooperation and choosiness. One of the main differences with the present approach, however, is that they assumed that the variability of cooperation is maintained at the genetic level, via a high mutation rate, rather than at the phenotypic level. Under this assumption, negative selection on inefficient mutants (either too choosy or too generous) generates linkage disequilibrium between cooperation and choosiness, resulting in a positive assortative matching. For this reason, their work is more similar to our second model where choosiness is plastic than to our first model where choosiness is hard-wired. In McNamara et al. (2008)'s simulations, however, in contrast with our results, cooperation never reaches the socially optimal level (in the model where they consider a payoff function with diminishing returns). In a complementary analysis (see SI), we showed that this could be a consequence of their assumption that the genetic mutation rate is very high, which prevents natural selection from fully optimizing social strategies.

Some scholars have already imported principles from matching theory into evolutionary biology, especially in the field of sexual selection. Johnstone et al. (1996) and Bergstrom and Real (2000) have used matching models, respectively with and without search frictions, to shed light on mutual mate choice. Both works focused on the evolution of choosiness with a given, fixed distribution of individual's quality. As we have previously shown, the intensity of assortment may have a dramatic impact on the evolution of the chosen trait (cooperation, in our case). For instance, further models could investigate the precise limits of the runaway processes that occur on weaponry, or on ornamental traits, in sexual selection. More generally, matching models could be helpful to analyse a large variety of biological markets (Noë and

Hammerstein 1994, 1995; Hammerstein and Noë 2016), including inter-specific mutualisms, such as mycorrhizal symbiosis or plant-rhizobia relationships (Simms and Taylor 2002; Kiers et al. 2003, 2011).

As for the human case in particular, several lines of evidence suggest that partner choice is a likely candidate as a key driving force in the evolution of cooperation. Numerous experimental studies have shown that human beings indeed do choose their social partners in function of their cooperative reputation (Barclay and Willer 2007; Sylwester and Roberts 2010, 2013; Barclay 2013, 2016; Baumard et al. 2013; Raihani and Smith 2015; Barclay and Raihani 2016; Efferson et al. 2016; Stovel and Chiang 2016; Wu et al. 2016). Anthropological observations show that defection in traditional societies is mostly met with a passive abandon rather than with more defection in return (see Baumard et al. 2013 for a review). Also, several theoretical studies have shown that partner choice can account for the evolution of other important properties of human cooperation, such as the fact that its benefits are often shared in proportion to everyone's respective effort in producing them (Chiang 2008; André and Baumard 2011a, 2011b; Debove et al. 2015a, 2015b, 2017; Takesue 2017).

Regarding the quantitative level of cooperation, observations show that humans have precise preferences regarding the amount of effort that shall be put into helping others. Daily life contains ample examples of these preferences. For instance, we hold the door for others in subway stations, but only when they are sufficiently close to the door already, not when they are very far from it. And this is true quite generally. As experiments in real settings demonstrate, we have preferences for specific amounts of cooperation, neither too little, nor too much (Santamaria and Rosenbaum 2011; Lange and Eggert 2015). Sometimes this preference is expressed in a purely quantitative manner. At other times, the same preference is expressed in a more qualitative way, determining the kinds of cooperative action that we are willing, or unwilling, to perform. In any case, our investment in helping is quantitatively bounded. Moreover, the precise level of effort we are willing to put in cooperation seems to be constrained by considerations of social efficiency. Individuals help one another only when it is mutually advantageous, that is when the cost of helping is less than the benefit of being helped. Additionally, recent evolutionary modellings of risk pooling have revealed the socially optimal nature of helping behaviours (Cronk 2007; Aktipis et al. 2011, 2016; Campennì and Schino 2014; Hao et al. 2015). They have shown that people's systems of mutual help correspond to the most efficient systems of risk pooling in a volatile environment.

In this paper, we have shown that partner choice can foster the evolution of such an intermediate and efficient amount of cooperation, neither too little nor too much.

But we have also shown that the precise evolutionarily stable amount of cooperation should depend on the fluidity of the biological market, and can range from a very low level of cooperation, up to the socially optimal level (Fig. 4.4). A number of anthropological studies suggest that contemporary hunter-gatherer societies exhibit high levels of spatial mobility (Baumard et al. 2013; Lewis et al. 2014). Therefore, it seems plausible that biological markets were highly fluid in the social structure that our ancestors experienced. Our model predicts that, in this case, the amount of effort invested into cooperation should become very close to the social optimum. Therefore, partner choice can account for the evolution of human preferences concerning social efficiency.

One could wonder, however, whether other models than partner choice could account for the evolution of a socially optimal level of cooperation as well. The most influential model on the evolution of quantitative cooperation among non-kin is the continuous version of the iterated prisoner's dilemma (Roberts and Sherratt 1998; Wahl and Nowak 1999a, 1999b; Killingback and Doebeli 2002; Lehmann and Keller 2006; André and Day 2007; André 2015). In this game, André and Day (2007) have shown that the only evolutionarily stable level of investment is the one that maximises the total benefit of the interaction, i.e. that natural selection does eventually favour the socially optimal amount of cooperation (see also Binmore 1990; Fundenberg and Maskin 1990; Robson 1990 and Binmore and Samuelson 1992 in a discrete version of the iterated prisoner's dilemma). Yet, in this approach, selection for efficient cooperation is only a second-order force, which plays a significant role only because André and Day (2007) assumed the absence of other first-order effects. For instance, a slight cognitive cost of conditional behaviour would have prevented the evolution of efficient cooperation in their model. In another related study, Akçay and Van Cleve (2012) have shown that socially optimal cooperation is favoured when individuals play a specific class of behavioural responses to others' cooperative actions. They have also shown that, for a specific case of their model, these behavioural responses can evolve by natural selection under low levels of relatedness Here, we have shown that, under the effect of partner choice, efficient cooperation is favoured by first-order selective effects even in the total absence of genetic relatedness. This occurs because, unlike reciprocity, partner choice is a *directional* enforcement mechanism. Whereas reciprocity merely stabilizes any given level of cooperation (a principle called the folk theorem, see Aumann and Shapley 1994; Boyd 2006), partner choice directionally favours the most efficient level.

One limit of our model is that we did not introduce an explicit mechanism for reputation. We simply assumed that, in a way or another, individuals have reliable information regarding the cooperation level of others, but we did not model the way in which they obtain this information. Costly signalling theory proposes that some cooperative behaviours are costly signals of an individual's quality or willingness to cooperate (Leimar 1997; Gintis et al. 2001; André 2010; Barclay 2015; Bird and Power 2015; Bliege Bird et al. 2018). Such signals could, in theory, be far from socially efficient (Gintis et al. 2001). However, further analyses are needed to rigorously model signalling in the context of a biological market.

4.6 Supplementary Information

"Hard-wired choosiness" model

We suppose a infinite size population of individuals with lifespan L. Solitary individuals randomly encounter each other at a fixed rate β (See section *"Linear" and "quadratic" search* for an alternative assumption). If both individuals accept each other, they leave the solitary state and enter an interaction state that dissolves at rate τ. Each individual i is genetically characterized by two traits: her cooperation level x_i, and her choosiness y_i. Cooperation level x_i represents the quantitative amount of effort that an individual i is willing to invest into cooperation. Choosiness y_i represents the minimal cooperation level that an individual i is willing to accept in a partner.

As in Song and Feldman (2013) or André (2015), we choose to maintain variability in cooperation at the phenotypic level by assuming that individuals do not perfectly control the level they express. The actual investment into cooperation an individual makes follows a truncated to zero normal distribution with mean $\bar{x}$ and standard deviation σ. The corresponding density function is:

$$f_{\bar{x}}(x) = \begin{cases} \frac{e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma\left(1-\frac{1}{2}\text{erfc}\left(\frac{\bar{x}}{\sqrt{2}\sigma}\right)\right)} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

We call cooperation level the genetically encoded cooperation level $\bar{x}_i$ that individuals aim for, and "phenotypic cooperation" the actual level x_i of cooperation that they express after phenotypic noise.

There are two ways to implement this variability. Cooperation can be randomized at birth and stay constant for the whole individual's life, or at every encounter she makes. For the sake of simplicity, we use the first case in our adaptive dynamics model, but we also run agent-based simulations for both cases (see section "Complementary agent-based simulations").

Individuals encounter at random, so their cooperation level are independent. When two individuals i and j encounter, the probability that they are mutually compatible is therefore:

$$\alpha_i = \int_{y_j}^{\infty} \int_{y_i}^{\infty} f_{\bar{x}_i}(x) f_{\bar{x}_j}(y) dx dy$$

We suppose that mutants individuals are rare, and therefore only interact with residents. When in the solitary state, a mutant m gains the solitary payoff normalized to zero per unit of time. When interacting with a resident r, she gains a mean social payoff $\bar{\Pi}_m$ per unit of time. $\bar{\Pi}_m$ is the conditional expectation of the social payoff, given that the mutant and the resident are mutually compatible. Let MC be the event "the mutant and the resident are mutually compatible" and 1_{MC} the indicator random variable of the event MC.

$$\bar{\Pi}_m = \mathbf{E}[\Pi(x_m, x_r) \mid MC]$$

$$\bar{\Pi}_m = \frac{\mathbf{E}[1_{MC}\Pi(x_m, x_r)]}{P[MC]}$$

$$\bar{\Pi}_m = \frac{\int_{y_r}^{\infty}\int_{y_m}^{\infty}\Pi(x_m, y_r) f_{\bar{x}_m}(x_m) f_{\bar{x}_r}(x_r) dx_m dx_r}{\alpha_m}$$

The mutant fecundity is supposed to be totally explained by her average cumulative lifetime payoff G_m. Let $P_S(t)$ be the probability for a mutant to be in the solitary state at time t. Accordingly, $P_I(t) = 1 - P_S(t)$ is the probability for a mutant to be involved in an interaction at time t. A mutant encounters a mutually compatible partner at rate $\alpha_m\beta$, and any interaction she can be involved in dissolves at rate τ. We can then write the following dynamics:

$$\frac{dP_S(t)}{dt} = \tau P_I(t) - \alpha_m\beta P_S(t)$$

Making the additional assumption that individuals are born solitary ($P_S(0) = 1$), we can integrate the differential equation:

$$P_S(t) = 1 - \frac{\alpha_m\beta}{\alpha_m\beta + \tau}(1 - e^{-(\alpha_m\beta+\tau)t})$$

In the sake of simplicity, we assume that the lifespan of individuals L is always very large in front of the durations of both social interactions and solitary states. Therefore, the fraction of time spent by an individual in the solitary state can be written as $\frac{\tau}{\alpha_m\beta+\tau}$. Now, we can derive the average cumulative lifetime payoff G_m:

$$G_m = \bar{\Pi}_m \frac{\alpha_m\beta}{\alpha_m\beta + \tau} L$$

Without loss of generality, we make the assumption that the lifespan of individuals is equal to unity. The cumulative lifetime payoff simplifies to

$$G_m = \frac{\bar{\Pi}_m \alpha_m\beta}{\alpha_m\beta + \tau}$$

Case without variability ($\sigma = 0$)

Here, every resident individuals have the same cooperation level $\bar{x} = x$ and choosiness y. We can study the following three situations.