Mechanisms and Benefits of Reciprocal Relationships

submitted by

Charlotte S. L. Rossetti

Research group Dynamics of Social Behaviour,

Max Planck Institute for Evolutionary Biology, Plön

A Thesis submitted in partial fulfilment for the

degree of *Doctor rerum naturalium*

to

the Faculty of Mathematics and Natural Sciences

at the Christian-Albrechts University of Kiel.

January 2024

What a life, what a night
What a beautiful, beautiful ride

Scarlet Pleasure

to my friends and family, for making every experience in this life memorable.

Thesis Examination Committee

Advisor:	Dr Christian Hilbe
Reviewer:	Prof. Dr Hinrich Schulenburg
Chair:	Prof. Dr Matthias Leippe
Examiners:	Prof. Dr Arne Traulsen
	Prof. Oliver Hauser

Date of oral examination: 27th February 2024

URN: urn:nbn:de:gbv:8:3-2024-00723-3

Declaration

I hereby declare,

1. that apart from the supervisor's guidance, the content and design of the thesis is all the candidate's own work;
2. that the thesis has not been been submitted either partially or wholly as part of a doctoral degree to another examining body nor has it been published or submitted for publication than indicated in the thesis;
3. that the thesis has been prepared with regard to the Rules of Good Scientific Practice of the German Research Foundation;
4. that prior to this thesis, I have not attempted and failed to obtain a doctoral degree nor have I withdrawn an academic degree.

Charlotte Rossetti

Date and time

Abstract

Cooperation is at the heart of life on Earth. It binds together organisms, families, and societies. One of the most ubiquitous tools for studying cooperation is game theory, which has revolutionised the field of social behaviour research. This thesis explores how theoretical and experimental methods can contribute to our understanding of the different mechanisms and benefits by which cooperation evolves.

The first three chapters focus on a core mechanism of cooperation: direct reciprocity. When individuals know that there is a high probability of future interactions, they benefit from maintaining a cooperative relationship. These repeated interactions allow individuals to exchange favours and build a mutually beneficial relationship that leaves everyone better off. Chapter 2 begins with a review of the theoretical and experimental literature, which highlights some important gaps in our understanding of direct reciprocity. In particular, we argue that theoretical studies often fail to predict which reciprocal strategies humans use to maintain cooperation. One possible reason may be that most of these models and experiments study interactions in isolation, whereas most of human social life is much more complex. Thus, in Chapter 3, we investigate direct reciprocity in multi-game settings, where individuals take part in two concurrent interactions with the same or different partners. Using both evolutionary simulations and a behavioural experiment, the results show that individuals are able to link games when it makes strategic sense. We also show that cognitive biases are crucial for predicting human behaviour. Chapter 4 develops a different type of reciprocal strategy that does not rely on explicit and exact memory of the game history. Instead, it embodies more realistic cognitive abilities of human players, as revealed by a behavioural experiment.

The last two chapters examine how the importance of finding cooperative partners can explain different social behaviours. Chapter 5 reviews the literature on how different types of cues signal cooperativeness. Behavioural cues based on past behaviour have been found to be accurate predictors of cooperativeness and they are perceived as such by third parties. One such behaviour is acting according to moral values. A behaviour is considered principled if individuals display it consistently, regardless of the costs and without compromise. The Chapter 6 builds on this finding by developing a signalling model to analyse principled behaviour. The act of consistently abiding to one's principles enhances an individual's reputations for trustworthiness and makes them a preferred cooperative partner. Given the social benefits of principled behaviour, game theory and evolutionary principles can show how the dynamics of partner choice can lead individuals to display seemingly maladaptive behaviour.

Overall, the thesis offers a comprehensive examination of cooperation in reciprocal relationships, bridging theoretical insights with empirical observations to deepen our understanding of human social interactions.

Zusammenfassung

Kooperation ist das Herzstück des Lebens auf der Erde. Sie schweisst Organismen, Familien und Gesellschaften zusammen. Eines der am weitesten verbreiteten Instrumente zur Untersuchung von Kooperation ist die Spieltheorie. Sie verbindet theoretische und experimentelle Methoden. Damit erlaubt sie es ein breites Spektrum sozialer Verhaltensweisen zu erklären, darunter auch Kooperation. In dieser Arbeit wird untersucht, wie theoretische und experimentelle Methoden dazu beitragen können, die verschiedenen Mechanismen und den Nutzen der Kooperation zu erklären.

Wenn Individuen wissen, dass es eine hohe Wahrscheinlichkeit für zukünftige Interaktionen gibt, profitieren sie von der Aufrechterhaltung ihrer kooperativen Beziehungen. Diese wiederholten Interaktionen ermöglichen es den Individuen, Gefälligkeiten auszutauschen und eine für beide Seiten vorteilhafte Beziehung aufzubauen, aus der alle Beteiligten Nutzen ziehen. Kapitel 2 beginnt mit einem Überblick über die theoretische und experimentelle Literatur, der einige wichtige Lücken in unserem Verständnis der direkten Reziprozität aufzeigt. Insbesondere zeigt dieses Kapitel auf, dass theoretisch vorhergesagte Strategien die empirisch nachgewiesenen Verhaltensweisen nur ungenügend wiedergeben. Ein möglicher Grund dafür könnte sein, dass die meisten dieser Modelle und Experimente Interaktionen isoliert untersuchen, während der grösste Teil des menschlichen Soziallebens sehr viel komplexer ist. Daher wird in Kapitel 3 die direkte Reziprozität in einer Mehrspielumgebung untersucht, in der Individuen an zwei gleichzeitigen Interaktionen mit demselben oder verschiedenen Partnern teilnehmen. Es wird gezeigt, dass Individuen in der Lage sind, Spiele miteinander zu verbinden, wenn dies strategisch sinnvoll ist, und dass kognitive Verzerrungen entscheidend sind, um menschliches Verhalten genau zu modellieren. In Kapitel 4 wird eine andere Art von reziproker Strategie entwickelt, die nicht auf einer expliziten und genauen Erinnerung an den Spielverlauf beruht, sondern stattdessen realistischere kognitive Fähigkeiten menschlicher Spieler verkörpert, wie ein Verhaltensexperiment zeigt.

Die letzten beiden Kapitel gehen darüber hinaus, indem sie untersuchen, wie die Bedeutung der Suche nach kooperativen Partnern verschiedene soziale Verhaltensweisen erklären kann. In der Literatur wird untersucht, wie die Kooperationsbereitschaft durch verschiedene Arten von Hinweisen signalisiert wird. Es hat sich gezeigt, dass vergangenes Verhalten, ob direkt kooperativ oder mit Kooperation verbunden, ein verlässlicher Prädiktor für Kooperationsbereitschaft ist und von Dritten auch so wahrgenommen wird. Ein Beispiel für solches Verhalten ist das Handeln nach moralischen Werten. Das 5. Kapitel baut auf dieser Erkenntnis und einer neuen Theorie des prinzipientreuen Verhaltens auf, um ein Signaling-Modell zu entwickeln. Es zeigt, wie die konsequente Befolgung moralischer Grundsätze das Ansehen des Einzelnen in Bezug auf seine Vertrauenswürdigkeit steigert und ihn zu einem bevorzugten Kooperationspartner macht. In Anbetracht der weitreichenden sozialen Vorteile prinzipientreuen Verhaltens, können die Spieltheorie und die Prinzipien der Evolution zeigen, wie die Dynamik der Partnerwahl dazu führen kann, dass Individuen scheinbar unangepasstes Verhalten zeigen.

Insgesamt bietet die Arbeit eine umfassende Untersuchung der Kooperation in reziproken Beziehungen. Sie verknüpft theoretische Erkenntnisse mit empirischen Beobachtungen, um unser Verständnis menschlicher sozialer Interaktionen zu vertiefen.

Contents

Chapter 1

Introduction

Contents

Prologue

When friends ask me what my PhD is about, I always struggle to come up with good, concise answer. I usually say that I study cooperation, but people immediately ask me if I mean it in the sense of international relations. I say yes, and explain that I study it at a more abstract level, trying to understand the mechanisms that allow cooperation to emerge, and why it is odd that we humans cooperate so much when it seems to defy evolutionary pressures.

At this point, I usually mention my true love and the subject that brought me to a small town in northern Germany: Game theory. The core of my research is to use the ideas, methods, and principles of game theory to understand human social behaviour. I explain that this fascinating mathematical tool from economics has spread through biology, where it mixed with evolutionary theory, before infiltrating back into research on human behaviour. The moment the outcome of your action also depends on the actions of others, game theory can be used to find the best strategy. Cooperation is a strategy, and I want to understand the mechanisms that make it possible. In our lives, cooperative relationships are crucial because they are beneficial. How do we keep track of cooperation in our interactions? How do we perceive others? cooperation? How do we reciprocate cooperation across the many social relationships that make up our world? And what other behaviours do we display in the hope of attracting cooperative partners?

To lay the foundation of my thesis, let me walk you through what game theory is, why it's so powerful, how well it works with evolutionary theory, how we can use experiments, evolutionary simulations, and mathematical models to explain reciprocity and cooperation, and why all scientists who study human social behaviour should learn about it.

1.1 How cooperation shapes our lives

Humans are world-class cooperators. We help strangers in need, work together to build civilisations, pair up to produce children, and unite to defend our territory. Cooperation is everywhere, it is woven into the very fabric of our society. It is also observed in a wide range of organisms [1]: from single cells to primates, including us humans. But humans cooperate on a much larger scale and with greater variability than has ever been seen in the animal world [2]. Remarkably, the underlying mechanisms that sustain cooperation between nation states [3] and within a village of hunter-gatherers [4] are essentially the same.

Cooperation is a puzzle as old as Darwin's theory of evolution. At its core, a cooperator is an individual who pays a cost c to give a benefit b to another individual, where $c > 0$ and $b > c$. Since only behaviours that increase the fitness of organisms are favoured by natural selection, how could such costly actions have evolved? In humans, and many non-human animals, cooperation is advantageous because we are social species. Cooperation may appear to be immediately costly, but when interacting with other cooperators in a social group, it pays off. Cooperation can be seen as adaptive problem solving for the hunter-gatherer lifestyle [5]. Cooperative breeding and foraging were key features of the social lives of our ancestors and are still common in small-scale societies scattered around the globe. Hunts tend to have a low success rate, with even the most skilled hunters often returning empty-handed. However, a successful hunt can feed many, and given the perishable nature of meat, sharing becomes the logical choice. In sharing, individuals can offset the risks associated with hunting by pooling both risk and benefits collectively. Humans have distinctly long childrearing periods, with short intervals between births and extended childhoods during which juveniles depend on their mothers for survival. This model is only sustainable if mothers receive help to care for their children and create future healthy and productive adults. These examples illustrate why cooperation is crucial to the evolution of complex societies that we see today, and also how cooperation can be beneficial when the costs and benefits are shared in a group.

However, where there is cooperation, there are cheaters [6]. In a group of cooperators, it pays to be a defector and reap the benefits of cooperation without paying the associated costs. This "free-rider problem", where defectors free-ride on the benefits of others' cooperation, can lead to the collapse of cooperation, as cooperators stop cooperating to avoid being exploited. How can cooperation evolve under these conditions? It is clear that cooperation can be highly beneficial to the species and to the self, but it is not obvious how it is sustained.

The evolution of cooperation has kept scientist of many fields busy: anthropologists [5], economists [7],biologists [8], political scientists [9, 10], psychologists [11, 12], sociologists [13] and even mathematicians [14] have all contributed to different aspects of this age-old puzzle. From them, a handful of core mechanisms have been uncovered. One of the first and most primal answers is kin selection [15]. When we cooperate with family members, we help our own genes to survive. But humans routinely engage in cooperative acts with non-related kin, so what mechanisms can explain this? For cooperation to evolve, any solution must ensure compliance and deter cheating. For pairwise interactions, direct reciprocity does just that. Reciprocal relationships, where individuals repay cooperation and defections in kind, can thrive in a population with cheaters [16]. But reciprocity in large groups is a different matter. When one person in the group cooperates but another free-rides, should we withdraw our own cooperation or continue? In such cases, punishments and rewards can maintain cooperative behaviour in large groups [17]. Individuals can bear the cost of punishing themselves, or it can be enforced by an external third party [18]. Third-party punishment ensures that defectors have an incentive to comply and leads to the formation of institutions to enforce such behaviour. Another mechanism that can bypass institutions is reputation [19, 20], where reciprocators and punishers develop a positive reputation [21–23], while defectors develop negative ones. In turn, cooperative individuals will assort among themselves and only cooperate with those who have a good reputation. This form of partner choice ensures that cooperative individuals are favoured, while defectors are excluded from cooperative relationships. Whatever the mechanism, to understand cooperation, we need two things: Evolution and game theory.

1.2 Game theory and cooperation

1.2.1 Game theory

Game theory is the branch of economics that studies strategic decision-making [24, 25]. Strategic means that decisions are made in an interactive setting, where what others do matters as much as what individuals do themselves. As such, it applies to individuals, groups, firms, or nation states [1], as long as there are interactions between the decisions of the different actors. We represent these strategic interactions as games with players, actions, and payoffs. As an example, let's consider the prisoner's dilemma [26], a static game with complete information. Static means that the players move simultaneously, and complete information means that the players have full knowledge of all the parameters of the game. The prisoner's dilemma is the most widely studied game in the field of cooperation and a popular tool in this thesis [27–29]. The prisoner's dilemma pits self-interest against social interest: players can cooperate and gain a mutual benefit, or defect and gain a higher benefit for themselves at the expense of the other player. We have seen above that cooperation, at its core, is the act of paying a cost c to provide a benefit b to another individual. This is exactly what the simplest form of the prisoner's dilemma does, as shown in Figure 1.1: both players simultaneously decide whether or not to pay this cost. If they pay it, the other player gets the benefit, otherwise they do not.

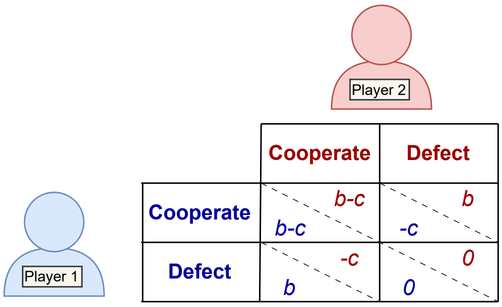


Figure 1.1: Payoff matrix for the two-players prisoner's dilemma. Each cell indicates the payoff for each player based on their respective choice. For example, if player one defects (bottom row) and player 2 cooperates (left column), player 1 gets the benefit b and player 2 the cost -c. Blue represents the actions and payoffs for player 1 and red the actions and payoffs of player 2.

This game is a prisoner's dilemma because $b > b - c > 0 > -c$ and $b - c > (b - c)/2$. The second inequality ensures that players cannot gain more than the reward of mutual cooperation by alternating defecting and cooperating with each other.

In economics, and therefor game theory, models assume that players are rational. This means that players have preferences about the possible outcomes that are complete and consistent, and that they will choose the action from a set of alternatives that leads to their most preferred

[1] People are indeed correct when linking cooperation to international relations, and we will see that the underlying dynamics are similar

outcome. In our prisoner's dilemma, players want the highest payoff. They must now choose the action that will lead to this preferred outcome. Since all players are rational and have complete information, each knows that they are themselves rational, and that the other player is rational, and they know that the other player knows that they are rational, and so on. We say that the players have *common knowledge*. With these assumptions, it becomes evident that the solution to the game, the action rational players will choose, is to defect. Let's say player 1 chooses to cooperate, then player 2 should defect since the benefit b is greater than the reward $b - c$. Similarly, if player 1 chooses to defect, then player 2 should also choose to defect, since 0 is better than $-c$. The strategy [2] defect is the only rational solution.

The outcome {defect; defect} is what we call a Nash equilibrium [30]. This concept from game theory means, by analogy with physics, that the system is in a *stable state*. The internal forces balance each other out and the system is at rest unless perturbed by a new external force. In game theory, the Nash equilibrium is the solution concept of a game. A set of strategies is a Nash equilibrium if no player can improve their payoff by changing their strategy. In our prisoner's dilemma, if player 1 changes her strategy to "cooperate', she will be worse off. The same applies to player 2. The system is at rest. This means that in situations that can be modelled with a prisoner's dilemma, individuals will not be willing to pay the cost to send a benefit, and so there will be no cooperation.

Game theory is historically and predominantly a mathematical tool. Nonetheless, because it makes predictions about behaviour, it should be tested experimentally. The prisoner's dilemma is one of many economic games or "social dilemmas" that have been adapted to be played by human participants [31]. Perhaps unsurprisingly, the results were not what game theory predicted. In the prisoner's dilemma, participants routinely deviate by choosing to cooperate: they are able to achieve the mutually beneficial outcome. These results have been replicated in a wide range of strategic games [32] and across cultures [33].

Does this mean that game theory is wrong? Rationality as described by economists maynot be representative of how humans reason and make decisions [34]. [3] First, most of our so-called preferences and beliefs are not directly chosen by us. We do not decide that we like chocolate, or value fairness, or believe in equality. It is strange to have models that assume we have done the necessary calculations to arrive at them [36]. Second, economics as a field chose many decades ago to abstract from psychological factors and focus on normative mathematical approaches [37]. These discrepancies between observed behaviour and theoretical predictions led to a revolution in the field of economics and the beginning of behavioural economics [38]. The aim of this new approach is to incorporate knowledge of psychological factors such as loss aversion

[2] An action is what the player does, while a strategy tells the player what to do given the information he has. We will see that the distinction becomes important when games are repeated.

[3] For those curious about the "anomalies" of rational choice theories, I warmly recommend reading the series of papers published in "The Journal of Economic Perspectives" by the Nobel laureate Richard Thaler [35]. In each article he focuses on a traditional effect in economics and shows how the real numbers don't match the model.

and inequality aversion to make models more predictive of actual behaviour.

1.2.2 Direct reciprocity

Beyond kin selection, one of the most fundamental mechanisms for the evolution of cooperation is reciprocity. Developed by Trivers [16] to explain what was coined "reciprocal altruism", direct reciprocity is the act of repaying another individual's cooperation. You scratch my back, I scratch yours. The term altruism was quickly lost as there is there is nothing altruistic about reciprocity if actions are conditional on returning the favour [39]. Examples of direct reciprocity in animals are not nearly as common as they are in humans (notable exceptions: the Norway rat [40], the vampire bat [41], and the stickleback [42]), but even some of these examples are contested [43]

However, most of life is not one-shot. We rarely interact with someone once, never to see them again. Repetition is a core feature of social interactions [39]. It is part of our language: the french verbs "rendre service" and "rendre visite" literally mean "to return service" and " to return visit" and are commonly used when providing help or when visiting someone. Friends who offer a drink one night expect to be offered the next one. Parents take turns watching each other's children and academics review each other's grant applications. Reciprocity can only work if interactions are repeated.

Does this mean that cooperation can be sustained in equilibrium when the prisoner's dilemma is played repeatedly? Well, it depends. According to game theory, there is no cooperative equilibrium in finite games with a known end. This is because, if the end of the game is known, by backward indiction, players will defect in the last round; knowing this, they will also defect in the penultimate round, so they will defect on the round before that as well, and so on, until the first round of the game is reached, leading to full defection. The Nash equilibrium is once more to defect on all rounds. This result may seem counterintuitive, and indeed, in experimental games, participants tend to cooperate for a few rounds, even if they eventually stop as they near the end of the game, indicating some awareness of the backward induction solution. This strategy of cooperating until the last few rounds of the game yields a higher payoff than a full defection, and participants are mostly likely aware of this fact and know that the other players know it too. [4]. However, infinitely repeated games are a different story. People do cooperate in infinite games. But this is not surprising, because in infinite games, anything goes: since there is an infinite number of rounds, there is an infinite number of strategies and thus an infinite number of equilibria. This is the Folk-theorem, so called because it was known long before it was formally proved. The Folk-theorem applies as long as the likelihood of another encounter is sufficiently high [45]. Theoreticians model this as a continuation probability δ which represents the shadow of the future. The higher δ, the higher the probability that the two individuals will meet again [46, 47]. So what kind of cooperative strategies are feasible in infinitely repeated games?

[4]The literature on the centipede game provides an excellent discussion on the paradox of backward induction [44].

Typical strategies that support cooperation are conditional [6], meaning that actions are based on the outcome of previous rounds. For example, the strategy "cooperate in all rounds" is not conditional. A strategy that says "If the other player cooperates, then cooperate. If they defect, then defect." is conditional. How many past rounds are considered varies, but the vast majority of the theoretical literature focuses on strategies that only remember the previous round, also called memory-1 strategies [48, 49]. A variation of the above strategy which is memory-2 would be be "only defect after the other player defected twice in a row".

The strategy described above is one of the most famous and successful strategies in the repeated prisoner's dilemma: Tit-for-Tat (TfT). In Axelrod's seminal tournament paper, when pitted against other strategies, TfT ended up with the highest payoff [50]. TfT is a highly reciprocal strategy that very closely resembles the adage "an eye for an eye". Axelrod's analysis of this strategy and the other successful ones revealed some key characteristics of what makes a winning strategy. 1) Be nice, by starting with a cooperation. 2) Beg forgiving, by restoring cooperation after a defection. 3) Be retaliatory, by punishing defection. 4) Be clear, by having a simple set of comprehensive rules. These characteristics demonstrate that it pays to be cooperative [51], but that it's important not to not be taken advantage of. This tournament paved the way to many decades of research on direct reciprocity, especially through evolutionary models. [5]

1.3 Evolutionary biology and (social) behaviour

1.3.1 From economics to biology

In the 1970s, biologists realised that game theory was the perfect tool to study the evolution of behaviour in animals [52, 53]. There are many situations in animal lives where the best course of action depends on what other individuals do. Animal conflict and territoriality [54, 55], sex allocation [56], bird mating [57], kin selection and inclusive fitness [15]. Organisms are the new players, and fitness is the new payoff. Several types of games are used to reflect the different problems that organisms face. We can see which strategies can survive and reproduce as natural selection drives organisms towards optimisation.

Importantly, biologists have realised two things: 1) Seemingly "altruistic" behaviours benefit the individual just as much as the species. We have seen that when it comes to cooperation, if the recipient of the help is kin or likely to reciprocate in the future, then it is advantageous for the individual to cooperate, even if the whole group benefits. This means that group selection is not necessary to explain these behaviours – individual self-interest can lead to widespread cooperative strategies. 2) Decisions need not be conscious and rationality is not required, [52, 58]. Players do not need to understand the game structure or have the ability to engage in complex reasoning (e.g. by using backward induction). Nash equilibria can be reached as long as players gather information about the relative advantages of different strategies simply by trial and error! The mechanism of natural selection naturally rewards the best strategies that have

[5] See Chapter 2 for a review of theoretical and experimental results of reciprocity in the repeated prisoner's dilemma

the ability to survive and reproduce. In this way, game theory can be applied to biology and animal behaviour without making any assumptions about the level of cognitive complexity of the organisms. It can even be used to model plant growth [59]. Strategies can simply be inherited rather than chosen.

With these two insights, evolutionary game theory was born [60, 61]. By applying game theory to evolving populations, a whole range of biological phenomena, not just behavioural ones, can be explained. We know from classical game theory that the success of a strategy in a game depends on the strategy of the other players. Just in the same way, the fitness of a phenotype depends on the other phenotypes present in the population [62]. Game theory is particularly well suited to modelling biological systems because such systems are often frequency dependent, meaning that the frequency of each phenotype matters as much its mere presence in the population. As we have seen in the tournaments in subsection 1.2.2, the environment selects the strategies, but the strategies are what make up the environment, meaning that they shape the environment themselves.

Different mechanisms exist to model evolutionary dynamics in a game, relying on processes of natural selection, mutation, or drift [62]. A population of strategies is simulated over generations that replicate through a particular mechanism: individuals interact in a game, the winner reproduces, the loser does not. Evolutionary games also have their equilibrium state. The classical Nash equilibrium is a state in which no player can benefit from deviating from their strategy, provided that others stick to theirs. This concept can be extended to a dynamic population of competitors. An *evolutionary stable strategy* (ESS) is a strategy that can survive all possible "mutant" strategies, meaning that no other mutant strategy can successfully invade the current state [54]. We find again the concept of stable state.

Returning to the case of cooperation, evolutionary game theory can help us identify the strategies that evolve, just like as in a tournament. Strategies can evolve through competition in a population, as we described above. Players interact repeatedly in a game and can compare their strategies against a mutant strategy. The strategy yielding the highest payoff is implemented and the process is repeated over many cycles. This is the method we use in Chapter 3 to investigate which strategy evolves when players interact in two games concurrently.

Nevertheless models of direct reciprocity make some simplifications that can be problematic: first, most of the literature on repeated games has focused on two-person interactions, but research on humans requires considering larger groups of interacting individuals [7]. There is some experimental and theoretical evidence that direct reciprocity fails in large groups [28, 63]. As group size increases, it becomes more difficult to target cooperation at those individuals who cooperate and to avoid free riders. Indeed, the only cooperative ESS is to respond with cooperation when everyone else cooperates – the basin of attraction is extremely small because a few defectors will undermine it. Second, these models assume forced interactions over fixed repetitions: players have no other choice but to interact with each other. In human social life, the option to stop interacting is often present, meaning that individuals can simply choose not to interact at all, or end a relationship mid-game. Third, all common reciprocal strategies such as TfT are what is referred as "bookkeeping" strategies. However, in reality, favours may not be so quantifiable. Individuals often differ in their possession of resources and personal traits, making it sometimes difficult for players to agree on what constitutes fair reciprocation [64, 65]. Also, in species with socially complex environments where individuals navigate multiple partners, reciprocity can often cross domains [66, 67], leading to the same difficulties. There is little evidence that animals have the cognitive capacity to preform such bookkeeping. Even in humans, many cooperative relationships are in the form of friendships or long-term partnerships in which people do not keep an exact count of cooperative actions [68].

We have seen that the cooperative strategies favoured by evolutionary models all too often fail to replicate in experiments. One of the aims of this thesis is to investigate cooperation and reciprocal mechanisms in domains closer to the richness of humans social life. Chapter 3 adresses the first problem of multiple interactions, while Chapter 4 explores a new strategy that is less cognitively demanding for players. Finally, Chapter 5 and Chapter 6 go beyond direct reciprocity to understand which behaviours are chosen for cooperative relationships.

1.3.2 Beyond cooperation: From biology back to human behaviour

We have seen that people rarely play the rational solution to the game in economic experiments. They are much more cooperative than predicted. A new generation of economists went back to incorporating human psychology into their models, taking into account cognitive mechanisms [69, 70], social preferences [71, 72], and beliefs [73, 74]. Through biology, however, game theory took a different route to predict behaviour, particularly social behaviour such as cooperation. As we have seen in our examples of non-human animals, individual self-interest can lead to group-beneficial strategies, even in the absence of conscious decision making. This gives us an entirely new lens though which to study human social behaviour. Game theory allows us to get to the ultimate function of a behaviour. Biologists are very familiar with the concept of different levels of analysis to explain behaviour [75]. Relevant to us, they distinguish between proximate explanations – the How?, and ultimate explanations – the Why?. *How* questions refer to the mechanism of the behaviour, our thoughts, feelings, beliefs, and preferences. *Why* questions refer to the underlying adaptive function of the behaviour, reproductive success, survival, fitness maximisation. The whole concept of ultimate answers comes from evolution.

When we relax the assumption of rationality and apply game theory with an evolutionary lens, we find that it has real predictive power. Evolution has shaped our behaviour to be optimal and is able to lead equilibrium [76]. Just as with non-human animals, we can explain complex behaviours without any conscious decision making on the part of the players. In this sense, it is worth thinking of human behaviour as "adaption-executioner" rather than "fitness-maximiser" [77]. People don't maximise themselves, they simply execute the strategy that has adapted to their lives. Individuals do not consciously compute the payoffs of their actions. Evolution leads them to behave as if they did, simply because behaviours that are not adapted to the environment have been selected against. Evolution is the fitness maximiser, not us [6]. We have evolved a psychology that makes us prefer fairness, and that does not magically disappear in the lab. Is it really so strange that participants still cooperate in a one-shot prisoner's dilemma when it is the mutually beneficial outcome? Is it really so strange that, in an ultimatum game, people choose to split the money equally rather than take the money themselves? Is it really so strange that people behave this way even when there is no chance of a future encounter and no opportunity for reputation? Game theory is a tool, and the parameters we input determine the output it gives. To quote computer scientists: it's garbage in, garbage out. If we start with assumptions that have no basis in reality, the result will not be predictive of reality.

When investigating the ultimate functions of social behaviour, it often comes down to cooperation and communication which are inherently game-theoretic [62]. Cooperative relationships are highly beneficial and therefore valuable. Being a good reciprocator, whatever shape reciprocity take, makes it more likely to interact with other cooperative individuals. There is a large body of literature on reputation-based partner choice [79–81], where people with a reputation for being cooperative or altruistic are favoured as social partners [81, 82], and receive more help even from people that they have not previously helped [83, 84]. By the same logic, can we show that other types of behaviour that confer a good reputation are sought by potential cooperative partners? My final two chapters address this very question from two different angles: Chapter 5 reviews which cues individuals use to infer the cooperativeness of others, and whther such cues are reliable. Chapter 6 uses game theory to find the ultimate function of a type of puzzling human behaviour through concepts of cooperation and communication.

1.4 Methods and objectives

This thesis consists of five chapters, each of which represents a project I worked on during my PhD. They range from the very basic of the mechanisms of reciprocity in social cooperative relationships, to the more complex use of evolution and game theory to explain how other types of social behaviour emerge to help individuals form these relationships. In Chapter 2, we review the experimental and theoretical literature on reciprocity in the repeated prisoner's dilemma. In the third and fourth chapters, we consider two projects on direct reciprocity which blend the two approaches; in Chapter 3, we investigate conditional cooperation across multiple concurrent interactions, while in Chapter 4 we look at a different kind of reciprocal strategy: cumulative reciprocity. The last two chapters look beyond the prisoner's dilemma and direct reciprocity to focus on how different social behaviours can be used as cues for cooperative and trustworthy types. Chapter 5 reviews the literature on social dilemmas to highlight the different cues people might use to infer the cooperativeness or trustworthiness of others, and whether such inferences are reliable. Finally, Chapter 6 shows that certain behaviours, such as principled behaviour, can

[6] It is worth noting that in the case of humans, natural selection can be replaced by mechanisms of cultural learning [36, 78]

be explained as signalling devices that have evolved to make a principled individual appear more trustworthy, and therefore preferred as a cooperative partner.

In my research, I use a diverse skillset. I run evolutionary simulations, economic experiments, and analyse game theory models. This mixed method approach helps me achieve more complete answers in my research. In the following, I describe each chapter in more detail.

1.4.1 Direct reciprocity in the prisoner's dilemma

Chapter 2 is a review, written with Christian Hilbe, of the literature on direct reciprocity in the repeated prisoner's dilemma. We chose to focus on this two-player game because it is by far the most widely studied game. As highlighted in subsection 1.2.1, studying idealised scenarios such as the prisoner's dilemma can help to capture the core features and dynamics of a behaviour. In this review we were particularly interested in contrasting results from theoretical analysis and experimental methods, as both make extensive use of the game.

We consider the environments in which reciprocal strategies can evolve and in which participants play cooperative strategies. The parameters of the game, such as the cost/benefit ratio, the continuation probability and the presence of errors, have a strong influence on cooperation. The effects on players predicted by theoretical methods are largely supported by empirical results. However, when it comes to predicting the exact strategy people use in this game, the results are more mixed.

1.4.2 Cooperation in concurrent interactions

Chapter 3 is based on the observation that most research on reciprocity examines a single type of interaction in isolation. However, human social life is far richer and more complex, with individuals engaging in multiple interactions simultaneously with many different partners. Studies on cooperation in networks do not allow participants to target their reciprocity to one neighbour, forcing them to interact in the same way with all game partners. A few papers investigate how reciprocity might spill over from one interaction to another when played concurrently, or how players learn to reciprocate across two different relationships. We combine these studies to investigate reciprocity in multiple interactions both theoretically and experimentally.

We consider two scenarios beyond the standard two-players prisoner's dilemma. In the first, players interact in two games simultaneously with the same partner, while in the second, they interact with a different partner in each game. To compare different levels of cooperation, and to see if our concurrent interaction design of two simultaneous repeated games is able to push cooperation up or down from a baseline, each game has a different benefit to cooperation (while the costs remain the same). We first run evolutionary simulations in which the players have the opportunity to update their strategy in each time step by comparing their current strategy to a new, randomly generated, mutant strategy. We complement these simulations by investigating how cognitive biases of memory errors and mistakes can influence behavioural spillovers across two games. In parallel, we run an online behavioural experiment based on the same design and analyse the conditional strategies used by the participants.

We find that playing two games lowers cooperation from the first round, even though people use different reciprocal rules in the two treatments. Cognitive biases play a crucial role in predicting actual player behaviour. This highlights the need to account for errors in evolutionary simulations.

1.4.3 Cumulative reciprocity: new strategy of direct reciprocity

In Chapter 2, we highlighted the different strategies that evolutionary models find capable of supporting cooperation. We pointed out two limitations: that the strategies favoured by theoretical models do not seem to be the ones people use when playing the games themselves, and that most of the literature focuses on a handful of strategies and their memory depth extensions. In Chapter 4, we collaborate with a team led by Prof. Haoxiang Xia to provide empirical evidence for a new strategy for the prisoner's dilemma: Cumulative Reciprocity (CURE).

CURE differs from the common strategies in the literature because it does not follow the standard structure of memory-1 strategies. Instead of remembering the exact actions of the players in previous rounds, the players only keep a tally of the imbalance in the interaction. If this counter is zero or low enough, they continue to cooperate. Once it reaches a certain threshold, they start defecting. This rule captures the dynamics of reciprocity in a novel way, as players will only start cooperating again once the counter goes back below the threshold, when the other player has started cooperating again. It also allows reciprocity to unfold over a longer period of time and without the need to closely the exchange. These features make CURE a very realistic strategy for people to use in everyday life. In close relationships, humans rarely keep an exact count of who owes whom a favour. Rather, they have a general sense of whether the relationship is equal and fair, or whether they are being cheated.

Theoretically, CURE is very strong. The strategy supports cooperation even in the presence of errors, enforces fair outcomes, and is able to evolve in hostile environments. Most importantly, these results are supported by empirical evidence from an economic experiment. To implement CURE experimentally, we created two versions of a repeated prisoner's dilemma: a standard one and one with errors. The treatment with errors is important because CURE is particularly robust to the presence of errors. The results show that CURE is better able to predict the behaviour of most participants in every round compared to other standard strategies. Cumulative reciprocity may be a more valid way to encapsulate reciprocity when taking into account the type of players humans are (error-prone) and the shape cooperation takes in our world (norms of fairness).

1.4.4 Cues that (reliably) signal cooperativeness

In the previous three chapters we have seen how reciprocity can sustain cooperation in a wide variety of situations. Chapter 5 goes beyond the mechanisms of reciprocity by reviewing how players perceive the cooperativeness of others. Reciprocating cooperation, unsurprisingly, makes individuals appear cooperative themselves. However, there are many other cues that can lead to different degrees of perceived cooperativeness. Other types of altruistic actions, such as donating to charity or punishing defectors, are generally perceived as indicators of good character. Sometimes, even seemingly more arbitrary cues are also used to infer cooperativeness, such as wealth, gender, or religious affiliation. To what extent do people rely on such cues and to what extent are they reliable indicators of how cooperative someone is?

In this chapter, we review the literature on perceptions of cooperativeness through different types of economic games. We look at how either a third party participant or the interacting partner rates the level of cooperativeness or trustworthiness of players displaying a range of cues. More importantly, we pay particular attention to studies that also test the reliability of such cues. Incentivised (or sometimes hypothetical) economic games are perfect for this, as they allow for the actual behaviour of players to be tested in parallel with how the other participant perceives that behaviour. Our review shows that behavioural cues based on past actions are strong predictors of cooperativeness and are correctly perceived as such. On the other hand, personal cues are much less reliable and can lead people to make incorrect assumptions, perhaps linked to preexisiting stereotypes. Nevertheless, it seems that people are aware of the unreliability of personal cues and only rely on them when no other cues are available. People are a good judges of others and are willing to infer cooperativeness from many different sources. As cooperative relationships are highly beneficial, the ability to find good cooperative partners is crucial.

1.4.5 A signalling model of principled behaviour

Chapter 5 shows how cooperativeness is often inferred from our actions, even when they are not directly related to reciprocating. In particular, cues that may not be directly associated with cooperative behaviour can still be perceived as such by others, such decisions in a moral dilemma, religious affiliation or pro-environmental preferences. If individuals have learned to use this information to infer cooperativeness, could it be that the performers have learned to display this behaviour to signal cooperativeness? The literature on reputation-based partner choice discussed in subsection 1.3.2 shows how individuals can acquire a reputation for being cooperative through their actions. One such action is the display of principled behaviour.

Principled behaviour is the act of abiding to certain moral and social principles. Individuals who behave in a principled manner display a number of characteristic behaviours, such as consistency at all costs, prioritisation beyond instrumental value, and rejection of any compromise. In Chapter 6, we show that the ultimate function of principled behaviour is to build a reputation for trustworthiness. This, in turn, makes them more likely to be chosen as cooperative partners. One of the key characteristics of principled behaviour is consistency: always sticking to the principle regardless of the situation or the cost. By acting consistently in cooperative contexts, individuals increase their trustworthiness with each iteration of the principle. Using a signalling model, we show that highly selective receivers lead senders to exhibit key features of consistency:

complementarity (the more I cooperated in the past, the more likely I am to cooperate again) and discontinuity (if I defect once, all my reputational capital is lost and further cooperation is futile). We compare this equilibrium with a competing equilibrium: moral licensing. Individuals who license are more likely to act morally ambiguously after an initial principled action.

Chapter 2

Direct Reciprocity in the Prisoner's Dilemma

This chapter is published under the title: Rossetti, C. S. L., Hilbe, C. (2023). Direct reciprocity among humans, *Ethology*, 130(4), eth.13407. https://doi.org/10.1111/eth.13407 [85]

Contents

Abstract

Direct reciprocity is the tendency to repay others' cooperation. This tendency can be crucial to maintain cooperation in evolving populations. Once direct reciprocity evolves, individuals have a long-run interest to cooperate, even if it is costly in the short run. The major theoretical framework to describe reciprocal behavior is the repeated prisoner's dilemma. Over the past decades, this game has been the major workhorse to predict when reciprocal cooperation ought to evolve, and which strategies individuals are supposed to adopt. Herein, we compare these predictions with the empirical evidence from experiments with human subjects. From a theory-driven perspective, humans represent an ideal test case, because they give researchers the most flexibility to tailor the experimental design to the assumptions of a model. Overall, we find that theoretical models describe well in which situations people cooperate. However, in the important case of "indefinitely repeated games," they have difficulties to predict which strategies people use.

2.1 Introduction

Cooperation is a fundamental component of many social interactions [1, 86]. It occurs when individuals share food and other commodities [40], contribute to a collective action [87], or when they use public resources responsibly [10, 88]. The common pattern behind these examples is that individuals incur a personal cost to benefit others. Such seemingly altruistic behaviors warrant an explanation: if cooperation is to evolve, it needs to give a fitness advantage to the cooperating individual or its kin [12]. Importantly, however, such a fitness advantage does not need to arise immediately. Instead it suffices if there is some advantage eventually, over the course of an individual's lifetime. This insight provides the basis for direct reciprocity [16], one of the key mechanisms for cooperation [14, 60]. When individuals interact in stable groups, their cooperative acts today may lead other group members to cooperate with them in future. Once future benefits are sufficiently valuable, (conditional) cooperation is what evolution selects for.

In nature, cooperation can come in various degrees, and it can involve many individuals. Yet when modeling behavior, it is often useful to consider idealized scenarios that capture a behavior's central features in the simplest possible way. One frequently used paradigm to study cooperation is the prisoner's dilemma [26]. It describes an interaction among two individuals (in theoretical studies, the prisoner's dilemma is often described as a "game," individuals are referred to as "players," and outcomes are called "payoffs"; however, the framework covers scenarios that are less innocent than these names might suggest). The rules of the interaction are as follows. Each individual can choose to cooperate or to defect. Mutual cooperation yields the highest payoff to the pair, yet defection yields a higher payoff to each individual. Because choices are made independently, the only reasonable and consistent outcome of the prisoner's dilemma – the only Nash equilibrium – is mutual defection. The prisoner's dilemma is widely used because it captures the essence of cooperation: the conflict between self-interest and group interest. At the same time, it is arguably the most simple model to do so: there are only two players (instead of many), and players can only choose among two discrete actions (there are no different shades of good and bad behavior).

While defection is the only equilibrium in a single interaction, predictions change when players interact over multiple rounds. In that case, players can adopt reciprocal strategies to enforce cooperation. They can cooperate with other cooperators, and they can stop cooperating against defectors. Importantly, this form of reciprocity can evolve even when individuals do not consciously compute the payoffs of their actions. In the end, evolution leads them to behave as if they did, simply because it wipes out behaviors that are not well-adapted to an individual's environment. After decades of research, there is by now a vast theoretical literature on evolution in the repeated prisoner's dilemma [89]. This literature explores how evolving cooperation rates depend on the parameters of the game and on the exact setup of the evolutionary process [90].

Importantly, these models in evolutionary game theory typically take an ultimate, not a proximate, perspective. They ask in which kinds of environments cooperative behavior would be adaptive. To this end, the models neglect any specific emotions that individuals might feel when making their decisions, or any values that individuals might hold. Rather the models ask which kinds of strategies allow for stable cooperation in a given environment, irrespective of the proximate mechanisms that might lead individuals to implement those strategies. Real behavior might not perfectly resemble the strategies predicted by this theory. Yet, we would hope that evolutionary theory can give us some clues on which behavioral patterns are essential for reciprocity to succeed. At the same time, it should be noted that evolutionary models do not have the aim of exactly predicting cooperation levels. Rather they allow us to explore which qualitative features of an interaction are favorable to cooperation and which are not.

In this article, we compare the theoretical predictions for the prisoner's dilemma with empirical evidence from behavioral experiments with humans. Humans represent an ideal test case for evolutionary models for various reasons. First, humans develop the capacity for reciprocity already at an early age [91], and a majority of adults engage in behaviors consistent with conditional cooperation [6, 92]. Second, online and laboratory experiments with human subjects are straightforward to implement and comparably cheap. Third, the experimental design and the instructions can be easily tailored to explore the impact of different payoff parameters, stopping conditions, and learning horizons. For some results, it is also useful that humans are capable to respond to hypothetical scenarios. For example, by letting participants interact with computerized opponents, one can explore how they would react to certain predefined strategies that are relevant for the theoretical literature [93–95]. Of course, the resulting insights on human subjects cannot be easily extrapolated to other species. can they be easily extrapolated to human interactions in real life, where reciprocity is more difficult to quantify (for a recent exception, see [3]. However, due to the flexibility of experiments with human subjects, they can serve as a first test case to determine which models of reciprocity might be sensible in principle. We use these insights to reflect on the success of evolutionary models, and to identify open problems that require more work.

The remainder of this article is organized as follows. In the next section, we briefly review the theoretical literature on the repeated prisoner's dilemma. We then comment on typical experimental implementations and describe their impact on observed average cooperation rates. Afterwards, we review common conditional strategies observed in human reciprocal interactions, and we discuss cognitive constraints and their impact on reciprocity. Finally, we provide a brief overview of reciprocal interactions captured by models different from the standard prisoner's dilemma.

2.2 Theoretical background

2.2.1 The repeated prisoner's dilemma

The prisoner's dilemma is a game among two players who independently decide whether to cooperate (C) or to defect (D), as illustrated in Fig.2.1a. Mutual cooperation yields a *reward* of R to both players, whereas mutual defection results in the *punishment* payoff P. If one player defects whereas the other cooperates, the defector obtains the *temptation* payoff T whereas the cooperator ends up with the *sucker's payoff* S. For the game to be a prisoner's dilemma, the payoffs have to satisfy the inequalities T > R > P > S. When these inequalities hold, game theory tells us that the rational choice for both players is to defect although mutual defection yields a lower payoff than mutual cooperation. This is because defection is the only "safe" choice where both players cannot do anything else that will make them better off. In addition to the above inequalities, most models also assume that 2R > T + S. This latter assumption ensures that it is the symmetric outcome of mutual cooperation that yields the highest total payoff, rather than the asymmetric outcome in which one player cooperates and the other defects.

There are two particular instantiations of the prisoner's dilemma that are often used as baseline examples. One is based on the payoffs $R = 3, S = 0, T = 5$ and $P = 1$ (Fig.2.1b). From a theoretical viewpoint, there is nothing special about these particular parameter values, other than that they were used in the seminal study of [50]. From an experimental viewpoint, however, it must be noted that all these payoffs are non-negative. While mathematical predictions typically only depend on the relative magnitudes of payoffs, not on their absolute values or signs, humans are known to be sensitive to negative framing. The other instantiation is the so-called donation game with payoffs $R = b - c, S = -c, T = b$, and $P = 0$, where b > c > 0 denote the benefit and the cost of cooperation, respectively (Fig.2.1c). While these two instantiations satisfy all of the above inequalities, they do not generate the entire space of all prisoner's dilemmas (which instead would require using the general payoffs R, S, T, P). However, in many cases the specific payoffs of Axelrod and of the donation game are easier to work with, which explains their wide use in many evolutionary models [90].

To explain direct reciprocity, we are interested to see what happens when the game is repeated. In a repeated prisoner's dilemma, we now have a social interaction with multiple encounters, such that players interact for several rounds (such iterated interactions are sometimes referred to as "supergames"). From a theoretical perspective, it is useful to distinguish two different kinds of repeated interactions. They are referred to as the finitely and the indefinitely repeated game, respectively. In the finitely repeated game, the two players interact for a commonly known number n of rounds (Fig.2.1d). Perhaps somewhat surprisingly, the standard prediction

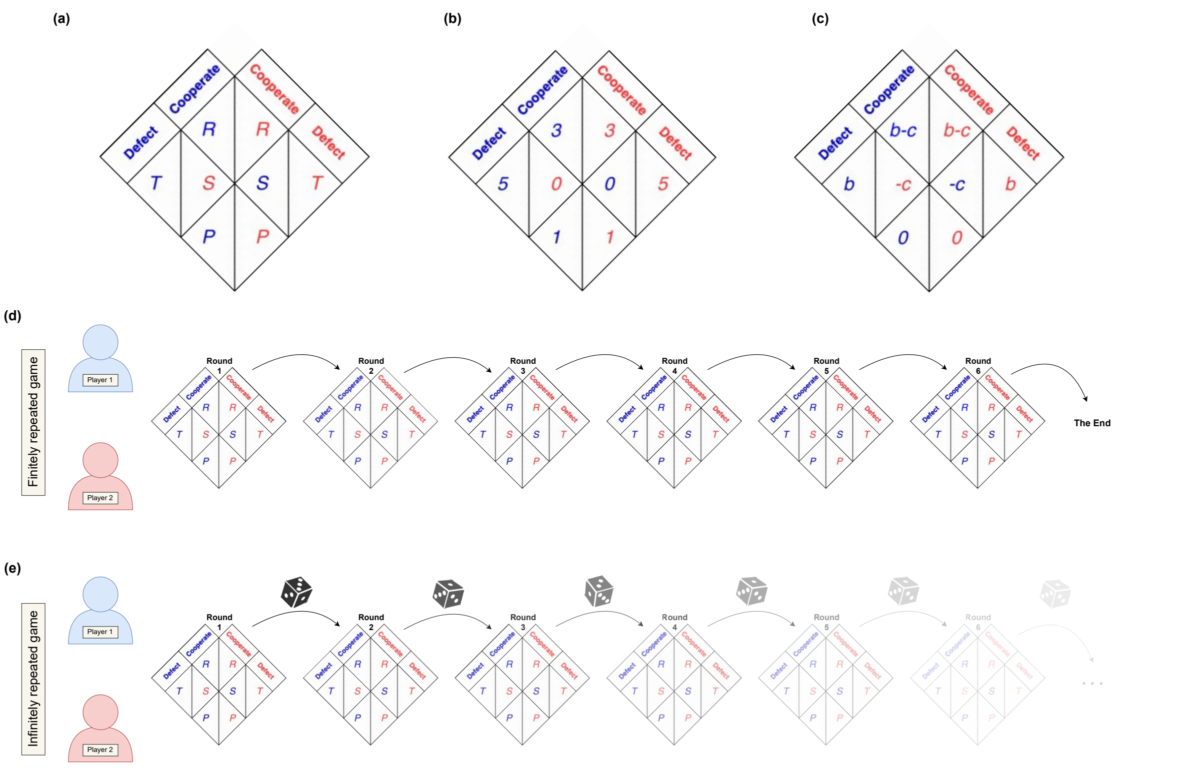


Figure 2.1: Dynamics of cooperation in concurrent games. Basic setup of the repeated prisoner's dilemma. (a) In the prisoner's dilemma, two individuals (here depicted as blue and red) independently decide whether to cooperate or defect. Mutual cooperation gives a reward R to both, whereas mutual defection yields the lower punishment payoff P to both. If one player cooperates and the other defects, the defector gets the highest payoff T (temptation), whereas the cooperator gets the smallest payoff S (the sucker's payoff). (b) Axelrod and Hamilton (1981) studied a particular variant of this game with payoffs $T = 5, R = 3, P = 1$, and $S = 0$, which has become a baseline since. (c) Another popular representation of the prisoner's dilemma is the donation game, in which payoffs are framed in terms of benefit b and cost c of cooperation. In the theoretical literature, it is common to distinguish two variants of repeated games: (d) In the finitely repeated prisoner's dilemma, the two players interact for a known number n of rounds. In particular, in the last round, players are aware that no further interactions will occur. (e) In the indefinitely repeated prisoner's dilemma, there is a constant chance that a further round occurs. In particular, players can never be sure that they will not interact again.

for finitely repeated game is the same as for the (one-shot) prisoner's dilemma (the one with n = 1). This result follows from *backward induction*: in the very last round n, players no longer have any incentive to cooperate, and hence they should both defect. However, given both players defect in round n anyway, it becomes optimal to already defect in round n -1, and by the same logic, in all previous rounds. This race-to-the-bottom logic no longer applies in the indefinitely repeated game (Fig.2.1e). Here, there is no commonly known last round. Instead, after any interaction, there is always a probability $\delta > 0$ of a further encounter. According to an equivalent interpretation, one may also imagine two players who interact for infinitely many rounds, but who discount future payoffs with a discount factor of δ. For this reason, indefinitely repeated games are sometimes also referred to as "infinitely repeated games" [27], even if $\delta < 1$. Once there is no predetermined last round, reciprocal cooperation becomes feasible. Hence it is the indefinitely repeated game that is considered in most (but not all) theoretical studies on the evolution of reciprocity.

When we study behavior in games, we look at what strategies players use. In a one-shot game,

Table 2.1: Some archetypical strategies of the repeated prisoner's dilemma.

Abbreviation	Name	Desciption	Type	Memory length	Reference
AllD	Always Defect	Play defect in all rounds	Unconditional	0	
AllC	Always Cooperate	Play cooperate in all rounds	Unconditional	0	
TfT	Tit-for-Tat	Cooperate on the first round, then copy the previous action of the other player	Conditional	1	Rapoport & Chammah, 1965; Axelrod and Hamilton , 1981
Grim	Grim trigger	Cooperate on the first round and continue to cooperate as long as the other player cooperates. If the other player defects once, then defect forever.	Conditional	1	
WSLS	Win-Stay/ Lose-Shift	Cooperate on the first round. Then if both players chose the same action in the previous round, choose cooperate. If both player chose different actions in the previous round, choose defect.	Conditional	1	Kraines & Kraines, 1993, Nowak & Sigmund, 1993
STfT	Suspicious-Tit-for-Tat	Defect on the first round, then copy the previous action of the other player	Conditional	1	Boyd & Lorberbaum, 1987
Tf2T	Tit-for-Two-Tats	Cooperate on the first round. Then play TfT but only defect after two defections.	Conditional	2	
CURE	Cumulative Reciprocity	Cooperate on the first round. Then on each round, add one point to your tally if the other player defected when you cooperated. Remove one point if you defected when the other player cooperated. Neither remove nor add points if both players did the same. When the tally exceeds some predetermined threshold, say 3, defect.	Conditional	∞	Li et al., 2022

Note: We describe some strategies that have been highlighted by the previous theoretical literature, either because they have important properties, or because of they represent interesting extreme cases.

there are as many pure strategies as there are actions: players can either cooperate or defect. In contrast, when the game is repeated, the number of strategies can be vast (it becomes infinite when the game is itself infinite). This is because strategies for the repeated game correspond to contingent plans of action. They tell the player what to do in any round, depending on what happened in all previous rounds. For example, always defect (ALLD) is a strategy. Choose at random is also one. Cooperate all the time but defect every fourth round as well. Only some strategies are interesting, either because they are played by human subjects or because of their theoretical properties. In particular, researchers tend to look at conditional strategies. Unlike the examples given just above, this set of strategies take into account the co-player's previous behavior. For example, a player may cooperate as long as the other co-player does, then defect every time. This strategy is called GRIM [60]. Another example is the strategy Tit-for-Tat (TFT), where players simply copy what the other player did on the previous round.

Because the space of possible strategies of the prisoner's dilemma is enormous, it is common in the evolutionary literature to artificially restrict the space of strategies that players can use. For example, some studies assume that players only react to the outcome of the last round, or more generally the last k rounds [51, 96–99]. Some other studies assume that individual strategies need to be representable by a so-called finite-state automaton [100]. The states of such an automaton can be thought of as the players' different mental states (such as being "satisfied" or being "angry"). The players' states in the current round determine which actions they choose, which in turn determine the players' state in the next round. A few instances of such strategies, including the strategies ALLD, GRIM, and TFT, are described in Table2.1. Restricting the players' feasible strategies (to either have finite memory or finitely many states) serves two purposes. On the one hand, it captures that humans rarely act as perfectly calculating machines that condition their behavior on the entire previous history of interactions. On the other hand, these restrictions allow researchers to more efficiently explore which strategies might evolve. For example, only when players are restricted to choose from a reasonably small set, one can hope to explore the dynamics with computer simulations.

A final modeling assumption that is often made is that people may commit errors. For example, they may commit implementation errors: in situations in which players would usually cooperate, they might instead defect with some probability ε, possibly because of a "trembling hand" [101]. Alternatively, it is sometimes assumed that individuals misremember past events, possibly due to a "fuzzy mind" [102]. Again, the assumption of errors serves two purposes. On the one hand, it makes models more realistic. After all, chance events do sometimes interfere with people's decisions: sometimes we misinterpret an action, or we simply forget what we meant to do, or we meant to do that in another interaction. As a response, people seem to have developed ways to cope with these different kinds of noise [103, 104]. On the other hand, models with errors are sometimes easier to tract mathematically and statistically [60]. For example, without noise it can be difficult to infer a player's strategy solely based on the player's previous sequence of actions. This problem arises, for example, when two players both cooperate in all rounds. Such a sequence of actions is consistent with the assumption that both players are unconditional cooperators (ALLC). However, it is equally consistent with the assumption that both players cooperate conditionally (e.g., TFT or GRIM). Games with noise make it easier to distinguish these two cases: once one player defects (possibly by mistake), the other player can show her true colors.

2.2.2 Theoretical predictions

After defining the rules and parameters of the game, we briefly discuss what kind of predictions have been derived from this model. These predictions can be based on several different ways how to analyze the repeated prisoner's dilemma, see Box 1. In the following, we summarize the general patterns that follow from this analysis, both for the finitely and for the indefinitely repeated game. In each case, we ask: how does a given parameter or assumption affect cooperation? In addition, we ask: which strategies are players predicted to adopt?

For the finitely repeated game, we have noted earlier that backward induction predicts that players fully defect eventually. This equilibrium prediction holds for all parameters (i.e., it is independent of the exact payoffs, or the exact number of rounds). There are, however, alternative models that predict some cooperation to emerge. These alternative models are based on the assumption that there is always a positive chance that a given co-player is conditionally cooperative — either because the co-player has social preferences [105], or because such strategies are occasionally introduced by mutations [106]. Once there is even a small chance that the opponent might cooperate, conditional cooperation can become self-enforcing: Rather than trying to preempt the co-player's defection, it becomes rational to adopt a conditionally cooperative strategy, and to only defect once the co-player did so. This mechanism can lead to a substantial increase in predicted cooperation rates, especially if cooperation yields high benefits and if players interact for many rounds.

For the indefinitely repeated game, evolutionary and equilibrium arguments suggest that game parameters should affect cooperation in intuitive ways, see also Table2.2. For example,

Table 2.2: Basic predictions for the repeated prisoner's dilemma.

Parameters	Effect	Desciption	Theoretical findings	Experimental evidence
Benefit	As the benefit of cooperation increases, cooperation increases.	For a given continuation probability, the game's payoffs determine whether or not cooperation is feasible. Intuitively, both theory and experiments find that the higher the benefit in comparison to the cost, the more cooperation.	van Veelen et al., 2012; Akin, 2016; Stewart and Plotkin, 2015	Dal Bó and Fréchette, 2019; Gill and Rosokha, 2020;
Continuation probability	As the probability of another round increases, cooperation increases.	Reciprocity is strongly influenced by the length of the interaction, or the likelihood that the current interaction will continue. High continuation probabliity lead to higher cooperation rates, especially for payoffs in equilibrium.	Hilbe et al, 2015; Schmid et al, 2022	Dal Bó and Fréchette, 2018
Error rate	As noise increases, cooperation first increases, then it decreases.	Several studies already find a decrease in cooperation at small levels of noise. However, when players are allowed a longer memory (take more than just the last round into account), they learn to be more forgiving to compensate this issue, maintaining cooperation.	Zhang 2018	Aoyagi et al., 2003; Fudenberg et al., 2012; Li et al., 2022

Note: In evolutionary models of the repeated prisoner's dilemma, the main parameters are the payoffs, the continuation probability, and the error rate. Here we summarize their predicted effects. The two last columns provide some references for the suggested relationships.

the larger the benefit-to-cost ratio b/c, the more profitable cooperation becomes, and hence individuals should be more likely to cooperate [107, 108]. A similar argument holds for the continuation probability δ. The more likely it is that people interact for many rounds, the more important it becomes to maintain cooperative relationships, and hence cooperation should increase (e.g. [109]. The effect of errors is predicted to be ambivalent. Small error rates ε can sometimes enhance cooperation [110], but frequent errors rates tend to be detrimental [111]. Moreover, cooperation can be further promoted if interactions are assorted rather than well-mixed (that is, when players are more likely to encounter co-players with the same strategy, [100]).

On the level of predicted strategies, there is a curious mismatch in predictions. Studies based on round-robin tournaments (when each contestant, here strategy, competes one-to-one with all others) often predict reciprocal strategies like TFT to be most successful (e.g., [50]). TFT cooperates if and only if the co-player did so in the previous round. This strict form of reciprocation can be advantageous in heterogeneous populations; by using TFT, a player can enforce that outcomes are fair, no matter what strategy the opponent adopts [112]. On the other hand, TFT is very sensitive to errors. When two TFT players interact, already one (mistaken) defection is sufficient for mutual cooperation to break down. For this reason, studies based on evolutionary simulations often find that TFT only plays a transient role, and that players eventually learn to adopt a strategy of win-stay lose-shift (WSLS, [99]). WSLS prescribes to repeat the previous action if the player's payoff was at least R, and to switch to the opposite action otherwise. Compared to strict reciprocation, WSLS has the strong advantage that it is robust with respect to errors. Indeed, even when one player defects by mistake, two WSLS players recover mutual cooperation after two rounds. Due to this property the strategy of WSLS is a Nash equilibrium in games with errors, whereas TFT is not [97]. As a result, most evolutionary simulations predict that individuals should use WSLS, not TFT, to enforce cooperation.

Box 1: Theoretical methods to explore optimal play in the repeated prisoner's dilemma. Most previous research uses one of three different methods to explore optimal behavior in the repeated prisoner's dilemma: equilibrium analysis, computer tournaments, or evolutionary simulations.

Equilibrium analysis is a direct application of game theory and uses analytical methods to characterize which Nash equilibria are possible [113]. These equilibria are important because they give us some indication about which outcomes may occur in principle (strategies that are not equilibria are unlikely to persist). In the case of indefinitely repeated games, however, the equilibrium approach is surprisingly inconclusive. The celebrated "folk theorem" guarantees that almost any outcome might arise as an equilibrium if only the continuation probability is sufficiently large. The only requirement is that each player at least receives the mutual defection payoff P (see, e.g., [45]. In some cases, however, the required continuation probability might be prohibitively large in practice.

Round-robin tournaments represent another way to gain insights into the repeated prisoner's dilemma. Here, the assumption is that we can pit all strategies against each other and see which ones finish with the highest payoffs. This approach has been pioneered by [50], who found Tit-for-Tat (TFT) to succeed. Their study has since been repeated (and challenged) by several other groups [114, 115]. In particular, whether or not TFT succeeds depends on the strategies that are allowed to take part in the tournament, and on the game's parameters — such as the error rate [49].

Finally, through *evolutionary simulations*, researchers can test which strategies emerge in evolving populations. Here, researchers assume that individuals repeatedly play against other population members, and successful players are more likely to reproduce. By exploring which strategies evolve eventually, researchers aim to identify behaviors that optimally support cooperation. Such evolutionary simulations often predict that WSLS or related strategies succeed [51, 96, 99, 116]. The evolutionary approach is naturally connected to the other two. For example, in large populations with strong selection and rare mutations, the strategies that emerge correspond to the Nash equilibria of the game [117]. On the other hand, when mutations are frequent, such that all strategies are played in almost equal frequencies, evolution favors the strategy that would also succeed in the round-robin tournament [118].

2.3 Impact of design choices and parameters on human cooperation

After discussing the central predictions of the theoretical literature, we compare them to the experimental evidence. Herein, we focus on data from controlled experiments with human subjects. These subjects have either been invited to interact in games in a laboratory, or they have been recruited through online platforms like Amazon Turk or Prolific [119]. In each case, participants are asked to repeatedly make decisions in a repeated prisoner's dilemma and they are paid in proportion to their performance in the game (for similar evidence on the repeated public goods game, see for example [72]. Moreover, in some of the studies, individuals do not only engage in one repeated prisoner's dilemma (one supergame). Rather they consecutively act in several supergames with changing partners. In this way, the corresponding studies can disentangle two concurrent effects that both lead to behavioral change: strategic conditional play (within each supergame) and learning (across supergames).

2.3.1 Dynamics in the finitely repeated prisoner's dilemma

For our summary of the experimental literature, we start with the finitely repeated prisoner's dilemma. Here, participants know the length of the game beforehand. Experimental outcomes seem to critically depend on how many rounds participants play. When there are only a few rounds, like two rounds or four, people seem to learn the logic of backward induction. As a result, they eventually start to defect early on in the game [46]. This picture changes, however, once there are more rounds. In that case, subjects seem to robustly cooperate until 3 – 4 rounds prior to the known ending of the game; only then there is a notable drop in cooperation [120, 121]. Cooperation does not seem to further unravel even if subjects have many opportunities to learn the specifics of the game. In an online game, participants interacted in 10–round prisoner's dilemmas with changing co-players over 20 days. Even by the end of the experiment, cooperation rates in each game remained high until round eight [122]. Such results impose strong limits on backward induction. However, they agree with models suggesting that people cooperate conditionally because they assume others might do so as well [105, 106]. In line with this view, cooperation is even more pronounced if participants are told that some of their interactions may take place against computerized opponents who implement TFT [123].

2.3.2 Dynamics in the indefinitely repeated prisoner's dilemma

Next, we consider experiments on the indefinitely repeated prisoner's dilemma. After every round, the game will stop there with some (known) probability $1 - \delta$, or continue for at least one more round with probability δ. If the game continues, the same termination rule applies to the new round. A probability of $\delta = 0$ means there is no other round, whereas a probability of 1 means there will be another round for sure. A probability of 0.5 means that there is a 1/2 chance of another round, and hence the expected number of rounds is $1/\ (1 - \delta) = 2$. All existing theoretical models suggest that cooperation ought to become more likely as δ becomes larger. Confirming this basic expectation, [27] find in an analysis of 15 new studies that cooperation increases with the probability of another round.

In addition to the continuation probability, cooperation is predicted to depend on the exact payoff parameters of the prisoner's dilemma. Most often payoffs are chosen to test a mathematical model [124]. A specific combination of continuation rule and payoff matrices allow for different possible equilibria. Economists are interested to see how the availability of different equilibria affects behavior [27]. But even if the set of possible equilibria is unchanged, different payoff matrices can lead to different choices from players. For example, research suggests that cooperation is more abundant when it is risk-dominant; in this case, risk-dominance means that players prefer to cooperate when they think it is equally likely that the co-player adopts ALLD or

GRIM [125, 126]. In particular, a higher reward R leads to more cooperation [127], and a higher temptation T leads to less cooperation. [128] noted that when the reward is low, players are more likely to open with a defection on the first round. This indicates that when the gains from bilateral cooperation are not very high, players are more suspicious. They start by defecting even though they may still try to establish cooperation afterwards. In addition to these expected effects of payoffs, we note that there are many factors that can influence cooperation in games played with human subjects that are difficult to account for with standard evolutionary models. We already mentioned the effect of negative payoffs, as individuals are notoriously loss averse [129]. But also the value of the payoffs, which can be manipulated through the conversion rate into real money payments, can have an effect.

The last component that has a significant impact on behavior is the addition of noise through errors. Experimentally, errors can be implemented by having a choice be executed as its opposite. While any participant can find out that their own actions have been misimplemented, they have no way of finding out whether the co-player's choice was intentional. Instead, participants are only informed about the general rate with which errors occur. The presence of such noise changes how people play and which strategies they use. Already with a small level of noise, cooperation decreases [130–132]. This might be because strategies in treatments with noise tend to look further backward in time. That is, players condition not only on the previous round but also on older rounds. For example, players in [131] stated that they tended to give their co-player a benefit of the doubt. They would attribute the first defection to an error, and they would only start defecting themselves after the co-player defected multiple times. The authors describe this "leniency" and "forgiveness" as key components of strategies in noisy treatments.

2.4 Evolving strategies

After having looked at the emerging cooperation rates, in the next step we wish to describe which strategies participants use. This endeavor, however, is non-trivial. After all, strategies are contingent plan — they tell a player what to do after any possible history of previous play. In contrast, in experiments participants often make decisions for one particular history, which makes it difficult to estimate how they would react to alternative scenarios.

In the literature, there have been different ways to deal with this problem. First, instead of asking participants to choose an action each round, we can ask them to choose their repeated-game strategies. Participants are informed that these elicited strategies are then used to determine how they act in the subsequent experiment. This is the so-called strategy method [133]. Here, participants either chose from a menu of predefined strategies, or they define their memory-1 conditional strategies (i.e., for any outcome of the previous round, participants define with which probability they wish to cooperate in the next round, see [127, 128]. This method has the advantage that the results are clear; we can see which existing strategies are preferred. However, we lose a lot of nuance as individuals are usually more messy and hardly stick to one such strategy completely. By letting subjects choose from a finite strategy set, we also risk missing a strategy that would be popular had it existed in the menu of possible strategies. The other, more common approach is to infer strategy from actual choices [27, 121, 122, 126, 131, 132, 134–138]. This method is more cumbersome and different techniques exist. One is to use Bayesian inference to ask which strategy (out of a given set) is most likely to reproduce a participant's observed behavior. Because the set of possible strategies is determined by the researcher, this approach is subject to similar criticisms as the strategy method. The other option is to estimate conditional responses based on previous play (most often assuming that individuals react to the last round only). This approach, however, requires that participants in fact experience all possible game outcomes for which a response is to be estimated.

2.4.1 Strategies in the finitely repeated prisoner's dilemma

After highlighting the difficulties that arise when estimating the participants' strategies, in the following we discuss which conclusions have been drawn with the above methods. In the finitely repeated prisoner's dilemma, conclusions are surprisingly clear. Here, the data suggests that a large fraction of participants can be accurately described by a particular class of conditional strategies. For a game of length n, these strategies define a threshold of rounds $k \leq n$ up to which they cooperate — unless the co-player defects before, in which case they defect for the remainder of the game [121, 122]. These estimated strategies are in good agreement with previous models of cooperation in finitely repeated games [105, 106].

2.4.2 Indefinitely repeated prisoner's dilemma

In games in which there is always a probability of another encounter, results are more mixed. It seems that the dominating strategies are TFT (including some variants thereof), ALLD and GRIM [27, 126, 128, 131, 136, 138, 139]. Papers that allow for longer memory either find that it is not necessary [128], or that players simply prefer a more lenient version of TFT. This can be modulated by the payoffs chosen as demonstrated by [127]: the higher the reward, the more lenient the strategies. One drawback of these results, especially when strategies are estimated from behavior, is that when players cooperate from beginning to the end of an interaction, there is no way to distinguish among several possible strategies. Adding noise can force a more diverse history of play, which makes it easier to tell strategies apart. When that is the case, longer-memory strategies and more lenient strategies become more popular [131, 132]. Surprisingly, however, these experiments give little support to WSLS, which usually emerges in evolutionary simulations [96, 99]. These observations suggest that when evolutionary simulations predict cooperation to evolve based on WSLS, they might not reflect the true dynamic that underlies human cooperation. One aspect of WSLS that might be particularly counter-intuitive to human subjects is how it continues after a deviation from mutual cooperation (possibly because of an error). In that case, the strategy comes back to co- operation only after the interaction fell into mutual defection, not before. As such it is a little bit more forward looking than GRIM or TFT, which either never return to cooperation (GRIM), or only after the other player cooperated (TFT). This subtlety of WSLS might make it hard for humans to understand the true intentions of someone using this rule of behavior, even though it makes WSLS more robust to errors.

2.5 Memory constraints

As discussed above, memory plays a key role in the strategies people play in a prisoner's dilemma. When strategies are constructed, we have a choice over how much memory we allow. Real humans are not so straightforward and simple. Already when estimating strategies, some papers limit memory by only considering memory-1 strategies. To some extent, there is a good argument to be made to limit memory, as many people would fail to remember exactly what happened in all previous rounds as length increases. In addition, in a real-life setting, interactions can span weeks, months, years, and people interact with many other interaction partners during that time-frame. All of this places some constraints on what can be realistically remembered of the details of the interaction.

A few papers have tested memory for cooperative actions explicitly. [102] and [140] have the participants take part in a memory task where pictures of hypothetical partners as well as their action in a hypothetical game is displayed on screen. Treatments vary the number of total partners in a memory set or the number of "in-between" partners between two viewings of the same partner [102]. These studies suggest that overall, memory is extremely poor. Moreover, error rates further increase drastically as the number of "in-between" partners increases. As a consequence, when the researchers perform evolutionary simulations based on these error rates, the dominating strategies tend to be ALLD and GRIM. Interestingly, the total number of different partners does not influence memory. The authors conclude that traditional conditional strategies such as TFT are not realistic because in a setting with multiple partners, memory is not sufficiently accurate.

However, these studies look at memory without having subjects actually interact in a game. The results are quite different when participants must recognize and type hypothetical partners that they actually played a prisoner's dilemma with (computer partners with photographs). [141] find that memory is highly accurate for recognition and categorization as cooperator or defector both immediately and 1 week after playing. The amount of cooperation with each type matches participants' memory performance: subjects cooperate less with partners that defected before. They also highlight that memory is best for rare types in the population, rather than best for defectors as previous literature has suggested. Similarly, [142] find that memory for both defectors and cooperators is accurate when playing a repeated prisoner's dilemma against 16 different computer partners they encountered six times. These studies show that even if exact actions may not be remembered perfectly, human subjects have an accurate feeling of the kind of partner they are facing.

These papers tested memory for partners and their actions. Another aspect of memory is simply how a high load might affect cooperative behavior and strategies. [135] investigated if strategy complexity is affected by memory load, which they find to be the case. When subjects play a memory game in parallel to the repeated prisoner's dilemma, they move from playing WSLS to TFT. However, the methods of this paper might not pass the test of time. Interactions were not anonymous and the entire lab saw the decisions of the participants. A more recent experiment by [143] using the same distracting memory task finds that low load subjects are better able to condition their strategy on previous outcomes. Players in both the low and high load condition conditioned their strategies on previous actions, but only low load players seem to consider older actions.

These empirical results demonstrate that individuals remember the information they need in order to reciprocate cooperation. When the interaction is real, they are attentive to player types even when encountering dozens of multiples partners in one session and can remember these players accurately for days. At the same time, when the demands on memory are high, player tend to use simpler strategies but still maintain a similar level of cooperation. Sophisticated strategies that require long memory do not seem crucial to the emergence of reciprocity. Instead, simple rules of behavior relying on remembering types of players is often sufficient.

2.6 Beyond the standard prisoner's dilemma

In the previous sections, we restricted our attention to a particular class of experiments on reciprocity. In all cases, participants interacted with a fixed co-player in a prisoner's dilemma over a series of multiple (discrete) rounds. In the following, we briefly mention two natural extensions that highlight the particular flexibility that researchers have when conducting experiments with humans. One extension deals with cooperation in networked populations; the other extension explores how people cooperate when they make decisions in real time.

Most human interactions happen within a social network where individuals have relationships with many others. Abundant theoretical work suggests that such non-trivial interaction structures can have an impact on cooperation through the mechanism of network reciprocity [144, 145]. This form of reciprocity argues that different network shapes and connectivity patterns allow players to cluster into cooperative groups. This natural occurring assortment makes cooperators less susceptible to exploitation. Several papers have tested this theory with human players in large to very large networks [146–148]. Assuming that individuals have to choose the same action (cooperate or defect) against all their neighbors (as in the models), these studies find little evidence of clustering. Moreover, they find a similar decay in cooperation independent of the size and exact shape of the network, unless the benefit of cooperation is sufficiently large [149]. As for strategies, a re-analysis of the three main papers found that players seem to ignore the payoffs of their neighbors when making decisions. Instead they simply chose their action based on how many cooperators are among their neighbors, as well as what they themselves did in the previous round [92]. These results highlight the importance of direct reciprocity, even when interacting with several connected players. However, when the number of interaction partners is more than just one, cooperation systematically decays, which is a common theme in multiplayer social dilemmas [72, 150]. Even when players use conditional cooperation and attempt to reciprocate, when the number of partners is too large, they struggle [151].

However, most social networks are not static. Humans are usually able to end relationships with defectors and instead initialize interactions with other cooperators. Direct reciprocity in large networks of connected individuals can lead to cooperation if players can adjust their ties. To address this, [152] and [153] investigate cooperation in dynamical networks. Here, players can cut their link to their neighbors when they are not satisfied with the relationship. Under this setup, cooperation is greatly enhanced as players learn to break ties with defectors. Moreover, this positive effect persists even if participants need to pay a substantial cost to cut ties [154]. These results suggest that the mere possibility to quit an interaction is effective in promoting cooperation.

Dynamic networks can be realized in many different ways. [155] allowed players to choose their new partner. Any new link had to be accepted by both parties and there was no upper limit on the number of connections of a player. The authors find that if the benefit of cooperative relationships is large enough compared to the cost of cooperating with a defector, players make the rational decision to create new cooperative ties rather than sever defective ones. Interestingly, this leads to a proliferation of defectors and lowers overall cooperation in the network. [156] look at players' movements in a grid where they can choose their location relative to their neighbors. They find that cooperators do indeed cluster together. However, those cooperators at the boundaries get tired of being exploited by their defecting neigbors and start defecting, too. This leads cooperation to unravel. Nevertheless, in the vast majority of network experiments, subjects have to choose one action for all neighbors (see [152] , for an exception). This design choice does not allow for proper one-to-one direct reciprocal relationships and does not treat the interactions as independent. The results of those studies could be very different if subjects were allowed to give targeted responses to each neighbor.

Another interesting variation on the classical prisoner's dilemma arises when people can make their decisions in real time [134, 150]. In corresponding experiments, players no longer make decisions in well-defined rounds. Rather they can choose with which action to start (cooperation or defection). After that, the game unfolds in continuous time, and people can revise their chosen action at any given point. Compared to the classical setup, this experimental design has several features that make it particularly attractive. For one, games tend to last shorter; players no longer need to make a sequence of decisions after which they need to be informed of the co-player's last decision. Rather decisions are made and information is provided in real time, such that supergames are typically finished in 1 or 2 min. At the same time, results from continuous-time experiments seem to be comparable to the classical setup. For example, for the finitely repeated prisoner's dilemma, [134] find that individual behavior is consistent with a conditional cutoff strategy. Participants cooperate until almost the end of the game, unless their opponent defected first, recovering similar results in the conventional repeated prisoner's dilemma [121, 122].

2.7 Discussion

Over the last decades, the repeated prisoner's dilemma has become the standard model for the evolution of direct reciprocity. It encapsulates the idea that individuals can maintain cooperation when they repeatedly interact in stable pairs, or small groups. By now, there is a rich theoretical literature that describes in which environments cooperation is to evolve, and which strategies are most effective in sustaining cooperation. In this article we compare these theoretical results to the empirical literature on human cooperation. Because the empirical literature on the prisoner's dilemma is vast, here we only present a selection of works. For a more comprehensive overview on the empirical literature, we recommend the invaluable resource of the cooperation databank [157], as well as other review articles [27, 158]. Perhaps somewhat surprisingly, our comparison shows that the predictive value of theoretical models is somewhat ambivalent.

On the one hand, models seem to describe reasonably well for which parameters cooperation is most likely to evolve. In particular, the effect of parameter changes is often accurately predicted by these models: for example, increasing the expected length of the game does tend to increase cooperation; and similarly, increasing the payoffs for mutual cooperation makes people on average more cooperative. On the other hand, models seem to be far less successful when it comes to predict the particular strategies that humans would use. For example, in indefinitely repeated games (those without a known end), evolutionary models often stress the success of strategies like WSLS [51, 96, 99, 117, 159]. In fact, this strategy has a number of appealing theoretical properties. It can resist invasion by unconditional cooperators, it is robust with respect to occasional errors and mistakes, and it is evolutionary stable when cooperation is sufficiently valuable[97]. However, most empirical studies find little evidence for behavior consistent with WSLS, even in parameter regions in which this strategy is supposed to be strongly favored (e.g., [131, 132].

There is a number of reasons that might account for this mismatch. For example, evolutionary simulations are often run under rather restrictive parameters assumptions. Most importantly, many studies assume that mutations are rare, which allows researchers to simulate evolutionary processes more efficiently [160]. When mutations are assumed to be rare, most populations tend to be monomorphic, which favors the evolution of equilibrium strategies like WSLS. On the other hand, data from experiments suggests that there is quite some variation in human behavior [131, 132]. In populations with many different strategies, more reciprocal strategies like TFT may have an advantage, because they are less prone to be exploited by any given opponent.

Another limitation of most evolutionary models is that people are often assumed to play each repeated game in isolation. In contrast, most human interactions do not happen in a strict uninterrupted sequence. Rather we engage in games with one individual at one time, only to interact with another group member a few minutes later. To date, there is little theoretical work that can describe how individuals keep an optimal record of their social interactions, and how they should react based on their record. While our discussion of memory constraints suggest that humans tend to remember the general nature of their co-player, there might be interesting interactions between the exact way how people memorize past interactions, and which strategies they use in response.

More generally, much of the previous research, both theoretically and experimentally, is restricted to constrained strategy sets. In particular, researchers often focus on memory-1 strategies, or on some of the simple strategies taken from the classic set described in Table 1. Even when more complex strategies are considered, they are typically longer memory extensions of essentially the same rules (for example, [131], considers nine variants of TFT out of a total set of 20 different strategies used in their analysis). Research might benefit from testing a more heterogenous set of strategies when investigating human behavior in the repeated prisoner's dilemma. These strategies should be explored in different environments, with different error rates and game lengths. [161] and [136] specifically look at very long games and find that the way players punish, exploit, or forgive can be predicted by how long the interaction has the potential to last. Longer games and the presence of errors allow for richer behaviors and strategies, and could make for interesting future research into the dynamics of reciprocity.

Chapter 3

Cooperation in Concurrent Games

This chapter is currently a draft manuscript titled *Dynamics of cooperation in concurrent games*. It is supported by a supplementary information in the appendices, with the details of the evolutionary simulations, empirical methods, statistical analysis, and screenshots of the online task.

Contents

Abstract

Humans frequently encounter situations in which individually optimal decisions run counter to the interests of the group. To navigate such social dilemmas, people often use simple heuristics based on direct reciprocity. They cooperate when others do and stop cooperating once interaction partners defect. Yet most of this work assumes that individuals only interact in one game at a time, or that they treat each game as independent. However, in most real examples, people engage in several games concurrently. In concurrent games, the outcome of one interaction may affect how individuals subsequently behave in a different one. Herein, we introduce a theoretical framework to study the resulting cross-over and spill-over effects. Individuals repeatedly engage in two independent stage games, either with the same or with different partners. They adapt their strategies over time according to an evolutionary learning process. We find that individuals often link their behavior across the two games. This linkage is particularly pronounced when we allow for cognitive constraints, such as imperfect recall or narrow-bracketing. Our theory shows that linkage can have both a positive or a negative effect on cooperation, compared to the standard case of independent games. An additional behavioral experiment, however, suggests that the overall effect of two concurrent games on cooperation tends to be negative. These results highlight how different kinds of strategic motives and spillovers jointly affect and interfere with reciprocity in concurrent games.

3.1 Introduction

Direct reciprocity is one of the core mechanisms enabling cooperation among unrelated individuals [14, 16]. This mechanism is at work when neighbors take turns picking up each others' children from school, when students correct each others' work, or when couples share domestic chores. Experimental work shows that reciprocal relationships emerge naturally if interactions occur repeatedly, provided the probability of another encounter is sufficiently high [27, 136]. Repetition allows individuals to condition their current actions on their interaction partner's past behavior [162]. When they adopt conditionally cooperative strategies such as Tit-for-Tat [50, 163, 164], Generous Tit-for-Tat [165, 166], or generalizations thereof [132, 167–171], even selfish opponents have an incentive to cooperate. Using models of evolutionary game theory, researchers have explored which kinds of strategies evolve, and in which environments reciprocal cooperation is stable [100, 107, 116, 117, 172–175].

Yet most of this work assumes that individuals either only engage in one repeated game at a time, or that they treat each game as independent. This means that both theoretically and experimentally, each ongoing strategic interaction is studied in isolation [27]. This assumption of independence greatly facilitates a theoretical analysis. It allows researchers to consider a comparably small set of possible strategies [176]. Once this assumption is dropped, a player's strategy does not only depend on the opponent's previous actions in the respective game any more. Instead, it may depend on the previous actions of all opponents, across all games. As a result, the cooperation dynamics needs to be described at a different level: instead of the standard game-perspective, models now need to take a population-perspective. This change in perspective drastically increases a model's computational complexity [177]. To circumvent these difficulties, most research is based on the implicit assumption that by analyzing different games individually, one can extrapolate (or at least approximate) how people behave when they engage in many games in parallel. Our aim is to explore to which extent this assumption is justified. We make two key contributions. First, we refute, both theoretically and experimentally, that people generally treat their different games as independent. Second, by taking into account linkage between games, we introduce a novel theoretical framework that gives rise to a richer and more realistic class of game-theoretic models.

Our inquiry is based on the notion of a concurrent game. A concurrent game arises when players engage in several, formally independent, repeated games in parallel. Players may have their different repeated games either with the same or with different interaction partners (**Fig.3.1**). We ask to which extent behavior in the concurrent game can be inferred from the constituent repeated games. While this question has received some attention before, respective models typically take a static equilibrium approach [178, 179]. This research shows, for example, that if players implement an equilibrium for each isolated repeated game, the resulting strategy profile also constitutes an equilibrium of the concurrent game. When all games are identical, symmetric, and played with the same partner, one can even derive a stronger result. In that case, full cooperation is feasible in the concurrent game if and only if it is feasible in each repeated game [178]. In the **Supplementary Information**, we provide a more detailed summary of the relevant literature. These studies greatly illuminate which behaviors are possible in equilibrium. Yet, they do not address which of these equilibria (if any) are most likely to emerge when strategies are not consciously chosen, but learnt over time. Moreover, this existing work does not attempt to study the consequences of several cognitive constraints and behavioral heuristics that might affect human play in concurrent games. For example, effects arising from imperfect recall [102, 135, 141], or from a drive to act consistently, may naturally introduce spillovers between games. Once such spillovers occur, behavior may spread from one game to another [180]. Herein, we study a simple but comprehensive theoretical framework to describe these effects.

We consider three idealized scenarios, to which we refer as treatments. In all treatments, players engage in two different repeated social dilemmas. The two dilemmas either result in a high or a low benefit of cooperation (**Fig.3.1a**). The three treatments differ in whether or not players treat each repeated game as independent, and in whether or not the two games are played with the same or with different interaction partners. In the first treatment, the *control*, we consider the baseline case typically studied in the literature (**Fig.3.1b**). Here, individuals play each repeated game in isolation. Hence they treat each repeated game as independent by design. Second, in the *same-partner treatment*, the two games are played simultaneously, and with the same opponent (**Fig.3.1c**). This setup has been previously termed a 'multichannel game' [176], since players can interact and influence each other through multiple channels. As a result, players can react to an opponent's defection in one game by defecting in the other. In this way, we aim to capture a players' strategic motives to link their behavior across different games. This linkage may provide players with a stronger leverage to enforce cooperation. Third, in the *different-partners treat-*

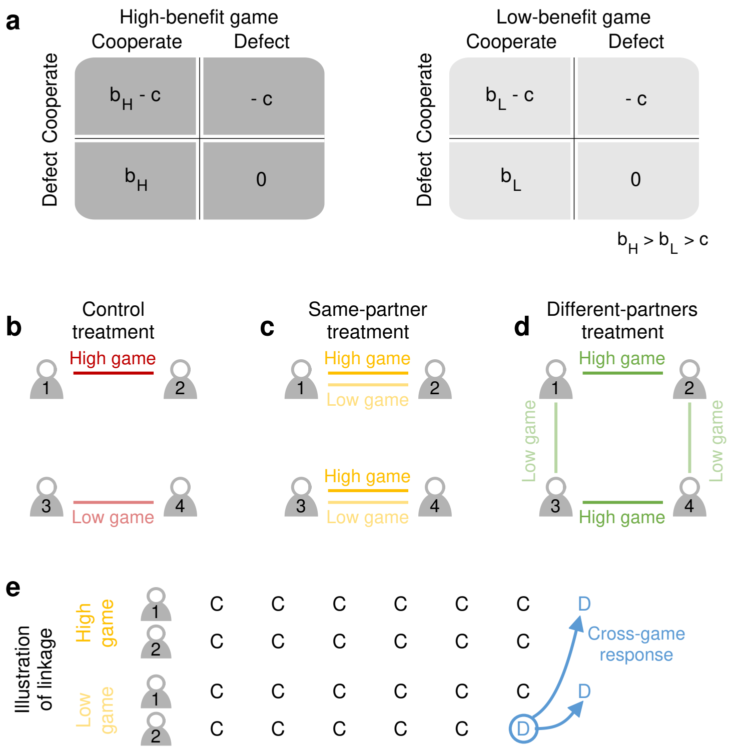


Figure 3.1: A framework of concurrent games. a, In concurrent games, players engage in two or more games simultaneously. Herein, we consider the case that players engage in two games, one with a high benefit of cooperation ('high game') and one with a smaller benefit ('low game'). Each payoff matrix describes the payoff of the player who picks a row, depending on the co-player's choice who picks a column. **b,** In the control treatment, players only engage in one repeated game at a time, as usually assumed in the literature. **c,** In the same-partner treatment, each player engages in both games but with the same co-player. **d,** In the different-partners treatment, players engage in both games but with different co-players. **e,** Concurrent games allow for linkage. Players might respond to a co-player's defection in one game by defecting in both games. Such linkage may arise both in the same-partner treatment (depicted here) and in the different-partners treatment.

ment, individuals play the two games simultaneously, but with a different co-player in each game (**Fig.3.1d**) . By comparing the control with the other treatments, we explore in which case players treat each game as independent. By comparing the same-partner with the different-partners treatment, we explore how strategic motives contribute to the evolution of cross-game effects. Finally, with various model extensions, we study the impact of several plausible cognitive constraints and behavioral heuristics.

In the same-partner treatment, and to a far lesser extent in the different-partners treatment, we find that a player's behavior in one game is linked to the previous outcome of the other game. This linkage can either result in more or less cooperation compared to the control, depending on the treatment and on the presence of cognitive constraints. To further explore these theoretical results, we have run a behavioral experiment that implements our three treatments, based on a similar design as in previous empirical studies [181, 182]. In the experiment, both the same-partner treatment and the different-partners treatment result in less cooperation than the control. For the same-partner treatment, our empirical data not only rules out that players treat each game as independent, but also calls into question a previous prediction that concurrently ongoing games among the same partners would enhance cooperation [176]. Our results have important implications for the effectiveness of direct reciprocity. People in their daily lives often engage in several games concurrently. For such concurrent games, we find that strategic motives, spillovers, and cognitive constraints can easily affect, and often undermine, cooperation.

3.2 Results

3.2.1 A model of concurrent games

We study cooperative interactions based on a variant of the prisoner's dilemma, the donation game [60]. In this game, players either cooperate (C) or defect (D). Cooperation means to pay a cost c for the partner to get a benefit b. Defection means to pay no cost and for the partner to get no benefit. We consider two different implementations of this game (**Fig.3.1a**). In one implementation, the benefit is high, and we accordingly speak of the high-benefit game, or *high game* (H). In the other implementation, the benefit is smaller, and we call it the *low game* (L). Assuming $b_{\mathtt{H}} \geq b_{\mathtt{L}} > c$ throughout, the dominant action if players only meet once is to defect in either game. However, we assume players interact for infinitely many rounds (an extension to finitely repeated games will be discussed later). We refer to each iterated donation game as a *repeated game*. When players engage in both donation games in parallel, such that players make two choices each round (one for each game), we speak of a *concurrent game*. A large literature shows that cooperation is feasible in repeated games [162]. This result naturally extends to concurrent games. Here, we are interested in how likely cooperation is to evolve in concurrent games, and which strategies are used to sustain it.

To this end, we discuss three different idealized scenarios (treatments) of how these games unfold. In each case, we consider four players. In the *control treatment*, players only engage in a single repeated game at a time, with a fixed partner (**Fig.3.1b**). One pair of players repeatedly engages in the high game, whereas the other pair plays the low game. Players use reactive strategies to make their decisions. This means that a player's choice whether or not to cooperate in a given round only depends on the co-player's decision in the previous round. Reactive strategies take the following form [60],

$$\mathbf{p} = (p_C^k, p_D^k) \in [0,1]^2. \tag{3.1}$$

Here, p_a^k is the player's probability to cooperate in game $k \in \{\mathtt{H}, \mathtt{L}\}$, depending on the co-player's previous action $a \in \{C, D\}$. For example, a player with strategy $\mathbf{p} = (1, 0)$ implements Tit-for-Tat (TFT). A player with $\mathbf{p} = (1, p_D)$ and $0 < p_D < 1$ uses Generous Tit-for-Tat [GTFT, see 165, 166]. Finally, a player with $\mathbf{p} = (0, 0)$ defects unconditionally (ALLD).

We contrast this control treatment with two different kinds of a concurrent game. In the first one, the *same-partner treatment*, players are again matched with a single partner, but the two players interact in both repeated games simultaneously (**Fig.3.1c**). In particular, their decision in either game may depend on how the co-player acted in the other game. Reactive strategies for the same-partner treatment take the form

$$\mathbf{p} = (p^{\mathtt{H}}_{CC}, p^{\mathtt{H}}_{CD}, p^{\mathtt{H}}_{DC}, p^{\mathtt{H}}_{DD};\; p^{\mathtt{L}}_{CC}, p^{\mathtt{L}}_{CD}, p^{\mathtt{L}}_{DC}, p^{\mathtt{L}}_{DD}) \in [0,1]^8. \tag{3.2}$$

Here, $p^k_{a^{\mathtt{H}}a^{\mathtt{L}}}$ is the player's probability to cooperate in game $k \in \{\mathtt{H}, \mathtt{L}\}$, depending on the co-player's previous decisions in both the high and the low game, $a^{\mathtt{H}}, a^{\mathtt{L}} \in \{C, D\}$. We say such a strategy treats both games as independent if the entries satisfy

$$p^{\mathtt{H}}_{CC} = p^{\mathtt{H}}_{CD},\;\; p^{\mathtt{H}}_{DC} = p^{\mathtt{H}}_{DD} \quad \text{and} \quad p^{\mathtt{L}}_{CC} = p^{\mathtt{L}}_{DC},\;\; p^{\mathtt{L}}_{CD} = p^{\mathtt{L}}_{DD}. \tag{3.3}$$

That is, the strategy only reacts to the co-player's previous action in the respective game, irrespective of the outcome of the other game. In the case of independence, strategies of the control treatment naturally map to strategies in the same-partner treatment. For example, if a player in the control were to use TFT in the high game and ALLD in the low game, that player could implement $\mathbf{p} = (1,1,0,0;\; 0,0,0,0)$ in the concurrent game. Thus, the same-partner treatment permits all strategic behaviors that are feasible in the control. In general, however, the set of feasible strategies is strictly larger in the same-partner treatment. For example, players with $\mathbf{p} = (1,0,0,0;\; 1,0,0,0)$ only cooperate in either game if the co-player previously cooperated in *both* games. When the constituent games are not treated as independent, we say players link their behavior across games. Accordingly, we also speak of linkage. Examples like the one above illustrate that linkage might arise because of strategic motives. By doing so, players may be able to enforce cooperation more effectively, by threatening to defect in both games after any deviation of the co-player (**Fig.3.1e**).

The last treatment is the *different-partners treatment*. Here, players again engage in both the high and the low game simultaneously, but now with different co-players (**Fig.3.1d**). Reactive strategies for this treatment have the same complexity as in the same-partner treatment, see Eq. (3.2). Also the definition of independence is the same, see Eq. (3.3). From a strategic viewpoint, however, this treatment differs from the same-partner treatment. With different partners involved, there is less of an immediate strategic motive to link behavior across games, unless players wish to adopt a strategy of community-enforcement [183, 184].

For all three treatments, we can compute the players' payoffs explicitly. To this end, we represent the interaction as a Markov chain that depends on the players' strategies. We describe the respective procedure in the **Methods** and in the **Supplementary Information**. However, we do not regard the players' strategies as fixed. Rather, as usual in evolutionary game theory, players update their strategies over time based on their payoffs. To model this updating process, we use introspection dynamics [64, 185]. According to this process, players regularly compare their current payoff with the payoff they could have obtained by using a (randomly sampled) alternative strategy. The higher the payoff of the alternative, the more likely players are to switch. If we apply this learning process to the three treatments, and if we artificially require players in the last two treatments to treat each repeated game as independent, all treatments yield equivalent results (**Fig. S1**). In particular, all treatments recover the qualitative findings of the previous literature on direct reciprocity [90]. In the following, we systematically explore the effect of linkage, by no longer imposing that players treat each game as independent.

3.2.2 Introspection dynamics of concurrent games

To get a first impression, we simulate the learning dynamics in the three treatments for fixed parameter values (in particular, we set $b_{\mathtt{H}} = 5$, $b_{\mathtt{L}} = 3$ and $c = 1$). Results in the control treatment recover the conventional wisdom established by previous work in direct reciprocity [90]. Repetition allows players to achieve some cooperation, and players are more cooperative when there is a high benefit (**Fig.3.2a**). These intuitive results differ from what we find in both other treatments. In the same-partner treatment, individuals frequently cooperate in both games (**Fig.3.2b**). These results confirm work by Donahue *et al.* [176] where players adopt strategies based on social comparisons. In contrast, in the different-partners treatment, cooperation rates are consistently low (**Fig.3.2c**). Because all three treatments yield equivalent results if players are artificially restricted to treat each game as independent (**Fig. S1**), these results indicate that linkage affects the cooperation dynamics. This effect is predicted to be favorable in the same-partner treatment, whereas it is detrimental when people play their games with different partners.

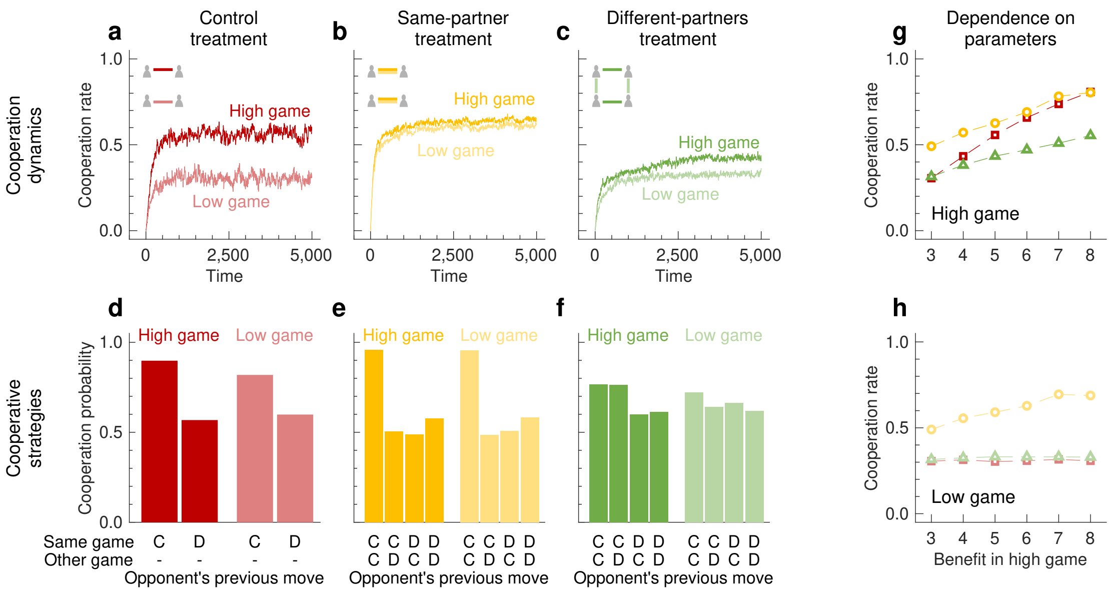


Figure 3.2: Dynamics of cooperation in concurrent games. a-c, We use introspection dynamics [64, 185] to model how people learn to cooperate in the three treatments. Here, we show average cooperation rates for both the high and the low game over time, averaged over 100 simulations. In the control treatment, there is substantially more cooperation in the high game than in the low game, as expected. In the same-partner treatment, players are generally more cooperative, whereas in the different-partners treatment, players tend to cooperate less. **d-f,** We have recorded which strategies players use when they cooperate in both games (at least 2/3 cooperation rate). In the control treatment, players adopt strategies consistent with Generous Tit-for-Tat [165, 166]. In the same-partner treatment, they only cooperate if the co-player previously cooperated in both games. In the different-partner treatment, individuals are still most cooperative after receiving cooperation in both games, but there is overall less cooperation (and very little linkage). **g,h,** Our qualitative results remain valid for a wide range of parameter values (see also **Fig. S3**).

To explore the magnitude of linkage, we record the players' strategies during the learning process. In **Fig.3.2d-f**, we report results for cooperative players (those with a cooperation rate of at least 2/3 in each game they participate in). In the control treatment, such players use strategies similar to Generous Tit-for-Tat [165, 166], as one may expect. They tend to fully reciprocate a co-player's cooperation, and they show some leniency with defecting co-players (**Fig.3.2d**). In contrast, evolving behaviors in the same-partner treatment are more strict. Here, players only fully cooperate in either game if the co-player previously cooperated in both games (they still show some leniency with respect to partial or full defectors, **Fig.3.2e**). Importantly, these strategies exhibit linkage. Players condition their behavior in one game on actions that occurred in the other. Further simulations suggest that such strategies evolve in the same-partner treatment because they are more stable, compared to a strategy that just uses Generous Tit-for-Tat in each game (**Fig. S2**). Finally, for the different-partners treatment, players are unlikely to cooperate in both games altogether. Even when both co-players cooperated in the previous round in their respective games, players are on average less likely to reciprocate, and there is also little linkage overall (**Fig.3.2f**).

Overall, the same-partner treatment results in more cooperation whereas the different-partners treatment leads to reduced cooperation rates. These qualitative findings are robust, independent of the exact benefit of cooperation (**Fig.3.2g,h**), of whether or not players commit implementation errors (**Fig. S3a**), or whether or not the game is infinitely repeated (**Fig. S3b**). These framework can be extended to describe scenarios beyond the donation game, and beyond the prisoner's dilemma more generally (**Fig. S4**). We conclude that linkage in concurrent games has substantial effects on cooperation, even for the most basic models of direct reciprocity.

3.2.3 Incorporating cognitive constraints and different learning heuristics

Our framework allows us to go beyond a mere comparison between concurrent games and classical models of direct reciprocity. Instead, we can also explore the consequences of several cognitive constraints that are impossible to study (or have no analogue) in classical single repeated games. In the following, we discuss four constraints and heuristics that may conceivably affect behavior in concurrent games. In each case, we briefly summarize how they can be incorporated into our framework and how they affect our results. For all derivations and a more detailed discussion, we refer to the **Supplementary Information**.

The first model extension addresses the impact of *imperfect recall*. Everyday experience and previous experiments [102, 135, 141] suggest that people with several interactions may confuse past outcomes. A co-player's cooperation in one game may be misremembered as having happened in a different game, possibly with a different co-player. To capture this form of imperfect recall across games, we assume players confuse past outcomes with probability $\varepsilon_{\text{IR}} \geq 0$. When such an error occurs, instead of correctly recollecting the previous actions in the high and the low game as $(a^{\mathtt{H}}, a^{\mathtt{L}}) \in \{C, D\}^2$, the player takes them to be $(a^{\mathtt{L}}, a^{\mathtt{H}}) \in \{C, D\}^2$. As

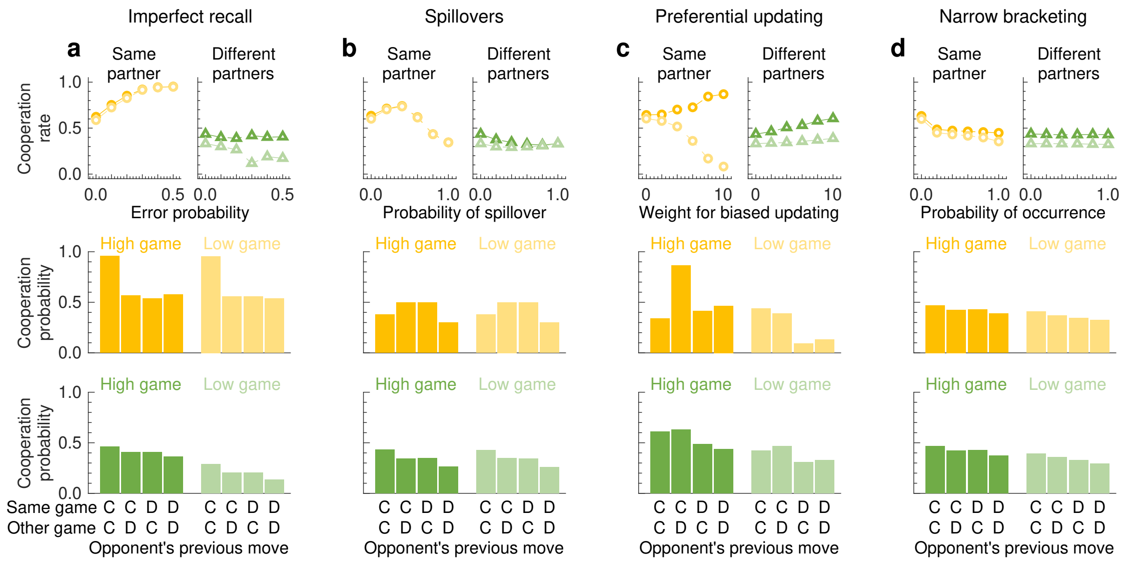


Figure 3.3: Modelling the effect of cognitive constraints in concurrent games. Our framework allows us to study the effect of various constraints, biases, and heuristics on cooperation in concurrent games. Here, we explore the impact of **a,** imperfect recall, **b,** spillovers, **c,** preferential updating in the game with lower payoffs, and **d,** narrow bracketing. In each case, we record the impact on average cooperation rates (upper panel). In addition, we also record the evolving average strategies in the most extreme case (lower two panels). For the same-partner treatment, we find that spillovers and narrow bracketing are most detrimental to cooperation. In that case, average cooperation rates may even be below the cooperation rates of the baseline control treatment (**Fig.3.2a**).

a result, the player cooperates with probability $p^k_{a^{\text{L}},a^{\text{H}}}$ instead of $p^k_{a^{\text{H}},a^{\text{L}}}$. Errors of this kind have no effect if the previous outcome is either (C,C) or (D,D), or if the player's strategy happens to satisfy $p^k_{CD} = p^k_{DC}$. In the first case, no confusion between the two games can arise, whereas in the second case, any confusion proves to be inconsequential. Errors of imperfect recall can arise both in the same-partner treatment and the different-partners treatment. Yet they may have more of an effect when interacting with different partners, as they might lead players to give misdirected responses [102, 141]. In line with this intuition, we find that such errors have a weakly negative effect on cooperation in the different-partners treatment (**Fig.3.3a**). Perhaps surprisingly, however, we find that imperfect recall reinforces cooperation in the same-partner treatment. Here, errors provide further incentives for players to link their behavior across games, and to only cooperate if the co-player previously cooperated in both games (**Fig.3.3a**).

The second model extension addresses (exogenous) *behavioral spillovers*. A spillover arises when an individual's action in one domain leads that individual to take the same action in a different domain. Such spillovers have been reported in various contexts, and they can have important policy implications [186, 187]. In our context, spillovers introduce additional correlations into a player's behavior. For any given history, they increase the chance that a player chooses the same action in each of the two games (rather than cooperating in one game and defecting in the other). For the same reason as before, such correlations seem particularly harmful when interacting with different partners because they undermine a player's ability to give targeted responses. Indeed, simulations again suggest a weakly negative effect of spillovers in the different-partners treatment (**Fig.3.3b**). In contrast, in the same-partners treatment, the effect can be both positive and negative, depending on how frequent spillovers are. Indeed, in some cases the resulting cooperation rates may even be below the cooperation rates of the control treatment (**Fig.3.2a**).

The next two model extensions address different ways how people might update their strategies in the two games. In our previous simulations, we assume that players are equally likely to update their strategy in either the high or the low game. Instead, players may be more inclined to update their strategy in the game in which they currently receive the smaller payoff (relative to the maximum feasible payoff in that game). Simulations suggest that such *preferential updating* has weakly positive effects in the different-partners treatment. In the same-partner treatment, it increases cooperation in the high game but it destabilizes cooperation in the low game (**Fig.3.3c**), presumably because players now update their low-game strategies more often.

Our last model extension addresses *narrow bracketing*. Narrow bracketing refers to situations in which people make decisions in one domain, without fully internalizing the consequences of those decisions in a different domain [188]. Such a bias may also affect how people learn in concurrent games. When players update their strategies in one game (high or low), they may not anticipate how these changes affect the dynamics of the other game. Narrow-bracketing has limited effects when people naturally treat their games as independent. In that case, changes in one game's strategy have no effect on the dynamics of the other. As a result, simulations suggest that narrow bracketing has no discernible impact in the different-partners treatment (**Fig.3.3d**). However, in the same-partner treatment, in which players naturally learn to link their behavior across games, the effects can be considerable. Here we find that narrow-bracketing undermines cooperation, both in the high and the low game.

Overall, our framework can readily capture each of the four cognitive constraints and learning heuristics discussed above. These model extensions highlight the additional complexities that arise when individuals engage in several repeated games in parallel.

3.2.4 Human behavior in concurrent games

These theoretical results indicate that concurrent games alter the dynamics of reciprocal interactions. But whether concurrent games lead to more or less cooperation depends on how they update their strategies, whether their decision-making is influenced by biases and heuristics, and whether they interact with the same or with different partners. To explore the actual cooperation dynamics among human participants in more detail, we conducted a behavioral experiment. The experiment directly implements the three treatments illustrated in **Fig.3.1b-d**. Participants are randomly assigned to treatments, and in the control treatment, they are randomly assigned to either play the high or the low game. In the high game, players can pay 2 points to give 4 points to the other player. In the low game, a player's 2 points are translated into 3 points for the co-player. Participants interact for at least 20 rounds, with a stochastic stopping rule implemented thereafter.

Fig.3.4a shows the resulting average cooperation rates across three treatments. In contrast to the predictions of the baseline model, but in agreement with some of our model extensions, we find that people are most cooperative in the control treatment. More specifically, the average cooperation rate in the high game is 78.6% in the control, compared to 60.5% in the same-partner and 62.4% in the different-partners treatment ($p = 0.004$ same-partner, $p = 0.006$ different-partners). Similarly, cooperation rates in the low game are 70.3% in the control, compared to 54.0% for the same-partner treatment and 52.2% for the different-partners treatment ($p = 0.056$ same-partner, $p = 0.007$ different-partners). Interestingly, these differences in cooperation rates are already present in the first round and they are stable throughout the experiment (**Fig.3.4b**). These results suggest that the simultaneous presence of two games interferes with the emergence of reciprocal cooperation in either game.

	Dependent variable:			
	Cooperation			
	Same-partner		Different-partners	
	High game	Low game	High game	Low game
Partner's previous decision in the high game ($C_{H,t-1}$)	0.450^{***}	0.157^{**}	0.637^{***}	0.023
	(0.050)	(0.041)	(0.103)	(0.059)
Partner's previous decision in the low game ($C_{L,t-1}$)	0.189^{***}	0.264^{**}	0.095^{*}	0.513^{***}
	(0.042)	(0.055)	(0.055)	(0.082)
Interaction ($C_{H,t-1}$) × ($C_{L,t-1}$)	0.068	0.275^{**}	−0.079	0.070
	(0.050)	(0.058)	(0.105)	(0.090)
Constant	0.198^{***}	0.181^{***}	0.205^{***}	0.210^{***}
	(0.040)	(0.040)	(0.044)	(0.033)
Observations	1,368	1,368	2,432	2,432
R^2	0.401	0.395	0.375	0.327
Adjusted R^2	0.400	0.394	0.374	0.326
Residual Std. Error	0.379 (df = 1364)	0.388 (df = 1364)	0.384 (df = 2428)	0.410 (df = 2428)
Note:				*p<0.05; **p<0.01; ***p<0.001

Table 3.1: A linear regression to estimate the magnitude of linkage in human participants. Based on the data of our behavioral experiment, we estimate how likely participants are to cooperate, depending on their partner's previous behavior. In total, we have run six regressions (three treatments, in which two games are played each round). If participants treat each game as independent, we would expect that only the constant term and the partner's previous decision in the respective game affect a player's cooperation probability. However, in the same-partners treatment, we observe that also previous decisions in the other game have a significant impact (in addition, in the low game we observe a significant interaction of cooperation in the two games). In the different-partners treatment, we only observe a weak impact of the low game on high-game decisions. All other indicators for linkage are insignificant.

To explore to which extent linkage might drive these results, we infer the participants' reactive strategies based on the experimental data. For any possible outcome of the previous

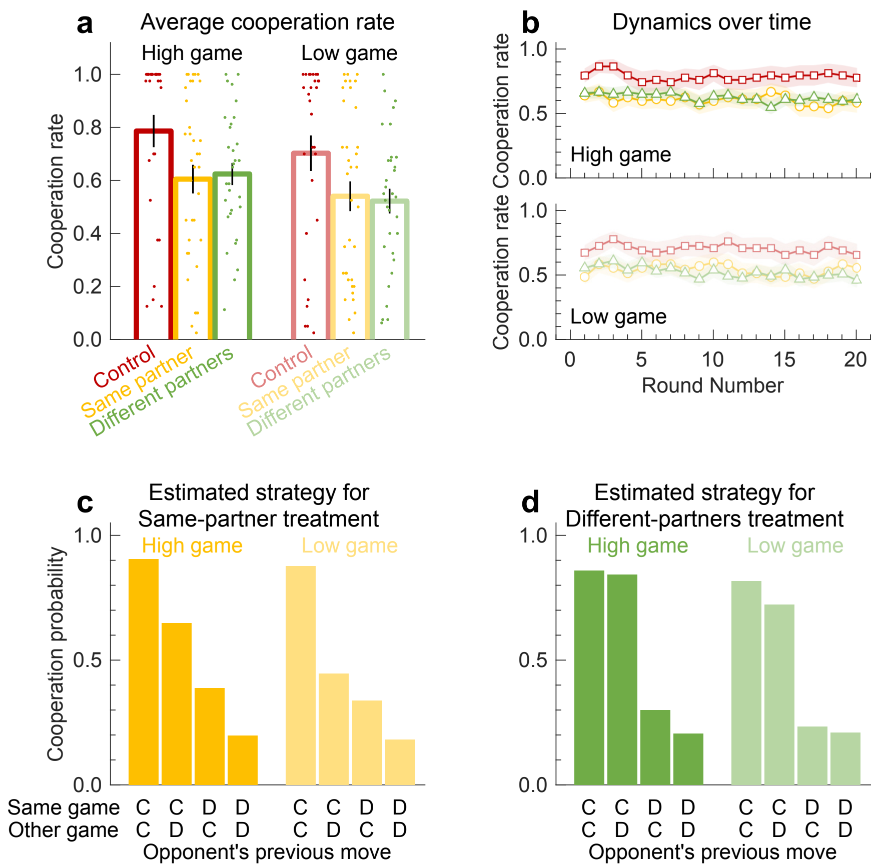


Figure 3.4: Concurrent games among humans. To explore how people act in concurrent games, we have implemented a behavioral experiment using the three treatments in **Fig.3.1**. Participants are randomly matched and interact for at least 20 rounds. After that, the game continues with 50% probability each round. **a,** Across all rounds, people were most cooperative in the control treatment, in both the high and the low game. Cooperation rates in the same-partner and the different-partner treatment are not significantly different from each other. **b,** These qualitative results are already present in the first round, and they are stable throughout the experiment. **c,d,** We use a linear regression to estimate the players' strategies based on the co-player's behavior in the previous round (**Table 3.1**). Here, we visualize the resulting conditional cooperation probabilities. In the same-partner treatment, participants link their behavior across games. As a result, a player's cooperation probability depends on the previous outcome of both games. In comparison, behaviors in the different-partners treatment are largely independent across the two games. In **a**, error bars depict standard errors and dots represent individual data points; in **b**, shaded areas depict standard errors.

round, we estimate how likely participants are to cooperate in the next round, both for the high and the low game. The results are summarized in **Fig.3.4b** and **Table 3.1**. In line with our earlier simulations, linkage is more pronounced in the same-partners treatment. For example in the high game, a linear regression suggests that participants cooperate with 90.5% probability if the co-player previously cooperated in both games. If the co-player only cooperated in the high game, this cooperation probability drops to 64.8%. In comparison, linkage is much weaker in the different-partners treatment. For example, people cooperate with 85.8% probability in the high game after receiving cooperation in both games. This number drops only marginally, to 84.2% when a player only received cooperation in the high game. More generally, **Table 3.1** suggests that linkage only has a minor effect in the different-partners treatment. Still, overall cooperation rates are below the control treatment, because players generally have lower cooperation probabilities (see **Table S1** for the regression results for the control treatment).

Overall, and in line with our theoretical results, we observe the strongest linkage effects in the same-partner treatment. However, we also find that participants do not benefit from this linkage. Instead of using it to better enforce cooperation, participants end up cooperating less often than participants in the control treatment. As a consequence, concurrent games result in reduced cooperation rates, independent of whether people have their games with the same partner or with different partners.

3.3 Discussion

People routinely engage in several social interactions at once [189]. They cooperate with their friends, their colleagues, and their families, possibly all at the same time. Moreover, with any given interaction partner, people often have several independent interactions in parallel. Colleagues might work on several projects concurrently, and whole nations routinely interact and negotiate over a wide array of different policies [3]. Despite this prevalence of concurrent games, the main paradigm for direct reciprocity is to study cooperation in (isolated) repeated games. Such an approach is justified (and from a computational perspective even preferred) when people treat each game as independent. However, herein we present a theoretical framework and experimental data that cast serious doubt on that assumption of independence.

For our theoretical analysis, we compare three idealized scenarios. In one scenario—the control—individuals only engage in one repeated game at a time, just as previously assumed by most of the literature. In the other two scenarios, individuals engage in two repeated games simultaneously, either with the same partner or with different partners. If individuals in these last two scenarios indeed were to treat each of their games as independent, all three scenarios yield indistinguishable results (**Fig. S1**). Yet as individuals learn to adopt more profitable strategies over time in an evolutionary process, we often find that they learn to link their behavior across games. This linkage is particularly pronounced when the different games take place among the same partners (**Fig.3.2**), in which linkage can come with explicit strategic benefits [176–179].

By shifting the perspective from individual to interconnected games, our framework serves as a starting point to better describe the effects of different cognitive constraints and biases (**Fig.3.3**). The previous literature on direct reciprocity focuses on implementation errors, or 'trembling hands' [60]. Such errors occur when individuals intend to cooperate but fail to do so, perhaps because of a lack of attention or of resources. Previous empirical research, however, has documented a plethora of other constraints that conceivably affect how humans cooperate. For example, the work of Stevens and colleagues [102, 141] shows how imperfect recall can undermine a person's ability to give directed responses. Our work suggests that the effects of imperfect recall depend on the previous history of interactions. People are only susceptible to this kind of error when they have made conflicting experiences, with cooperation in one game and defection in another. Moreover, the precise effects of imperfect recall also depend on an individual's strategy. While some strategies are sensitive to false recollections, others remain completely unaffected. In addition to imperfect recall, our framework can also capture several other plausible constraints and heuristics, such as spillovers between games, preferential strategy updating, and of narrow-bracketing. In this way, our framework systematically increases the scope of models of direct reciprocity.

While these model extensions add further realism, they make predictions more complex. For example, cooperation rates in concurrent games may be higher or lower than in classical single repeated games, depending on whether games take place among the same or with different partners, and depending on the constraints that affect individual play. To explore the impact of concurrent interactions on human behavior we conducted a behavioral experiment. In line with recent work by Laferrière and colleagues [182], we find that overall cooperation rates in the same-partner and different-partners treatments are surprisingly similar. Yet the strategies that people apply are markedly different (**Fig.3.4**, **Table 3.1**). While participants in the same-partner treatment routinely learn to link their behavior across games, linkage is comparably weak if games take place among different partners. Moreover, both of these treatments result in less cooperation than the control with a single repeated game (a treatment that Laferrière *et al.* do not consider). These empirical results suggest that strategic considerations, spillovers, and cognitive constraints in concurrent games may overall hamper the evolution of reciprocal relationships.

These empirical results also put some natural bounds on previously suggested mechanisms for cooperation. First, in contrast to previous models of games among the same partners, concurrent games do not seem to promote reciprocity [176–179]. They rather make cooperation more fragile. Second, in contrast to previous work on generalized reciprocity and community enforcement [183, 184], people in the different-partners treatment do not seem to be prepared to exploit their network structure to promote cooperation in the group. Here, too, the effect of multi-game contact appears to be negative.

Overall, our results suggest that models of direct reciprocity based on (single) repeated games only provide an incomplete picture of the reciprocal interactions around us. In concurrent games, individual experiences in one game can affect future behaviors in another. Such linkage between games lead to a richer dynamics, but they also make the emergence of reciprocal altruism more complex.

3.4 Methods

In the following, we briefly summarize our theoretical and experimental methods. All details can be found in the **Supplementary Information**.

3.4.1 Calculation of payoffs

For each treatment, we compute the players' payoffs by representing the game as a Markov chain. The possible states of this Markov chain are the possible outcomes of a given (repeated or concurrent) game. For example, in the control treatment, consider players 1 and 2, who interact in a repeated donation game with high benefits. The possible outcomes of a given round are the four possible realizations $\mathbf{a}=(a^1,a^2)\in\{C,D\}^2$. Given the players' strategies, and given the action profile $\mathbf{a}$ of the previous round, we can compute the probability $m_{\mathbf{a},\tilde{\mathbf{a}}}$ that players choose actions according to the profile $\tilde{\mathbf{a}}=(a^1,a^2)$ in the next round, for each $\tilde{\mathbf{a}}\in\{C,D\}^2$. By computing all the possible transition probabilities, we derive a 4×4 transition matrix $M=(m_{\mathbf{a},\tilde{\mathbf{a}}})$ that captures the dynamics of the repeated game. The respective invariant distribution $\mathbf{v}=(v_{CC},v_{CD},v_{DC},v_{DD})$ describes how often we are to observe each possible outcome $(a^1,a^2)\in\{C,D\}^2$ on average. Given this invariant distribution, payoffs are given by

$$\begin{aligned}\pi^1 &= (v_{CC}+v_{DC})b_{\mathtt{H}} - (v_{CC}+v_{CD})c.\\ \pi^2 &= (v_{CC}+v_{CD})b_{\mathtt{H}} - (v_{CC}+v_{DC})c.\end{aligned} \tag{3.4}$$

In the other treatments, payoffs can be calculated similarly, even though they require more computation. In the same-partner treatment, the possible outcomes of an interaction between players 1 and 2 are now given by a 4-tuple $\mathbf{a}=(a^{1\mathtt{H}},a^{2\mathtt{H}},a^{1\mathtt{L}},a^{2\mathtt{L}})$. Here, an entry $a^{ik}\in\{C,D\}$ represents player i's action in game k. Because each entry can take one of two values, there are now 16 possible outcomes. Hence the corresponding transition matrix is of size 16×16.

In the different-partner treatment, all four players need to be considered simultaneously. Therefore, the current state is now represented by an 8-tuple $\mathbf{a}=(a^{1\mathtt{H}},a^{2\mathtt{H}},a^{3\mathtt{H}},a^{4\mathtt{H}},a^{1\mathtt{L}},a^{2\mathtt{L}},a^{3\mathtt{L}},a^{4\mathtt{L}})\in\{C,D\}^8$. It follows that the state space has $2^8=256$ elements. Hence, calculating the players' payoffs requires the invariant distribution of a 256×256 transition matrix.

Throughout the main text, our model is based on the assumption that people use reactive strategies to make their decisions, and that each game is infinitely repeated. Neither of these assumptions is strictly required. In fact, the computational complexity of the model is unchanged if we assume players to use so-called memory-1 strategies instead [60]. In that case, a player's action does not only depend on the co-players' actions in the previous round, but also on the own previous actions. Similarly, the computational complexity is unchanged if we assume games are repeated with a constant continuation probability δ. Also in that case, payoffs follow from computing the invariant distribution of a 4×4, 16×16, and 256×256 matrix. The respective algorithm for the case of the control treatment is described, for example, in Ichinose and Masuda [190]. In **Fig. S3**, we show simulation results for $\delta<1$.

3.4.2 Description of the learning process

For our theoretical analysis, we take an evolutionary approach. Players adapt their strategies over time, depending on their payoffs (which in turn depend on the strategies of the other players). To model this adaptation process we use introspection dynamics. Compared to other processes such as pairwise imitation [191], introspection dynamics has computational advantages and it is easier to simulate [64, 185]. Moreover, it is the more natural dynamics when players are asymmetric (for example, because they differ in the co-players they interact with). In such cases, strategies that yield a high payoff for one player are not necessarily advantageous for another player, which makes pairwise imitation less plausible.

In the following, we describe our learning dynamics in detail. Learning happens in discrete time steps. For a given treatment, we assume that at time $t=0$, players defect unconditionally, $\mathbf{p}^i=(0,\dots,0)$ for all players i. At each subsequent time point t, the following elementary updating procedure happens. First, one of the players, say player j, is chosen at random. This player is then given an opportunity to revise its current strategy $\mathbf{p^j}$ in one of the two games. In the control treatment, the revision occurs in the one game the player is involved in. In the other two treatments, this is done by randomly choosing one of the two games $k\in\{\mathtt{H},\mathtt{L}\}$. In each case, we replace player j's strategy for game k with a random strategy sampled from a uniform distribution (j's strategy for the other repeated game is left unchanged). The player then compares this alternative strategy $\tilde{\mathbf{p}}^{\mathbf{j}}$ to the current strategy $\mathbf{p}^j$. To this end, let π^j be the payoff the player obtained with the current strategy. Similarly, let $\tilde{\pi}^j$ denote the payoff player j would have got in the previous interaction when adopting strategy $\tilde{\mathbf{p}}^{\mathbf{j}}$ instead (keeping the strategies of the other players unchanged). Player j switches to the new strategy with a probability given by the Fermi-function [192, 193],

$$\rho=\frac{1}{1+\exp\left[-\beta(\tilde{\pi}^j-\pi^j)\right]}. \tag{3.5}$$

The parameter $\beta\geq 0$ is the strength of selection. It measures to which extent strategy updates depend on payoffs. For $\beta\to 0$, payoffs are irrelevant and the updating probability approaches one half. In this limit of 'weak selection', updating occurs at random. In the other limit of 'strong selection', $\beta\to\infty$, only those alternative strategies are adopted that yield at least the payoff of the original strategy.

We iterate this elementary updating procedure for many time steps. For any finite selection strength β, this generates an ergodic stochastic process. In particular, the players' average cooperation rates (over the course of the learning process) converge in time, and they are independent of the players' initial strategies. For our study, we use simulations to numerically estimate these average rates for all three treatments.

3.4.3 Computational methods used for the figures

For the simulations in the main text, we use the following default parameters. The benefits in the two games are $b_{\mathrm{H}} = 5$ and $b_{\mathrm{L}} = 3$, respectively, and the cost is $c = 1$. In addition, we consider a selection strength of $\beta = 200$, and no trembling-hand (implementation) errors, $\varepsilon_{\mathrm{TH}} = 0$. In **Fig.3.2a-c** we show average trajectories. To this end, we have run 100 independent simulations for each treatment. To keep timescales comparable, simulations are run for 20,000 elementary time steps in the control treatment, and for 40,000 elementary time steps in the other two treatments. This implies that there are 5,000 updating events per player and game on average in each treatment. In **Fig.3.2d-e**, we display which strategies cooperative players tend to use. To this end, we use the data of the simulations in **Fig.3.2a-c**. We define a player's strategy to be cooperative if the player's cooperation rate in each game is at least 2/3 against the given co-player (other cut-offs give similar results). The panel then shows the arithmetic mean of all strategies classified as cooperative. Finally,**Fig.3.2g,h** shows the impact of the benefit of cooperation in the high game. Here, each point corresponds to the time average of one long simulation (10^6 time steps). We explore the impact of other model parameters in **Fig. S3**.

In **Fig.3.3**, we explore four model extensions that describe the impact of different cognitive constraints and heuristics. The top row in **Fig.3.3** describes how the evolving cooperation rates are affected as we change (A) the probability $\varepsilon_{\mathrm{IR}}$ of experiencing imperfect recall, (B) the probability $\varepsilon_{\mathrm{SP}}$ of experiencing a spillover, (C) the weight κ that measures the strength of preferential updating, and (D) the likelihood λ that a player engages in narrow bracketing. Each data point is the average of a simulation run for 10^6 updating steps. The middle and the bottom row of the figure show the player's average strategies. This figure is based on all strategies used during the simulation (not only the cooperative strategies). For a more detailed discussion of each model extension and the respective results, see **Supplementary Information**.

3.4.4 Experimental methods

For our experiment, we recruited 316 participants (161 females, mean age: 21) from the University of Exeter student pool FEELE (Finance and Economics Experimental Laboratory at Exeter). The experiment was implemented in oTree [194]. Sessions were programmed for one of the three treatments and players only participated in one session. Participants were matched in groups of four all playing the same treatment. All participants were anonymous and only referred to by numbers from 1 to 4. Each treatment lasted for a minimum of 20 rounds of the repeated game(s). After the 20th round, each subsequent round had a 50% chance of occurring, to avoid end-game effects. Participants received £3 for participating and could earn a bonus payment based on their decisions in the game. The points earned during the game were converted at a rate of 20 points = £0.26. The average bonus payment across all treatments was £1.39. The experiment was approved by the Ethics Committee of the Medical Faculty of Kiel University (D 571/20). In the same-partner and different-partners treatment, participants make their decision for both repeated games simultaneously, round by round. In the baseline treatment, they only take part in one repeated game. Payoffs in each repeated game are based on the payoff matrix of the donation game, with $b = 4$ points and $c = 2$ points in the high game, and $b = 3$ points and $c = 2$ points in the low game. These values were chosen based on preliminary simulations and pilots to avoid ceiling effects and to obtain the largest difference in cooperation rates between the two games in the control treatment.

We analysed the data using two-tailed non-parametric tests as well as statistical regressions, using interacting pairs as statistical units, with the exception of the different-partner treatment where groups of four interacting participants are used (due to the nature of the design, these groups cannot be separated into pairs). This gives us 36 groups of 2 for the same-partner treatment, 32 groups of 4 for the different partners treatment, and 29 groups of 4 for the baseline control. The sample size was estimated from past research [182]. Four participants dropped out during the repeated game, two in the same-partner treatment and two in the control treatment. We calculated the average value for each pair/grouping of players and we then compared this average value between treatment with a Mann-Whitney U-test, or within each treatment with a Wilcoxon signed-rank test. We report the outcome uncorrected for multiple testing. We only use the first 20 rounds of each game for our analysis and only take into account groups without a drop-out. For more details on the experimental setup, see **Supplementary Information**.

Chapter 4

Cumulative Reciprocity

This chapter is published under the title: Li, J., Zhao, X., Li, B., Rossetti, C. S. L., Hilbe, C., Xia, H. (2022). Evolution of cooperation through cumulative reciprocity, *Nature Computational Science*, 2(10), 677–686. https://doi.org/10.1038/s43588-022-00334-w [132]. It is supported by a supplementary information in the appendices, with the details of the theoretical analysis and empirical methods and analysis.

Contents

Abstract

Reciprocity is a simple principle for cooperation that explains many of the patterns of how humans seek and receive help from each other. To capture reciprocity, traditional models often assume that individuals use simple strategies with restricted memory. These memory-one strategies are mathematically convenient, but they miss important aspects of human reciprocity, where defections can have lasting effects. Here, we instead propose a strategy of cumulative reciprocity. Cumulative reciprocators count the imbalance of cooperation across their previous interactions with their opponent. They cooperate as long as this imbalance is sufficiently small. With analytical and computational methods, we show that this strategy can sustain cooperation in the presence of errors, it enforces fair outcomes, and it evolves in hostile environments. With an economic experiment, we confirm that cumulative reciprocity is more predictive of human behavior than several classical strategies. The basic principle of cumulative reciprocity is versatile and can be extended to a range of social dilemmas.

4.1 Introduction

Evolutionary game theory provides a formal framework to study the evolution of cooperation, which is a far-reaching problem that has attracted great attention [14, 16, 50]. The simplest and most widely used model to study this problem is the prisoner's dilemma [26, 50]. In the prisoner's dilemma, two individuals independently decide whether to cooperate. Mutual cooperation is optimal for the pair, yet each individual is tempted to defect. Although the basic premise of the game is simple, it approximates the logic of many cooperative interactions in biological, societal and artificial worlds, including friends who exchange favors [195], animals who exchange food or other services [196], or nations that coordinate their policies [197]. When there is only a single round of the prisoner's dilemma, defection is the only Nash equilibrium. However, if individuals interact repeatedly, they can sustain cooperation with conditionally cooperative strategies [162]. The respective mechanism of cooperation is called direct reciprocity [14, 16].

The key to sustain cooperation in a repeated prisoner's dilemma is to act like a 'partner' [198]. As long as the opponent is cooperative, a partner should go along. However, once an opponent defects, a partner strategy needs to make sure that the opponent cannot gain a lasting advantage. Examples of such partner strategies comprise tit-for-tat (TFT) [50], generous tit-for-tat (GTFT) [145, 165], or win-stay-lose-shift (WSLS) [99, 199], among many others [96–98, 107, 117, 117, 151, 200]. Although each of the above strategies can succeed in certain environments [50, 99, 145], they also have well-known weaknesses. For example, TFT is unable to sustain cooperation in the presence of errors [201] 22. GTFT typically fails to evolve when individuals have access to a richer strategy space [48]. Finally, WSLS is stable only when the benefit of mutual cooperation is sufficiently large [99]. The problem of identifying successful strategies of direct reciprocity becomes even more complex when interactions take place among more than two individuals [94, 96], or when the benefit of cooperation can change in time

[90, 202].

Traditionally, much of the existing work on direct reciprocity is confined to individuals with restricted memory. The most common assumption is that all individuals have one-round memory [60], but researchers have also identified several promising strategies that take into account the last two or three rounds [97, 98, 203]. While some of these strategies have remarkable robustness properties (especially when errors can be assumed to be vanishingly rare), strategies with a pre-defined memory length often fail to capture certain important aspects of human behavior. For example, individuals might often find it easier to forgive a defecting opponent if this opponent is generally cooperative. To encode these more nuanced responses, individuals need to resort to an opponent's cumulative behavior, during the entire course of their previous interactions.

Perhaps the most natural way to introduce such behaviors is to let individuals count how often each of them has defected so far. To formalize this idea, consider a repeated prisoner's dilemma between two individuals named 'Alice' and 'Bob'. Suppose during their first k-1 interactions, Alice defected in n_A rounds, whereas Bob defected in n_B of the rounds. Alice may choose to cooperate in round k, unless Bob defected substantially more often in the past than she did. In other words, Alice would cooperate unless $n_B - n_A > \Delta_A$, where $\Delta_A \geq 0$ can be interpreted as Alice's tolerance level. Herein, we refer to this kind of strategy as 'cumulative reciprocity', or more briefly, as CURE.

While cumulative reciprocity is straightforward to define, analytical results are more difficult to obtain, compared with the case of memory-one strategies. Nevertheless, such results are feasible. First, we show that CURE is indeed a partner strategy in the absence of implementation errors. Second, similar to previous work on zero-determinant strategies [111, 112, 177, 190, 204, 205], individuals can use CURE to enforce fairness: if one player is a cumulative reciprocator, both players are guaranteed to get the same payoff, independent of the opponent's strategy. Third, even in the presence of (rare) errors, the payoff of CURE against itself is approximately optimal. At the same time, unconditional defectors cannot invade. Further simulations suggest that individuals are most likely to adopt CURE when most well-known memory-one strategies fail. We further support these theoretical findings with a behavioral experiment. According to this experiment, cumulative reciprocity is better able to explain human behavior than many classical strategies like TFT [50] , WSLS [99, 199], or previously proposed memory-k strategies [97, 98, 203].

Overall, our findings suggest that when cooperation is particularly costly, simple strategies based on an opponent's last behavior do not suffice. In such environments, it takes a more cumulative assessment of the players' past actions to sustain cooperation. Here, we combine various mathematical, computational, and experimental methods to facilitate the analysis of cumulative reciprocity. Although many of our analyses focus on the classical prisoner's dilemma, the main

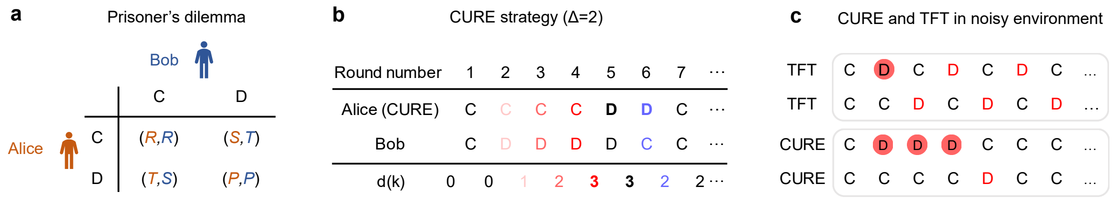


Figure 4.1: CURE in the repeated prisoner's dilemma. **a**, The payoff matrix of a prisoner's dilemma game. For simulations, we use $R = 3, S = 0, T = 5$ and $P = 1$, unless specified otherwise. **b**, CURE defects when the defection difference statistic d(k) exceeds the tolerance threshold (in this case, Δ = 2). Otherwise, CURE keeps cooperating. **c**, When two TFT players meet and one player makes an error, cooperation breaks down. When two CURE players meet, cooperation is robust even when one player mistakenly defects in three of the rounds (Δ = 2). The blue circle indicates the occurrence of an error, and the red letter indicates the defection caused by the error.

principles of CURE extend to multiplayer interactions, as well as to stochastic games in which payoffs fluctuate in time. In all these applications, cumulative reciprocity proves to be a simple mechanism to sustain fairness and cooperation.

4.2 Results

4.2.1 The repeated prisoner's dilemma with cumulative reciprocity

We first consider pairwise interactions between only two individuals, as shown in Fig.4.1a. We refer to the players as Alice and Bob. In each round, both Alice and Bob independently decide whether they want to cooperate or defect. If they both cooperate (denoted by CC), they get the reward R. If they defect (DD), they both receive the punishment payoff P. Finally, if one player cooperates and the other defects (CD or DC), the cooperator gets the sucker's payoff S whereas the defector gets the temptation payoff T. Payoffs satisfy the characteristic conditions of the prisoner's dilemma, $S < P < R < T$ and $2R > T + S$. That is, mutual cooperation is the best outcome for the pair, yet each player individually prefers to defect. For our numerical simulations, we use the payoffs of Axelrod2 ($R = 3, S = 0, T = 5, P = 1$), unless stated otherwise.

The players' actions may be subject to 'trembling hand' errors [201, 206]. That is, a player who intends to cooperate may instead defect with some probability $0 \leq \epsilon < 1/2$ (similarly, a player who wishes to defect may cooperate with the same probability). For analytical results, we suppose the game is infinitely repeated; after each round, there is another round. We complement these analytical results with simulations for long but finitely repeated games. The overall payoffs of Alice and Bob are defined as their expected payoffs per round. For details, see Methods.

Traditionally, much of the work on reciprocity assumes that players either have finite recall [97, 170, 173, 207, 208] 20,36-39, or that their decisions can be encoded with finite state automata [100, 209]. Instead, here we propose a strategy of cumulative reciprocity. We formalize this strategy by introducing two counter variables, $n_A(k)$ and $n_B(k)$. These variables record how often Alice and Bob have defected before round k. Let $d(k)$ denote the difference, $d(k) = n_B(k) - n_A(k)$. We refer to $d(k)$ as the defection difference statistic. We say Alice adopts the strategy CURE if in any given round k she cooperates if and only if this defection difference statistic is below a pre-defined threshold $d(k) \leq \Delta_A$. We interpret $\Delta_A \geq 0$ as Alice's tolerance level. If it is zero, Alice demands that Bob is at least as cooperative as Alice. For larger values of Δ_A, Alice becomes increasingly more lenient. Fig.4.1b depicts the basic logic of CURE.

In contrast to memory-one strategies, a cumulative reciprocator takes the entire history of the game into account. As a result, the mathematics becomes more intricate. Whereas games between two memory-one players can be represented as a Markov chain with four possible states (the possible outcomes of any given round), the state variable $d(k)$ of a cumulative reciprocator can assume arbitrary integer values. Perhaps somewhat surprisingly, it is still possible to derive analytical results. To this end, we represent the dynamics between a CURE player and its opponent by an infinitely dimensional linear system. In many important cases, this system can be solved. We summarize our results in the following. All details and proofs are in Section 1 of the Supplementary Information.

4.2.2 Payoffs against selected strategies

To gain some first insights into the performance of cumulative reciprocity, we first study games between two CURE players. In case they both use the same tolerance level $\Delta \geq 1$, each player's average cooperation rate ρ_{CURE} becomes

$$\rho_{CURE} = 1 - \frac{(2 - 3\epsilon + 2\Delta(1 - 2\epsilon))}{(1 - 2\epsilon^2 + 2\Delta(1 - 2\epsilon))}\epsilon. \tag{4.1}$$

In particular, the cooperation rate approaches one as errors become rare, $\epsilon \to 0$. Therefore, although CURE follows a similar basic principle as TFT, it is much more robust with respect to noise [60] (Fig.4.1c). Overall, the resulting average payoff $\pi(CURE, CURE)$ is

$$\begin{aligned} \pi(CURE, CURE) &= \frac{(1-\epsilon)^2 \times (1-2(1-\epsilon)\epsilon+2\Delta(1-2\epsilon))}{1-2\epsilon^2+2\Delta(1-2\epsilon)} \times R \\ &+ \frac{2\epsilon(1-\epsilon) \times ((1-\epsilon)^2+\Delta(1-2\epsilon))}{1-2\epsilon^2+2\Delta(1-2\epsilon)} \times (T+S) \\ &+ \frac{\epsilon^2 \times (3-2(3-\epsilon)\epsilon+2\Delta(1-2\epsilon))}{1-2\epsilon^2+2\Delta(1-2\epsilon)} \times P. \end{aligned} \tag{4.2}$$

As one may expect, this payoff is monotonically decreasing in the error rate, but increasing in the strategy's tolerance level: the more tolerant the two players are, the better they can cope with each other's unintentional errors. Analogous formulas can also be derived if the two cumulative reciprocators use different thresholds Δ_A and Δ_B.

Of course, a strategy's payoff against itself is only one possible measure of a strategy's ability to sustain cooperation. In a next step, we consider interactions between CURE and other well-known strategies. We start by matching a cumulative reciprocator with an unconditional defector (ALLD). We find that in that case, both players' cooperation rates assume the theoretical minimum ϵ (which is also how often ALLD cooperates against itself). The result is intuitive: if Alice adopts CURE and Bob adopts ALLD, Alice is only expected to cooperate in the first Δ_A

rounds. From then on, both players effectively implement an ALLD strategy. At that point, either of them only cooperates in case of an error. Because the first Δ_A rounds are negligible in infinitely repeated games, both players obtain the same payoff. As a result, we find that CURE weakly dominates ALLD for all error rates and all tolerance levels (see Supplementary Information Section 1). Cumulative reciprocators can therefore cope with unconditional defectors even better than TFT (TFT never dominates ALLD in the presence of errors [201]). Similarly, we can also consider games between cumulative reciprocators and unconditional cooperators (ALLC). Here, cooperation rates are maximal, $1 - \epsilon$, which is also the cooperation rate of ALLC against itself. In the absence of errors, $\epsilon \to 0$, the payoff of two CURE player thus matches the payoff of ALLC against CURE. For positive error rates, unconditional cooperators have a slight payoff advantage when they interact with a CURE opponent, compared to the interaction between two CURE players. We discuss the consequences of this advantage in more detail in the next sections.

In addition to ALLC and ALLD, we use a similar approach to derive the payoff of CURE against arbitrary memory-one opponents. We no longer solve the respective infinitely dimensional system explicitly, but use the equations to approximate payoffs numerically (Section 1 of Supplementary Information and Supplementary Tables 1–6). To validate these results, we implement independent computer simulations to estimate the players' payoffs and cooperation rates. To this end, we consider a CURE player and an opponent who adopts one of nine selected strategies. The simulation results match our analytical calculations (Section 2 of Supplementary Information and Supplementary Tables 1–8). The results also suggest that although CURE is generally cooperative, it does not cooperate with any other cooperative strategy in the presence of errors. As an example, we show that when CURE interacts with WSLS, all four game outcomes (CC, CD, DC, DD) occur equally often over time (Supplementary Fig.1).

4.2.3 Fairness and stability of cumulative reciprocity

The previous analysis implies that when cumulative reciprocators either interact with ALLC or ALLD, both players obtain the same payoff. This holds more generally. We can prove that for arbitrary 2×2 games, CURE always enforces an equal outcome. More precisely, irrespective of the co-player's strategy σ, a cumulative reciprocator always reacts in such a way that eventually

$$\pi(CURE, \sigma) = \pi(\sigma, CURE). \tag{4.3}$$

Equation (4.3) holds irrespective of the precise error rate and CURE's tolerance level. Using the terminology of Duersch et al [164, 210], we conclude that cumulative reciprocity is unbeatable: no opponent is able to gain a lasting advantage in a direct interaction with a cumulative reciprocator. CURE shares the property of enforcing fairness with another classical strategy for the prisoner's dilemma, TFT (for which an equation like Eq. (4.3) has been first derived by Press and Dyson [112]). This similarity between CURE and TFT is not a coincidence. As an illustration, consider a game without errors, and suppose Alice adopts CURE whereas Bob uses an arbitrary strategy. Then Alice cooperates until the defection difference statistic $d(k)$ hits her tolerance level. At that point, Alice effectively implements a TFT strategy; she cooperates as long as Bob does, and she switches to defection once Bob defects. If Bob then again resumes to cooperate, so does Alice.

By combining the previous results, we show that in the absence of errors, CURE forms a Nash equilibrium in the repeated prisoner's dilemma. To see this, we note that for $\epsilon = 0$, the payoff of CURE against itself simplifies to the mutual cooperation payoff R. If a single deviating player could achieve a larger payoff than R, Eq. (4.3) would imply that also the remaining CURE player obtains a payoff larger than R. However, because $2R$ is the maximum payoff that the two players can possibly achieve, this yields a contradiction (in the Supplementary Information, we slightly strengthen this result, by showing that CURE is in fact a subgame perfect equilibrium [25]).

For positive error rates, the above argument is no longer true. Here, players can gain a payoff advantage by deviating to ALLC. However, the respective payoff advantage is often negligible. In particular, for sufficiently small error rates CURE remains an approximate Nash equilibrium (see Section 1 of Supplementary Information). This means that the payoff advantage from deviating to any other strategy is bounded from above, and it vanishes completely as errors become rare.

4.2.4 CURE and population dynamics

For the previous results, we considered games among players with fixed strategies. This kind of analysis is useful to explore a strategy's basic properties. However, it does not take into account whether players have an incentive to adopt their respective strategies in the first place. To explore this latter question, we now consider a large population of players and let their strategies evolve (see Methods for the precise setup of our evolutionary simulations).

We first examine whether cumulative reciprocity has an evolutionary advantage when populations only contain two strategies. We compare CURE with a tolerance level of $\Delta = 1$ to nine well-known strategies for the repeated prisoner's dilemma (Fig.4.2, Supplementary Figs.2-7 and Supplementary Information Section 3; for exact descriptions of the nine strategies, see Supplementary Information Section 2). In the simulations in Fig.4.2, CURE is initially adopted by 0.1% of the population (accordingly, we speak of the other strategy as the 'resident'). The results show that cumulative reciprocity invades six of the nine considered resident populations (Fig.4.2a-f). Against WSLS, we observe CURE is risk-dominant (Fig.4.2g) [62]: the critical frequency of cumulative reciprocators required to invade is below 50%. Only in resident populations that tend to cooperate unconditionally, CURE does not evolve (Fig.4.2h,i). Here, CURE suffers from its slight payoff disadvantage discussed earlier. However, once we additionally include defectors into the population, CURE becomes again essential. In that case, we observe that ALLC, ALLD, and CURE can stably coexist (Supplementary Figs.8,9; see Section 4 of the Supplementary Information for a detailed analysis).

To explore the evolutionary performance of CURE when many strategies compete, we have

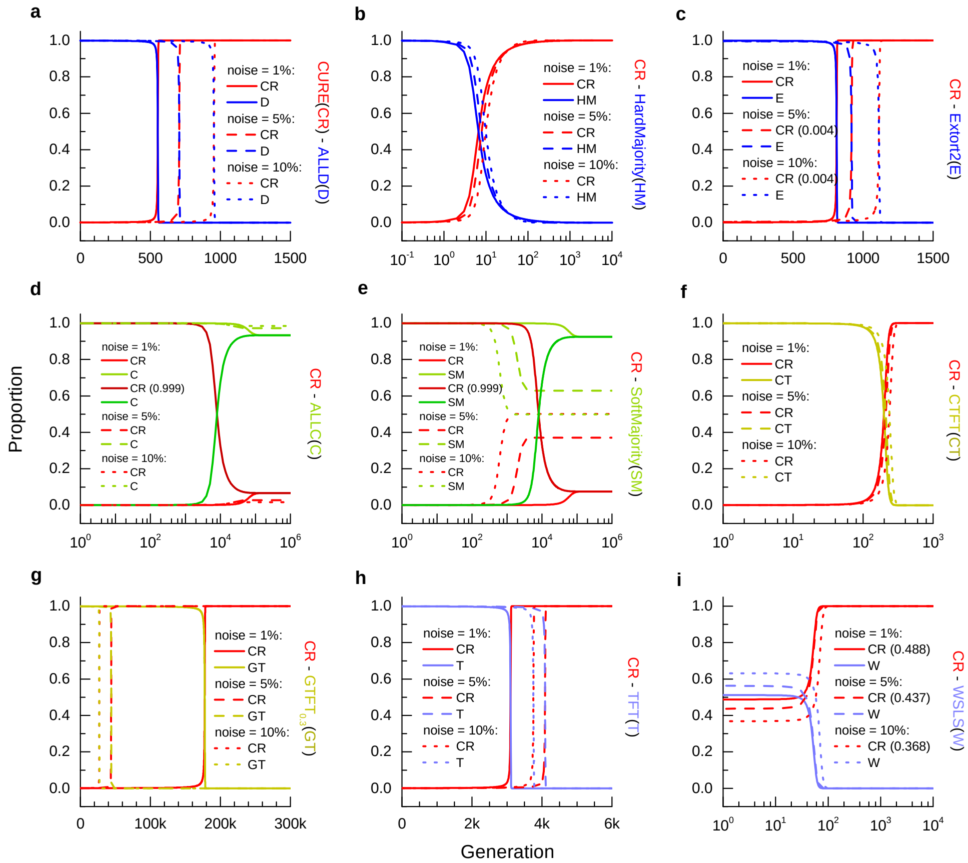


Figure 4.2: CURE in the repeated prisoner's dilemma. **a**, The payoff matrix of a prisoner's dilemma game. For simulations, we use $R = 3, S = 0, T = 5$ and $P = 1$, unless specified otherwise. **b**, CURE defects when the defection difference statistic d(k) exceeds the tolerance threshold (in this case, $\Delta = 2$). Otherwise, CURE keeps cooperating. **c**, When two TFT players meet and one player makes an error, cooperation breaks down. When two CURE players meet, cooperation is robust even when one player mistakenly defects in three of the rounds ($\Delta = 2$). The blue circle indicates the occurrence of an error, and the red letter indicates the defection caused by the error.

run additional simulations for arbitrary memory-one strategies. Each memory-one strategy is represented by a four-dimensional vector, (p_1, p_2, p_3, p_4). Here, p_1, p_2, p_3, p_4 refer to the player's probability to cooperate given that the outcome of the previous round is CC, CD, DC, DD, respectively. For each entry of the vector, we consider eleven possible values, equally distributed between 0.01 and 0.99 (corresponding to a noise rate of 1%). Overall, we thus allow for $114 = 14,641$ memory-one strategies, to which we add a single strategy of cumulative reciprocity. Initially, all strategies are equally abundant. We use the same parameters and simulation techniques as before (see Methods).

We examine the evolutionary dynamics under different scenarios, as shown in Fig.4.3. When individuals can only choose among memory-one strategies (not CURE), evolution eventually leads to a coexistence of different GTFT-like strategies (Fig.4.3a). The respective strategies are of the form $(0.99, g_1, 0.99, g_2)$, with $g_1, g_2 \geq 0.1$. That is, all of them aim to reciprocate a co-player's cooperation, but they would occasionally also cooperate against defectors. The dynamics change completely if CURE is added. After an initial transitionary period, we observe that a vast

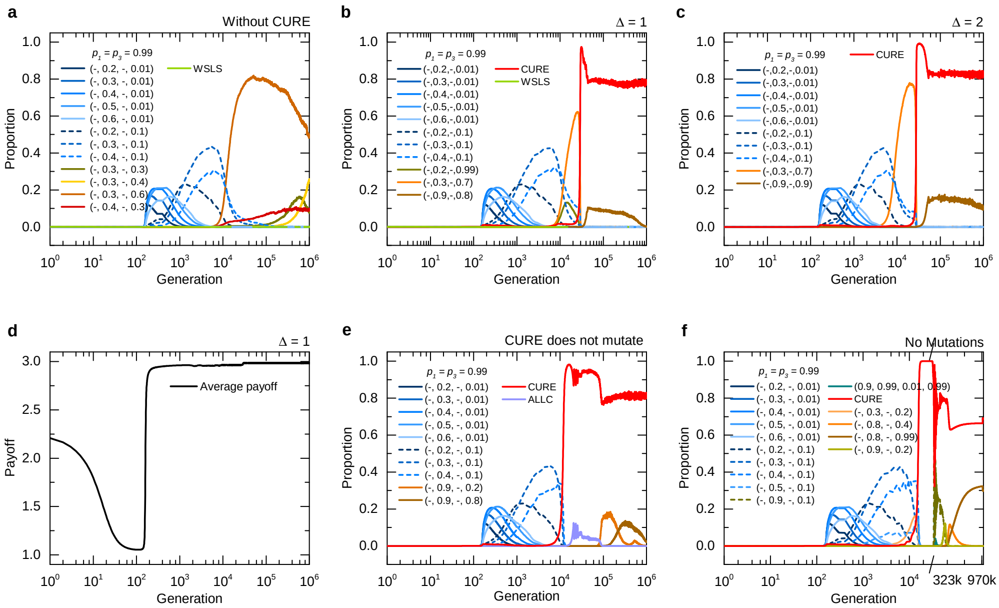


Figure 4.3: Evolution of CURE in populations of memory-1 players. We consider the evolutionary dynamics when individuals can choose among a general set of memory-1 strategies and CURE. Initially, all available strategies are equally abundant in the population. As the simulation proceeds, we first often observe the emergence of TFT-like strategies. The long-run dynamics depends on the scenario considered. **a**, Without CURE, players eventually tend to adopt GTFT- like strategies. These players always reciprocate a co-player's cooperation, but they occasionally also cooperate if the co-player defected. **b-d**, Once CURE is available, it becomes predominant, irrespective of the tolerance threshold Δ values of 1 **(b,d)** or 2 (**c**), with overall payoffs remaining stable and close to the theoretical maximum (**d**). **e,f**, While the previous results allow for (rare) mutations, we obtain similar results if there are either no mutations at all (**e**) or if players are unable to learn CURE by mutation (**f**). For clarity, we only depict strategies that reach a frequency of at least 0.1 at some point during the process. We represent the evolving memory-1 strategies as vectors (p_1, p_2, p_3, p_4). The entries correspond to the player's cooperation probability after CC, CD, DC and DD, respectively.

majority of players engages in cumulative reciprocity, irrespective of CURE's tolerance level or of the exact mutation scheme used (Fig.4.3b-f). Although both GTFT and CURE are overwhelmingly cooperative, CURE seems to be more robust with respect to subsequent invasions, and it leads to slightly larger average population payoffs.

In a next step, we explore how the dynamics depend on the exact payoffs. To this end, we vary the reward for mutual cooperation R between $(T+S)/2$ and T. We identify three different regimes (Fig.4.4 and Supplementary Figs.10-17; see Sections 5 and 6 of Supplementary Information for a detailed description). (i) When mutual cooperation yields substantial rewards ($R \geq 3.675$), WSLS dominates the population. This result is largely in line with earlier work. In the presence of errors, WSLS becomes evolutionarily stable once R is sufficiently large20. In that case, it also readily evolves among memory-one players13. (ii) For a small window of intermediate rewards, $3.625 \leq R < 3.675$, GTFT-like strategies are predominant. Among those strategies, most of them have a generosity of 0.6 (i.e., after a co-player's defection, they cooperate with around 60% probability, see Inset of Fig.4.4). (iii) When mutual cooperation only generates a comparably modest reward ($R < 3.625$), CURE is most abundant. CURE is thus particularly likely to succeed in those environments that are traditionally considered as hostile to the evolution of cooperation.

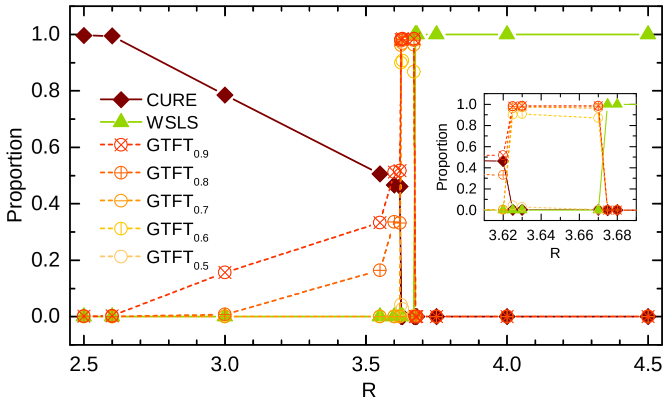


Figure 4.4: CURE, GTFT and WSLS dominate the population in different payoff regions. Our previous results are based on the classical payoffs used by Axelrod (ref 2). Here, we use the same basic setup as in Fig.4.3 to explore which strategies are most successful as we vary the reward R for mutual cooperation. For each value of R, we record the final frequency of CURE, WSLS and the class of GTFT-like strategies of the form (0.99, p_2, 0.99, p_4). We denote by $GTFT_x$ the setof all such strategies for which p2 and p4 are at most x. These sets are nested, $GTFT_{0.5} \subset GTFT_{0.6} \subset ... \subset GTFT_{0.9}$. The graph suggests that there are two transition points, $R1 = 3.625$ and $R2 = 3.675$. When $R < R_1^*$, CURE dominates the population. When $R_1^* \leq R < R_2^*$, GTFT-like strategies are predominant. When $R \geq R_2^*$, WSLS is most abundant. For details and depictions of some of the evolutionary trajectories, see Supplementary Information Section 6.

These evolutionary results are remarkably robust with respect to various model extensions. For example, in Sections 7-9 of the Supplementary Information, we additionally discuss the evolutionary dynamics when different variants of CURE compete, or when CURE competes against a selection of memory-2 and memory-3 strategies, or competitions between a discounted version of CURE and selected memory-one strategies. Moreover, in Fig. 5, we illustrate how the concept of cumulative reciprocity extends to stochastic games25,26 in which payoffs can fluctuate in time, and to repeated games that involve more than two players [173, 208] 38,39 (see also Section 10 of the Supplementary Information).

4.2.5 CURE and human play

The above results highlight the strong theoretical properties of CURE. In a next step, we explore the relevance of cumulative reciprocity for human decision making, by implementing a simple behavioral experiment with two treatments (see Methods). In the treatment without errors, the decisions of the human participants are executed perfectly. In the treatment with errors, the decisions are mis-implemented with a 10% probability. The basic results of this experiment are as one may expect [128, 131]. The game without errors yields more cooperation, and in both treatments, cooperation rates slightly decrease in time (Fig.4.6a,b).

To explore to which extent human decisions are accurately predicted by cumulative reciprocity, we consider one particular instance of CURE (with threshold $\Delta = 3$; other thresholds yield similar results). For comparison, we consider the four pure memory-one strategies that can sustain cooperation in the absence of errors [107], namely GRIM [190], TFT [50], firm-but-fair

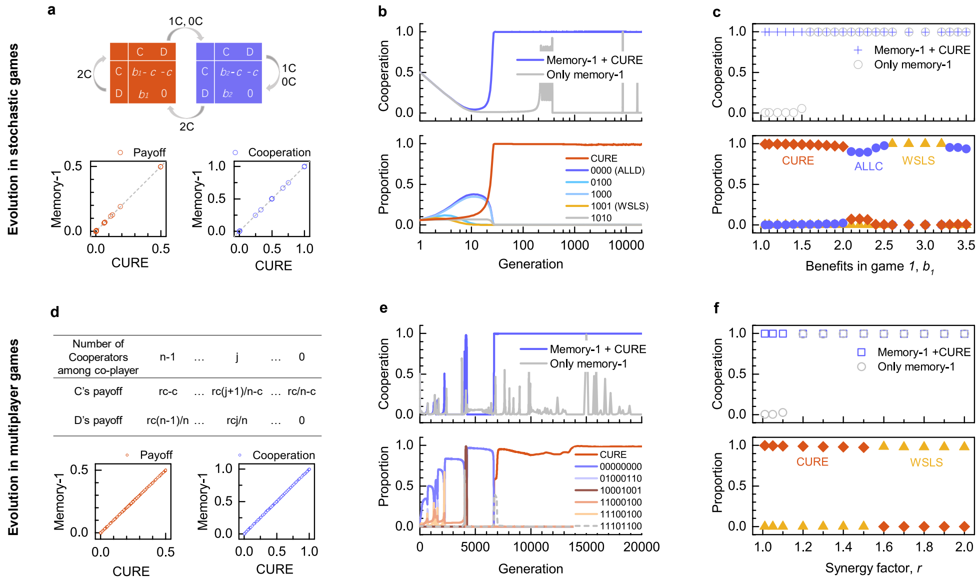


Figure 4.5: CURE in stochastic games and the repeated public goods game. In addition to the repeated prisoner's dilemma, we also explored examples in which individuals interact in a stochastic game [90, 202] or in a public goods game [173, 208]. **a**, As an example of a stochastic game, we suppose that players can be in one of two possible states. In both states, they interact in a donation game. The benefit of cooperation b_s depends on the current state s, with cooperation in state 1 being more valuable, $b_1 > b_2$. Players only find themselves in state 1 if they both cooperated (i.e. 2C) in the previous round. Also in this stochastic game, CURE is able to enforce equal payoffs, independent of the co-player's strategy. **b,c**, When the benefit of cooperation in the first game is comparably small, cooperation cannot evolve among memory-1 players. It evolves readily when CURE is added to the population. **d**, As an example of a multiplayer game, we explore a public goods game with four players. Again, CURE enforces fairness. Irrespective of the strategies of the other group members, a cumulative reciprocator gets the same payoff as the other group members obtain on average. **e,f**, As before, CURE is particularly strong when cooperation is difficult to achieve otherwise (for small multiplication factors r). The parameter values in the stochastic game (unless explicitly varied) are b1 = 1.5 and b2 = 1.01 with a threshold of $\Delta = 1$. In the public goods game, we use $n = 4, r = 1.5, c = 1, \Delta = 0.5$.In both cases, the noise is ϵ = 0.001 and strategies are subject to mutations for the entire simulation. See Supplementary Information Section 10 for details on how these games were implemented and our respective implementation of CURE.

[60], and WSLS [99, 199]. In addition, we also include three memory-2 and memory-3 strategies that have been highlighted recently, AON220, TFT-ATFT [203], and CAPRI [98]. For every participant in the experiment, we compare the participant's actual decisions with the decisions the participant would have made when using any of these eight strategies. For the treatment without errors, we find that all eight strategies predict human behavior equally well. However, these results change in the treatment with errors, in which even cooperative participants defect occasionally. Here, only CURE correctly predicts the behavior of a substantial number of participants across all rounds (Fig.4.6c; Section 11 of the Supplementary Information). Compared to traditionally considered rules of reciprocity, cumulative reciprocity thus seems to be a more sensible guiding principle for cooperation.

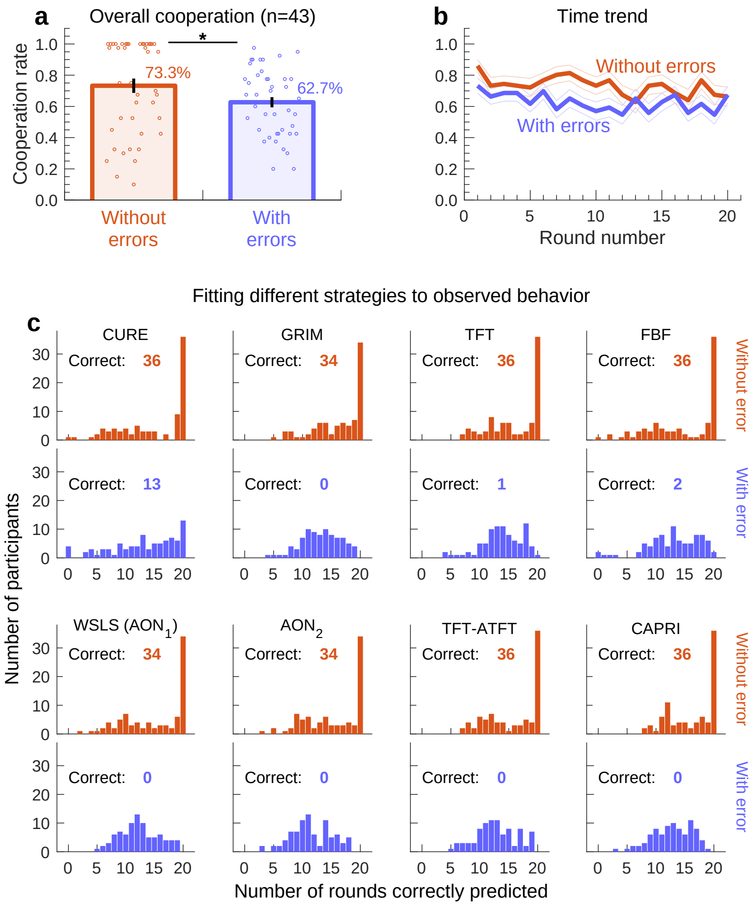


Figure 4.6: CURE in an economic experiment. To explore the relevance of CURE for human decision-making, we implemented an economic experiment based on the repeated prisoner's dilemma (see Methods for details). The experiment consists of two treatments: one treatment without errors, and one treatment with a 10% error rate. For each treatment, we report data from $n = 43$ pairs of participants. Here, we show the results with respect to the players' implemented decisions. **a**, As one may expect, there is more cooperation in the absence of errors (73.3% versus 62.7%, Mann'Whitney U test, $P = 0.018$). **b**, In both treatments, cooperation rates are slightly decreasing in time. **c**, To capture the participants' behaviour, we explore how many of their 20 decisions are explained by one of eight possible strategies. Without errors, all the considered strategies predict the participants' behaviour equally well. With errors, only CURE correctly predicts the behaviour of a substantial proportion of the subjects for all 20 rounds (this difference is significant, see Supplementary Information Section 11 for details). In **a**, dots represent the outcome of individual groups, error bars represent s.e.m. and the asterisk represents significance at $P < 0.05$. In **b**, the thick line represents averages, whereas the thin lines again represent s.e.m. All statistical tests are two-tailed and non-parametric. We do not adjust for multiple comparisons.

4.3 Discussion

Reciprocal cooperation requires that individuals are prepared to help others; yet they also need to be ready to fight back when their help is exploited2. Most of the previously known strategies with these properties, including TFT [50] or WSLS [99, 199], only react to the very last round, while ignoring the entire previous history of interactions. In contrast, friends are often comfortable with temporary imbalances, as long as they are on equal terms on average [211]. To capture reciprocal behaviors that unfold on such a longer timescale, we have introduced a strategy of cumulative reciprocity. Individuals with that strategy keep a mental record of how often each party has defected in the past. They cooperate as long as this record is sufficiently balanced, and they defect otherwise.

The strategy of cumulative reciprocity has a number of remarkable properties: its payoff is robust with respect to errors, it enforces fairness, and it is a Nash equilibrium in the absence of errors. Yet it does not satisfy the notion of evolutionary stability [54]. Evolutionary stability is generally difficult to achieve in repeated games [172, 212, 213]. However, in large and heterogeneous populations, the notion of evolutionary stability seems overall less important. In such populations, the success of a strategy may be rather determined by how well it performs against a wide variety of strategies. Our simulations suggest that CURE fares particularly well in that regard. These positive results are largely independent of how often errors occur, or how frequent mutations are. In this way, CURE may revive a more general discussion on the effect of memory on the stability of reciprocity. As Press and Dyson [112] pointed out, longer memory does not give a player an immediate advantage against shorter-memory opponents. Although their assertion holds for pairwise interactions, it is no longer true when players need to find strategies that respond well to many different co-players [49]. In such mixed populations, cumulative reciprocity can be remarkably successful.

The analytical properties of cumulative reciprocity are perhaps less straightforward to derive than the properties of more conventional strategies with finite recall. For example, the dynamics among two memory-one players can be captured by solving a comparably simple formula [112]. In contrast, a mathematical description of the average cooperation rate of CURE leads to an infinite system of linear equations (see Section 1 of Supplementary Information). It is probably for this reason that related previous work either completely relied on simulations [167], or focused on a simpler case in which players always execute their actions without errors [214]. With our study, we offer a mathematical framework to analyze cumulative reciprocity in more general environments. Our results are applicable to classical two-player repeated games like the prisoner's dilemma, but they also apply to multiplayer games, or games in which the feasible payoffs vary in time. Future work could explore, for example, how cumulative reciprocity extends to games with continuous degrees of cooperation, or it could identify alternative strategies based on cumulative assessments of the players' past history.

Even though the mathematical analysis of CURE may be elaborate, the strategy itself is straightforward to implement. Cumulative reciprocators do not need to memorize the exact outcome of the last k rounds [97], let alone the precise history of all past interactions. Instead, players only keep track of a single variable, the defection difference statistic. This statistic is easily updated, and it has an intuitive interpretation: It simply counts the imbalance in the players' past cooperative actions. A comparable bookkeeping strategy seems to be at work when people keep a loose record of obtained favors that need to be repaid. Such bookkeeping strategies may be rare among close friends, but they can play an important role in the early stages of a social relationship [68]. Especially when no mutual trust has been established, a mental record of the overall cooperation balance can be important to avoid exploitation. At the same time, it allows individuals to forgive peers who only defected by mistake. With these advantages, cumulative reciprocity serves as an effective mechanism to maintain fairness and mutual cooperation.

4.4 Methods

The theoretical aspects of our study raise no ethical concerns. For our behavioral experiments, we have obtained IRB approval by the Ethics Committee of the Medical Faculty of the University of Kiel (D 613/21, October 29, 2021).

4.4.1 Simulation-based estimation of payoffs and cooperation rates

In addition to our analytical results and numerical approximations (Supplementary Tables 1-6), we employ simulations to estimate the players' payoffs and cooperation rates (Supplementary Tables 7,8). For these simulations, we conducted 1,000 independent computer experiments between any given pair of players (repeated prisoner's dilemma games) and any given group of four players (repeated public goods games). In each experiment, players interact in the game for 10,000,000 rounds. We calculate the average payoffs of the players across all rounds and we count how often each player cooperates. The payoffs and cooperation rates are then averaged over the 1,000 experiments (over 100 experiments in the case of stochastic games). We have run these simulations for different parameter combinations. Unless noted otherwise, we implement CURE with a tolerance threshold of $\Delta = 1$. For the payoffs of the repeated prisoner's dilemma, we take the values used by Axelrod2 as a default (i.e., $T = 5, R = 3, P = 1$, and $S = 0$). Furthermore, we considered three different noise rates – low (1%), medium (5%), and high (10%). The parameters of the stochastic games and the multiplayer games are illustrated in Fig.4.5; for the exact setup of the simulations used for this model extension, see Supplementary Information Section 10.

4.4.2 Simulating the frequency dynamics of strategies

To explore the evolution of strategy frequencies in populations of players, we consider several different scenarios: (i) Pairwise competitions between CURE and one other strategy (Fig.4.2, Supplementary Figs.2-7). The other strategy is either ALLC, ALLD, TFT, GTFT, CTFT, WSLS, an extortionate strategy, SoftMajority, or HardMajority. (ii) A 3-strategy competition between

ALLC, ALLD, and CURE (Supplementary Fig.8,9). (iii) Simulations of heterogeneous populations consisting of memory-one strategies and CURE in repeated games (Figs.3,4, Supplementary Figs.10-17). (iv) Simulations of heterogeneous populations consisting of memory-one strategies and CURE in stochastic games (Fig.4.5b,c). (v) Simulations of heterogeneous populations consisting of memory-one strategies and CURE in multiplayer games (Fig.4.5e,f). (vi) Simulations in which different variants of CURE (with different tolerance levels) compete (Supplementary Fig. 18) (vii) Simulations in which CURE competes with selected memory-2 and memory-3 strategies (Supplementary Fig. 19).

Each simulation consists of two steps. First, we obtain the payoffs $\pi(\sigma_i, \sigma_j)$ between two strategies σ_i and σ_j. The previously described simulation-based estimation is used to obtain the payoffs in most cases, including the payoff between two cumulative reciprocators, and those between a cumulative reciprocator and a player who uses either an arbitrary memory-one strategy or one of the 9 strategies selected in 2-strategy competitions. For the payoffs between two memory-one strategies we use the analytical solution of Press & Dyson [112].

Second, we calculate the strategies' frequencies during the process through the "survival of the fittest" in a noisy environment based on the obtained payoffs between pairs of strategies. In each generation, the evolutionary fitness of each strategy is calculated. Following Nowak & Sigmund's approach [145], the fitness of a strategy σ_i is defined by its cumulative payoff when playing the repeated game with the entire population (i.e., $f(\sigma_i) = \sum_{j=1}^{n} x_j \pi(\sigma_i, \sigma_j)$), where n is the number of strategies in the population, and x_i and x_j are the frequencies of σ_i and σ_j, respectively. We denote the overall fitness of all strategies by $\bar{f} = \sum_(i = 1)^n x_i f(\sigma_i)$. The frequency of σ_i in the next generation is determined to be $x_i^{'} = x_i * \frac{f(\sigma_i)}{\bar{f}}$. This elementary updating process is repeated for many generations.

To simulate evolution in 2-strategy populations, each simulation ends if the frequency of each strategy no longer changes, indicating two possible steady states (i.e., either the full invasion of one strategy into the other or the coexistence of the two strategies). One million generations are executed in each simulation for 3-strategy populations and populations of multiple strategies. When simulating the frequency dynamics of strategies in heterogeneous populations, we sometimes allow for mutations. In that case, mutations are introduced after 2,000 generations of a simulation. The mutation rate is set to 10%. When a mutation happens, all other strategies decrease their proportions to 99.9%, while a strategy is randomly selected to increase its proportion by 0.1%.

4.4.3 Behavioral experiments

The data in Fig.4.6 displays results from an economic experiment. For this experiment, we recruited subjects through the online platform Prolific (www.prolific.co). In total, we report data from 172 subjects, who all gave their informed consent to participate. Participants were randomly allocated to one of two possible treatments. In both treatments, participants engage in a repeated prisoner's dilemma based on our baseline payoff values (in UK pence: R=15p, S=0p, T=25p, P=5p, which corresponds to the values by Axelrod multiplied by a factor of five). Moreover, in both treatments, participants engage for at least 20 rounds. After the 20th round, there is a constant stopping probability of 1/2 after each round, to avoid end-game effects. For better comparison, we only use data of the first 20 rounds for our statistical analysis.

The two treatments differ in how the players' intended actions are implemented. In the treatment without errors, all the players' decisions are implemented faithfully. In the treatment with errors, the players' decisions are mis-implemented with a probability of 10%. In case of an error, an intended cooperation is executed as a defection, and vice versa. Participants know the overall error probability, and they learn if their own decision was mis-implemented. However, they do not know whether or not the co-player's decision was implemented faithfully.

In each round, participants learn how often each player has cooperated so far (in the treatment with errors, this number refers to how often a player's decision was implemented as cooperation). Thereafter, they make their decision whether or not to cooperate in the next round. After both players have made their decisions, the (implemented) outcome of that round is displayed.

To explore to which extent human play is predicted by various well-known cooperative strategies, we consider eight possible template strategies. These eight strategies are GRIM, tit-for-tat (TFT), firm-but-fair (FBF), win-stay lose-shift (WSLS), AON2, TFT-ATFT, CAPRI, and CURE (with threshold $\Delta = 3$). The first four strategies cooperate in the first round; thereafter, their response is given by the memory-1 vectors

$$p_G RIM = (1,0,0,0), p_T FT = (1,0,1,0), p_F BF = (1,0,1,1), p_W SLS = (1,0,0,1).$$

For the higher-memory strategies AON220, TFT-ATFT28, and CAPRI21, we provide the definitions of the implemented strategies in Section 11 of the Supplementary Information. For each participant, we compute how many of the participants' decisions are correctly predicted by each of these eight template strategies. The corresponding results are displayed in Fig.4.6c. For a more detailed description of the experimental methods, see Supplementary Section 11. The collected data and screenshots of our interactive game software are available online.

Statistics and reproducibility. For our behavioral experiment, all statistical tests are two-tailed and non-parametric. No statistical method was used to predetermine sample size. Results are based on the data of all 172 subjects that finished the experiment, not considering fifteen subjects who dropped out during the instructions, and two more subjects who dropped out during the experiment. Assignment to treatments was randomized. For further details on the study design, see Supplementary Section 11. The computer code for our simulations and for our behavioral experiments, as well as the resulting data, are available online.

Chapter 5

(Mis)perceiving Cooperativeness

This chapter is published under the title: Rossetti, C.S.L., Hilbe, C., Hauser, O.P. (2022). (Mis)perceiving cooperativeness, *Current Opinion in Psychology*, 43, 151–155.
https://doi.org/10.1016/j.copsyc.2021.06.020 [86].

Contents

Abstract

Cooperation is crucial for the success of social interactions. Given its importance, humans should readily be able to use available cues to predict how likely others are to cooperate. Here, we review the empirical literature on how accurate such predictions are. To this end, we distinguish between three classes of cues: behavioral (including past decisions), personal (including gender, attractiveness and group membership), and situational (including the benefits to cooperation, and the ability to communicate with each other). We discuss (i) how each cue correlates with future cooperative decisions, and (ii) whether people correctly anticipate each cue's predictive value. We find that people are fairly accurate in interpreting behavioral and situational cues. However, they often misperceive the value of personal cues.

5.1 Introduction

When people cooperate, they make an individual effort to benefit others [215]. Such voluntary acts of costly cooperation are crucial for most social interactions, affecting the well-being of families, the success of scientific collaborations, and the output of team work. Cooperative behavior can be explored from various angles. Theoretical studies analyze which social environments are most conducive to cooperation [5, 60, 216]. Similarly, experimental work explores under which conditions people actually cooperate, and how they adapt their own behavior to the actual or expected behavior of their peers [7, 73, 74, 217]. Here, we review this literature, focusing on the following two questions: How do people use available cues to predict the cooperativeness of their interaction partners ahead of an interaction, and which cues are most reliable for predicting cooperativeness?

We distinguish between three broad categories of cues that may be used to make such predictions. (i) Behavioral cues refer to the interaction partners' past actions. Examples of such cues include whether these partners cooperated on previous occasions, or whether they enforced cooperative social norms. (ii) Personal attributes comprise of, for example, the interaction partner's gender, or perceived attractiveness, among other characteristics. (iii) Situational cues define the environment in which the individual's next interaction takes place, including how costly cooperation will be, and whether pre-play communication is possible. For each of the three categories, we ask whether the given cue is in fact a reliable predictor of cooperativeness, as well as whether people judge the cue's predictive value accurately. While the existing literature covers a wide range of relevant cues, there also exist notable gaps in the literature, which we encourage scholars to explore further.

There are various ways to elicit how individuals perceive each other's cooperativeness. For the purpose of this review, we consider evidence from three approaches. The first approach is to elicit perceived cooperativeness directly by asking participants to estimate how likely others will cooperate. The two other approaches are more indirect, by either asking participants to choose a group member for future interactions (i.e. partner selection), or by asking participants how much money they would transfer to the respective group member in a trust game [218]. Our assumption is that the more inclined a participant is to choose an interaction partner, or to transfer more money to a partner, with whom they will interact in the future, the more cooperative they perceive the partner to be – and indeed, some empirical work has found this to be the case [219].

5.2 Behavioral cues of cooperativeness

Perhaps the most immediate cue to predict future cooperative behavior is whether, and how often, the respective individual cooperated in the past. Experimental research suggests that people exhibit a stable and consistent 'cooperative phenotype': an individual's decision in one cooperative game is indicative of what that individual will subsequently do in different games [220, 221]. Participants in laboratory experiments, in turn, seem to expect others to be consistent in their cooperative behavior. When people need to choose an interaction partner, they strongly prefer partners who have been cooperative in the past [222]. Similar evidence comes from studies on charitable giving and pro-environment behavior. Donors to charity are trusted more as well as chosen more often as interaction partners, and in many cases, they indeed turn out to be more cooperative in subsequent social dilemmas [223, 224], but see also [225].

Another – more indirect – cue of cooperative behavior is whether individuals previously engaged in the enforcement of social norms. According to this account, people who punish selfishness may signal that they are not selfish. To test this hypothesis, [226] consider an interaction that consists of two stages. In the first stage, a 'signaler' witnesses a transgressor who refuses to help a recipient. The signaler can then decide whether to engage in third-party punishment by reducing the payoff of the transgressor. In the second stage, a 'chooser' decides how much money to send to the signaler in a trust game. The experiment shows that signalers who punish transgressors are indeed entrusted with more money, which turns out to be justified: these signalers also return more money to the chooser. Interestingly, when in the first stage signalers have a choice between punishing the transgressor or helping the recipient, signalers are less likely to punish. Instead, helping turns out to be the more frequently chosen (and more accurate) signal of trustworthiness. This result is in line with an earlier experiment by [227]: When individuals need to choose group members for a cooperative task, they place more weight on how often group members cooperated, rather than how often they enforced cooperation. Overall, punishment is not necessarily taken as a cue of altruism, as it may also imply aggressiveness. As a result, punishment is judged more appropriate if it is implemented by the entire group, rather than by a single individual [228].

Finally, another potential cue may come from how a person makes cooperative decisions. For example, based on a game-theoretic model, [229] suggest that people who collect additional information to carefully compare the advantages and disadvantages of a cooperative decision are considered as less reliable and less cooperative. According to this account, people who deliberately refuse to learn payoff-relevant information may seem more committed to cooperation even when defection happens to be profitable. In line with this view, [226] show that study participants who ignore the precise costs of cooperation are indeed (and accurately) predicted to be more trustworthy. Participants in turn seemed to be well aware of the reputational benefit of strategic ignorance: when the cooperation costs can be learnt secretly, participants were more likely to do so. Similar evidence comes from [230] who compare the reputational consequences of emotion-based versus reason-based decision-making. Players who state having made a decision based on emotion are perceived as more cooperative by their partner, and indeed turn out to be more cooperative. Interestingly, however, players who state having used reason where not perceived as any less cooperative than a control group.

5.3 Personal cues of cooperativeness

A second category of cues pertains to personal characteristics of individuals. The idea that visible characteristics are used as signals of cooperativeness has its roots in evolutionary biology: computer simulations and game-theoretic models suggest that individuals can use visible cues to identify potential cooperation partners. This theoretical work suggests that partner choice based on visible cues can in turn be an important mechanism for the evolution of cooperation [231].

Arguably, one of the most salient personal cues is gender. While some scholars document gender differences in cooperativeness in the Dictator Game [232] and the Prisoner's Dilemma [233], the literature remains notoriously mixed [234–236]. Indeed, [237] show that gender only inconsistently predicts cooperativeness across seven economic games. Nevertheless, across all economic games, they find robust evidence that people consistently believe that women are expected to be fairer, more generous, and more cooperative than men, in line with [238] and [239].

Other cues about a person include their physical appearance, such as an interaction partner's face and facial expressions [240–242], their voice [243] and, in particular, their attractiveness [244, 245]. While there is no evidence that attractiveness is a reliable predictor of cooperativeness, good looks nonetheless positively impact people's perception of others in many domains of economic life, known as the 'beauty premium' [61]. Indeed, [244] and [245] find that people expect more cooperation and reciprocation from attractive partners. Interestingly, when such expectations are not met, attractive interaction partners incur a 'beauty penalty', receiving less reciprocation compared to less attractive players.

Another factor that may affect a partner's cooperativeness is wealth. Here, the experimental evidence remains inconclusive. For example, [246] find that subjects who consider their own socio-economic rank to be low tend to be more generous and charitable. In contrast, [247] report that millionaires are considerably more generous in dictator games than usual participants, especially if they are paired with a low-income partner. A similar conflicting picture emerges on the level of perceived cooperativeness. Some experiments find that wealthy participants are perceived to be more trustworthy and cooperative [248, 249], while at the same time, people seem to systematically underestimate the generosity of the extremely rich [250].

A person's political, religious, or ethical convictions can also serve as potential cues of cooperativeness. Research on political affiliation finds that left-leaning participants tend to cooperate more than right-leaning participants. However, the effect is small at best [251, 252], and it seems to be moderated by the fact that left-leaning participants expect more cooperation from others [253]. Political ideology in turn shapes how people are perceived. [254] find that among US participants, Democrats are perceived as more cooperative by both sides of the political spectrum, even though this belief is inaccurate. Similar to political ideology, religiosity appears to be correlated with prosocial behavior [255, 256]. As a result, when Christians show overt religious cues (e.g. a necklace with a cross), they are perceived as more trustworthy [257]. Finally, people are considered more trustworthy when they make deontological rather than consequentialist judgments [258, 259]. Interestingly, however, deontological participants are not necessarily more cooperative [260].

More often than not, when it comes to social attributes, perceptions of cooperativeness are partially shaped by in-group bias, as group membership is itself an important cue. A large literature demonstrates that participants cooperate more with, and preferentially reward, 'in-group' members over 'out-group' members, both in the laboratory [261] and in the field [262]. Democrats and Republicans both tend to cooperate more with in-group members [254], and participants cooperate more with a partner that shares their nationality [263]. However, group membership may largely serve as a coordination device: while participants do not believe that in-group members are intrinsically more cooperative than others, [264] find that more cooperation is expected from in-group members.

5.4 Situational cues of cooperativeness

Our final category of cues covers situational aspects surrounding a cooperative decision. These include factors outside an individual's control that are often determined by the structure or context of the interaction. For example, some social interactions create more mutual benefits to the participants than others. In laboratory studies, these factors can be studied in isolation, holding everything else constant. For instance, in the repeated Prisoner's Dilemma, people tend to be more cooperative when the mutual benefit of cooperation increases [126]. This is consistent with evidence by [265]: as the mutual benefit of cooperation increases, participants also expect to see more cooperativeness, especially those participants who later choose to cooperate. This suggests that individuals are able to "read" a situation, and that they adjust their willingness to cooperate accordingly.

There has been a debate on whether cooperativeness is affected by whether or not decisions need to be made under time pressure. One account holds that when individuals are forced to decide quickly, they tend to be more cooperative [266]. However, the causal evidence is mixed [267]. Alternatively, it has been argued that fast decisions may not necessarily result in more cooperation. Instead, it may result in more extreme outcomes, either towards cooperation or defection [268]. This is also what participants themselves seem to expect: when asked to predict the outcome of a fast cooperation decision, participants are more likely to expect an extreme (but not necessarily a cooperative) outcome [269].

Finally, communication has long been found to enhance cooperation [270]. However, there is debate as to what is the precise mechanism that allows communication to be favourable. [271] rule out several potential explanations, namely that communication reduces social distance or that it offers opportunities to make promises. Instead, they argue that most importantly, communication allows people to recognize the other person as a cooperative type. Indeed, [272] find that people can accurately predict cooperative behaviour in a prisoner's dilemma after only a short in-person interaction, even when participants do not discuss the game itself.

5.5 Conclusion and future directions

People are quick to form an impression, sometimes in just a few milliseconds [240], but these impressions do not need to be reliable. In this review, we have summarized how individuals perceive different cues in order to predict others' cooperativeness. Among the three categories of cues we considered, people most accurately use behavioral and situational cues to predict future cooperative behavior. In contrast, predictions seem to be least accurate when they are based on personal attributes. In fact, with the possible exception of political affiliation and religiosity, personal attributes are often not a good predictor of actual cooperation behavior, yet many such attributes'such as gender and attractiveness'are nonetheless (and inaccurately) perceived as predictors of cooperativeness.

These findings raise a number of interesting questions. For example, theoretical work could explore which kinds of environments and cues allow people to form reliable expectations. Similarly, experimental work could investigate how persistent certain misperceptions are, and whether they disappear with more experience. While we reported results on a number of different cues, several others seem to have received limited attention, including age and ethnicity.

Further interesting problems arise when several (possibly conflicting) cues are available, or when people need to aggregate cues from different domains. For example, [263] show that when people learn both their interaction partner's nationality and gender, the former cue becomes dominant: people cooperate more with own-nationality partners, and they expect own-nationality partners to be more cooperative.

More generally, any given cue seems to become less relevant once more salient information

– such as actual cooperative behaviour – is available [244, 273–275]. These findings suggest that people rank different cues according to each cue's predictive value. They only make use of unreliable cues when no other cues are available.

Chapter 6

A Model of Principled Behaviour

This chapter is part of an ongoing theoretical and experimental project with Moshe Hoffman and Christian Hilbe.

Contents

Abstract

In this chapter, we provide a reputation-based theory of principled behaviour. We refer to a behaviour as principled if individuals display it consistently, regardless of the associated costs in each context. Here, we hypothesize that commitment and consistency may function as a signal of trustworthiness that in turn attracts cooperative partners. Individuals who act principled benefit from these cooperative relationships, which outweigh the costs of such a rigid behaviour. We create a game-theoretic signalling model with two types of senders. First, there are committed senders for whom it is generally worthwhile to display consistent behavior. Second, there are "impersonators" who try to pass as committed individuals to reap the benefits of those cooperative relationships. We characterise a "zero-tolerance" profile and show the conditions under which principled behaviour occurs in equilibrium. We compare these results to another competing theory of moral behaviour, moral licensing. Finally, we discuss how to experimentally test the predictions from the model.

6.1 Introduction

Principled behaviour, characterised by unwavering adherence to moral values, presents a fascinating puzzle oh human behaviour. People who act principled exhibit a range of specific behaviours: (1) they consistently abide by the principle, (2) they refuse to compromise or consider trade-offs, (3) they prioritise the principle above all else, even when it is of little value, and (4) they feel that their principle is right for transcendental reasons rather than material incentives. People abide by values such as never lying or cheating, never eating pork, or beef, or any meat at all, never taking the plane, never to monetize love, sex, bodies, etc. In this chapter, we do not dwell on the philosophical question of what are moral values, what is morality or the different shapes it might take. We care about the behaviour of consistently abiding by stated moral principles regardless of costs and benefits and with no intrinsic value or reason for such behaviour.

Such a strict rule of behaviour is puzzling. Human behaviour is evolved. The same pressure from natural selection and sexual selection that shape the different functions of an organism shape the human mind [276]. Only behaviour that brings fitness benefits through reproduction or the acquisition of material resources, will evolve. However, all the features of principled behaviour describe a rigid behaviour that does not vary according to the costs and benefit of the situation, and as such may lead to direct negative impact. That individuals exhibit a behaviour that is blind to payoffs and potentially costly to themselves defies evolution. Adding to the complexity, the values individuals abide by vary considerably within and across cultures [277]. Different religions and culture proscribe different foods, while people themselves have different individual values within the same society. The moral values of right and left wing people are fundamentally different [278], and which set one person adheres to might change over the lifetime. Despite this variation, which types of values and behaviour individual become principled about might not be random.

Why do individuals act principled and what kind of behaviour are they principled about? To answer these questions, we look at the ultimate function of principled behaviour. Building on the literature of signalling games [113, 279], we develop a model to test a reputation-based theory of principled behaviour [280]. Following a growing body of literature that shows that many puzzling behaviour humans display are signals of trustworthiness aimed at potential cooperation partners [80, 281–283], we use evolutionary game theory to build a signalling model [229, 284–286]. Signalling theory has roots in both biology [287, 288] and economics [289]. Both approaches seek to understand the ultimate cause of a behaviour by analysing which information is conveyed by a behaviour. In their theory, Singh & Hoffman [280] reason that by acting principled, individuals signal consistent commitment to cooperative behaviour regardless of the context. This in turn makes them look trustworthy, as potential partners can rely on their cooperation even when the costs and benefits of a situation change.

There is evidence that individuals who display features of principled behaviour tend to be trusted more. Individuals who, in a trolley problem, choose according to moral rules of what is right rather than weighing the consequences by balancing the cost and benefit of an action, are rated more trustworthy and are trusted more in trust games [258–260]. Jordan *et al* [226] show that individuals who choose not to know the costs of of helping are trusted more and that players are willing to act in this way to signal trustworthiness. Similarly, individuals who choose to help intuitively and uncalculatingly are trusted more [290]. Moreover, we value consistency in ourselves greatly. Cognitive dissonance is a widely known effect in psychology by which people will do everything in their power to change actions or ideas that are not psychologically consistent with each other [291] Similarly, foot-in-the-door effects, where individuals are more likely to agree to a big demand after having accepted a smaller one, stem from a willingness to remain consistent [292]. These effects can be explained as an aversion to dissonance between one's actions and beliefs. But why this aversion? An aversion to inconsistency is only a proximate psychological mechanism that evolved to ensure individuals avoid displaying such behaviour. Here, we hypothesise that its ultimate function is to regulate individuals' reputation of trustworthiness.

In our model, principled behaviour is a non-conscious impersonation of a 'committed individual' [280]: a committed individual is an individual with a psychology so extreme that they never deviate from their principles under any circumstances. The "never" is crucial. It means even in difficult situations or when the stakes change, they can still be trusted. This leads individuals to interact with them differently as they would with others. If someone never deviates from being loyal, people will select that person as a cooperative partner over someone who considers costs and benefits. If someone never deviates from violently retaliating, even at a cost to themselves, people will avoid attacking them. Those individuals are rare, and most of the population are simple impersonators who attempt to pass as them to reap the benefits of this reputation. As such, impersonators should exhibit the same behaviour as committed individuals: consistently abiding by the principle, refusing trade-offs, and prioritising the principle.

In particular, we focus on one key feature of principled behaviour: consistency. It is the rigidity that principled individuals display that specifically garners trust. Consistency in our model has two manifestations: complementarity and discontinuity. Complementarity means that the more cooperative one has been in the past, the more cooperative one will be in the future. Discontinuity means that after one has defected once, one will no longer bother cooperating in the future. Complementarity is similar to investing in trustworthiness. The more one invests, the more one has to lose from deviating. Discontinuity implies that once a principle is betrayed, one has revealed that one is not truly principled and trust can never be fully regained. These are the mechanisms hiding behind cognitive dissonance and the foot-in-the-door effect. As people already exhibited principled behaviour once, they already invested in their reputation. Betraying the principle now will completely and irrevocably undo this reputation, so people stick with what they already did.

However, there is another recognised effect in consecutive moral decisions known as moral licensing. This psychological effect occurs when individuals who already behaved morally or cooperatively feel subsequently entitled to act opposite to the moral values they already upheld, because they have provided evidence that they abide by them. Essentially, once an individual has already done something moral, they have a license to do something immoral, as they have now earned "moral credits". For example, participants who disagreed with a strongly sexist statement were more likely to favour a male candidate for a job position [293]. The evidence for moral licensing is debated, with meta-analyses finding a small but real effect [294] and a recent study failing to replicate the effect [295]. There is evidence that the effect is sensitive to several moderators, perhaps explaining why the effect is hard to reproduce. In their review of the literature, Mullen & Monin [296] highlight a range of moderators which distinguish when individuals are more likely to be consistent or to license. Overall, the evidence shows that the likelihood of maintaining consistency in actions increases when individuals demonstrate a strong commitment to a particular value or strongly identify with the principle underlying the action. Conversely, actions perceived as less concrete, with tangible consequences or less indicative of moral principles, are more likely to lead to moral licensing. Our model helps to shed light on the conditions necessary for either consistency or licensing. Our reputation-based theory states that moral licensing can be explained by a desire not to appear as a negative committed individual. In the same way that one breach of principle is sufficient evidence to show that one is truly committed to a positive action, one positive action can demonstrate that a person is not committed to a negative one. There is evidence that moral licensing is a reputation mechanism: once an individual has established that they are good, they no longer have to behave as morally anymore [295]. To summarise, we expect a person who needs to be seen as trustworthy not to display moral licensing, while an individual who only wishes to establish they are not the worst possible person will license.

In this chapter, I present the signalling model and characterise equilibrium conditions for strategy profiles representing consistency and moral licensing. We hypothesise that principled behaviour is indirectly beneficial because acting consistently increases trust in potential partners. To explore this hypothesis, I explore under which conditions there is an equilibrium that features consistent behavior, and how these conditions differ from the respective conditions of the moral licensing equilibrium. Finally, I sketch a behavioural experiment that can put the resulting model predictions to the test.

6.2 The model

R	Receiver
S	Superior sender
I	Inferior sender
p^S	Probability of a receiver to be of superior type
p^I	Probability of a receiver to be of inferior type
a^S	Benefit for the superior type if accepted by a receiver
a^I	Benefit for the inferior type if accepted by a receiver
b^S	Benefit from accepting superior sender
b^I	Benefit from accepting inferior sender
c_h	High cost
c_l	Low cost
q_l^S	probability for the superior type to face a low cost
q_l^I	probability for the inferior type to face a low cost

Table 6.1: Parameters

To model principled behaviour, we develop a repeated signalling game between a sender and a receiver. The sender can be one of two types, referred to as superior (S) and inferior (I). The type of the sender is randomly determined before the game starts according to the probability distribution $\mathbf{p} = (p^S, p^I)$ with $p^S + p^I = 1$. Senders know their own types, whereas receivers only know the general probability distribution **p**.

After the sender's type is determined, the sender faces n rounds in which she can signal her type to the receiver. Each of these n rounds proceeds as follows: First nature randomly determines the cost of the signal; this cost can be either low or high, $c \in \{c_l, c_h\}$ with $c_h > c_l > 0$. The probability of having a high cost is independent of the round number and of previous realised costs. However, it depends on the type of the sender. Superior senders experience a high cost with probability q_h^S, whereas this probability is q_h^I for inferior types. We use $q_l^S = 1 - q_h^S$ and $q_l^I = 1 - q_h^I$ for the probability that the respective cost is low. Moreover, we use $\bar{c}^\theta := q_l^\theta c_l + q_h^\theta c_h$ to denote the expected cost that a sender of type $\theta \in \{S, I\}$ experiences in any given round. In the following we assume $q_h^S \leq q_h^I$. That is, superior types are less likely to experience a high cost. The sender (but not the receiver) learns the exact cost. Then the sender decides whether or not to pay the cost in order to send the signal. After the sender has made her decisions, the

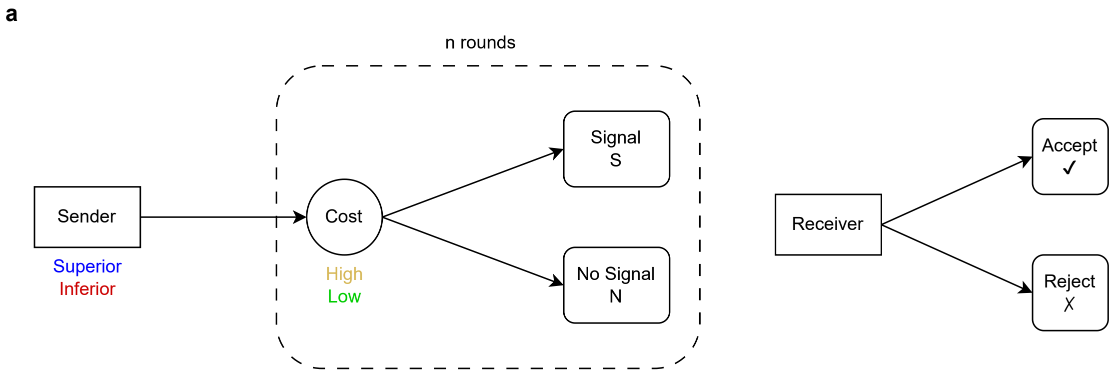


Figure 6.1: The repeated signalling game. We consider a signalling games between a sender and a receiver. Sender can be of two types: superior or inferior. To indicate their type, senders have the opportunity to send a signal on multiple rounds. On each round, they may pay a cost to send the signal (S). Otherwise no signal is sent (N). The cost can be either high or low and is randomly selected in each round. Based on the number of signals they see, the receivers choose to accept or reject the sender for some benefit.

receiver learns in how many of the n rounds the sender paid the cost. The receiver then decides whether or not to accept the sender. If the receiver rejects the sender, the payoffs in that final stage are zero for both players. If he accepts the sender, the payoffs depend on the sender's type; the sender's payoff is a^θ and the receiver's payoff is b^θ with $\theta \in \{S, I\}$ (Fig.6.1).

For the game to be interesting, we make the following assumptions on the players' incentives:

A1 Receivers only want to pair with superior senders

$$b^S > 0 > b^I. \tag{6.1}$$

A2 Senders are generally willing to signal

$$n \cdot c_l < a^\theta \text{ for all } \theta \in \{S, I\}. \tag{6.2}$$

A3 Some signalling is required to get accepted

$$p^S b^S + p^I b^I < 0. \tag{6.3}$$

A4 Inferior senders do not signal at all costs

$$n \cdot c_h > a^I. \tag{6.4}$$

Assumption (A1) states that receivers do not want to accept any sender but only superior senders. If the benefit from accepting an inferior sender is also positive but still smaller than the benefit of accepting an superior sender, receivers would simply accept anyone. The assumptions (A2) and (A3) rule out certain scenarios where no player ever sends the signal. Senders may gain a benefit from being accepted and therefore signalling is worthwhile. Assumption (A4) implies that inferior senders will not send the signal when the cost is high in each round.

This model has two key features: there is more than one opportunity to signal, and the related costs are stochastic. There needs to be different cost levels to distinguish a perfect sender – superior, from an imposter sender – inferior. Specifically, a perfect sender has no or very little consideration for the costs, while an imposter is incentivised to take the costs into account. The presence of a high cost gives the inferior sender the opportunity to reveal their true colours. There also needs to be multiple rounds so that the effect of previous actions can be modelled and consistency tested. Both our behaviours of interest, principled behaviour, and the competing moral licensing, are about consecutive moral decisions where past actions influence the current action. For principled behaviour in particular, we want to test two features of consistency: complementarity and discontinuity. In the following, we consider the case of $n = 2$ rounds, as this is sufficient to test for these two features. In a two-rounds model, complementarity means that senders send the signal the second round if and only if they already sent the signal in the first round. Discontinuity means that if the sender did not send the signal in the first round, they will not send it in the second round either.

To analyse the model, we characterise strategies for each type of player in the repeated signalling game. A strategy is a plan of actions. For example, a strategy for the receiver could be to never accept regardless of how many signals are observed. Another strategy would be to accept when at least one signal is sent. We assume the senders' strategies take the form $\sigma^\theta = \sigma^\theta(i, k, c)$. Here, the input $i \in \{1, \dots, n\}$ is the given round for which a decision is to be made; $k \in \{0, \dots, i-1\}$ is the number of times a signal has been sent in the past, and $c \in \{c_l, c_h\}$ is the current cost. The output takes the form $\sigma^\theta(i, k, c) \in \{0, 1\}$ where 0 refers to not sending the signal for the given input. The receiver's strategy takes the form $\rho = \rho(k)$, where $k \in \{0, \dots, n\}$ refers to the number of times the sender has sent the signal. Here, the output again takes values $\rho(k) \in \{0, 1\}$, where 0 means to reject the sender.

6.3 Equilibrium analysis

To explore the implications of this model we characterise the relevant Perfect Bayesian Nash equilibria (PBNE). The notion of PBNE is the standard way game theorists solve signalling games. They require that at each stage of the game, individuals act optimally given their beliefs. The beliefs in turn are formed using Bayes' rule. In our signalling game, observing that a signal was sent or not is new information that receivers use to update their beliefs about the type of the sender. In a PBNE, the strategy of each type of each player is specified in such a way that no player can gain from deviating given her preferences and her information. If a strategy profile is not a PBNE, then some player could benefit from deviating. Equivalently, if strategies are learned or evolved, a mutation or experimentation that leads her to behave differently would succeed and propagate.

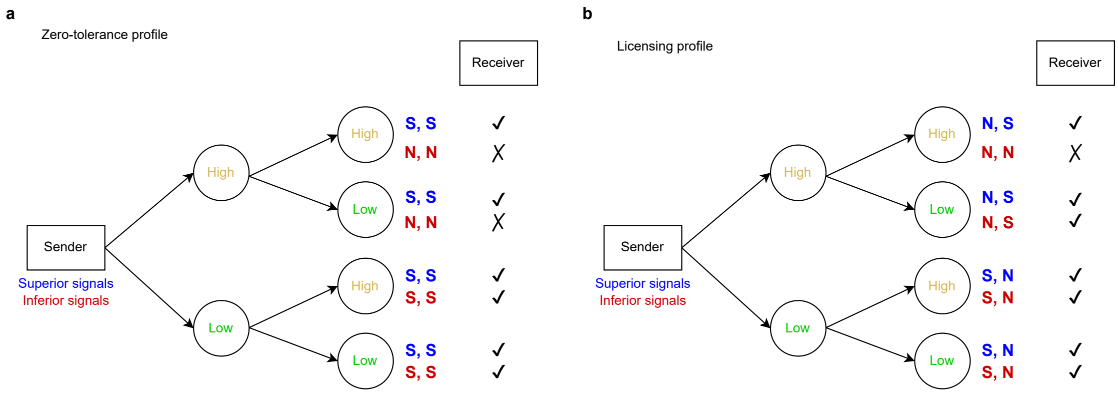


Figure 6.2: The strategy profiles. a. We define a zero-tolerance equilibrium in which receivers only accept senders who send the signal in all rounds (twice), superior senders send the signal in every round regardless of the cost, and inferior senders send the signal on the first round if the cost is low or as long as they have sent it before. **b.** We define a licensing equilibrium in which receiver accept sender who send the signal at least once, superior senders will always send the signal in the last round if they have not before, and inferior sender only send the signal in the last round if they have not before and the cost is low.

Zero-tolerance equilibrium. We first characterise a *zero-tolerance equilibrium* where receivers only accept "perfect" senders who send the signal in all rounds. Specifically, we consider the following strategy profile

$$\sigma^S(i,k,c) = \begin{cases} 1 & \text{if } k = i-1 \\ 0 & \text{otherwise} \end{cases} \tag{6.5}$$

$$\sigma^I(i,k,c) = \begin{cases} 1 & \text{if } (i\!=\!1 \text{ and } c\!=\!c_l) \text{ or } (i\!=\!2 \text{ and } k\!=\!1) \\ 0 & \text{otherwise} \end{cases} \tag{6.6}$$

$$\rho(k) = \begin{cases} 1 & \text{if } k = n \\ 0 & \text{otherwise} \end{cases} \tag{6.7}$$

We refer to the above strategy profile as the *zero tolerance profile* (Fig.6.2a). In this profile, superior senders send the signal in all rounds, regardless of costs. Inferior senders send the signal in the first round if the respective cost happens to be low; in the second round they send it as long as they sent it in the round before. The receiver only accepts senders who send the signal in all rounds (twice). The name zero-tolerance refers to the high standards of the receiver as they wish to only be paired with truly committed individuals.

This strategy profile displays the two features of principled behaviour we are interested in: complementarity and discontinuity. It shows complementarity because inferior types who send the signal in the first round also send it in the second. It shows discontinuity because senders who failed to signal in the first round also send no signal in the second. Inferior senders will signal in both rounds, and be accepted by receivers, when the conditions (the costs) are right. This way, the inferior sender can pass as a superior sender and be a successful imposter. By checking all possible deviations, we characterise when we can expect such behaviour in equilibrium.

Claim 1. *There is a perfect Bayesian Nash equilibrium in which players act according to the zero* *tolerance profile if and only if the following conditions hold:*

$$-\frac{b^S}{b^I} \geq \frac{1-p^S}{p^S}(1-q_h^I),$$
$$a^S \geq c_h + \bar{c}^S, \tag{6.8}$$
$$c_h + \bar{c}^I \geq a^I \geq c_l + \bar{c}^I.$$

The first condition states that the receiver finds it worthwhile to accept a sender who sent the signal in both rounds. This condition is harder to satisfy if the probability of encountering a superior type is low (i.e., if $p^S \approx 0$), or if inferior senders are unlikely to face a high signalling cost (i.e., if $q_h^I \approx 0$). The second condition ensures that superior senders are willing to pay the high cost even if it occurs in the first round. The last condition ensures that in the first round, inferior senders are willing to pay the low cost but not the high cost.

Moral licensing equilibrium. Moral licensing is a psychological effect by which individuals do not bother doing the 'right thing' when they already have done so before. The initial good deed gives them a "license" not to act good a subsequent time. Here, we attempt to capture moral licensing through the following strategy profile.

$$\sigma^S(i,k,c) = \begin{cases} 1 & \text{if } (i=1 \text{ and } c=c_l) \text{ or } (i=2 \text{ and } k=0) \\ 0 & \text{otherwise} \end{cases} \tag{6.9}$$

$$\sigma^I(i,k,c) = \begin{cases} 0 & \text{if } c=c_h \text{ or } (i=2 \text{ and } k=1) \\ 1 & \text{otherwise} \end{cases} \tag{6.10}$$

$$\rho(k) = \begin{cases} 1 & \text{if } k \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.11}$$

We refer to the above strategy profile as the *licensing profile* (Fig.6.2b). With this profile, we capture the features of moral licensing in both senders. The difference between the two sender types is their sensitivity to the cost. The superior type will always 'license' by sending the signal in the second round regardless of cost if they did not send it before. In contrast, the inferior type only 'licenses' in the second round if the cost is low. Receivers now accept senders who sent at least one signal. For this licensing profile we obtain the following equilibrium condition.

Claim 2. *There is a Perfect Bayesian Nash equilibrium in which players act according to the licensing profile if and only if the following conditions hold:*

$$-\frac{b^S}{b^I} \geq \frac{1-p^S}{p^S}\left(1-(q_h^I)^2\right),$$
$$a^S \geq c_h \geq a^I. \tag{6.12}$$

Similar to before, the first condition ensures that receivers have no incentive to deviate. This condition is now harder to satisfy than the equivalent condition for the zero-tolerance equilibrium, Eq. (6.8), because the right hand side of the inequality is larger. This is because in the licensing equilibrium, inferior types find it easier to pool with the superior senders. The only way they are rejected by the receivers is if they face a high cost in both rounds (whereas in the zero tolerance equilibrium, they are already rejected if they face a high cost in the first round).

The second condition ensures that senders find it worthwhile to follow the licensing profile. This condition states that the inferior sender does not have an incentive to pay the high cost, whereas the superior sender does. Compared to the respective constraints in Eq. (6.8), these conditions on a^S and a^I are now easier to satisfy.

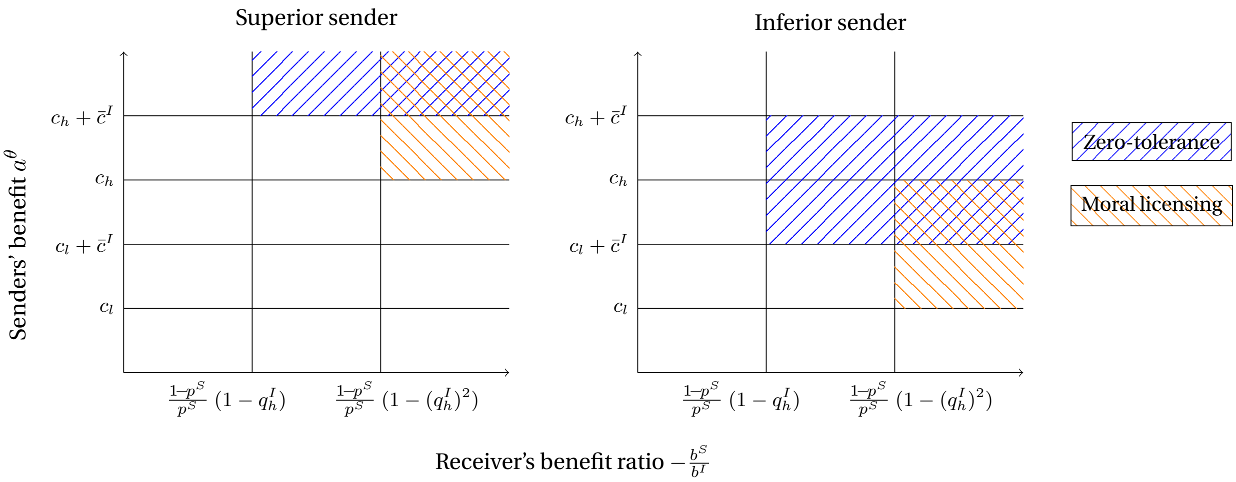


Figure 6.3: Comparison of zero-tolerance with competing equilibria. The conditions that allow for each equilibrium configurations depend on the benefit that the receiver gets (x axis) and the benefit that senders get (y axis). In the case of the superior sender, both the zero-tolerance and the moral licensing equilibria are only possible if the benefit a^I is greater than the high cost. For the inferior sender, the benefit a^I must be smaller than the high cost but greater than the low cost. In the case of the receiver, the conditions on the benefit ratio $-\frac{b^S}{b^I}$ are harder to satisfy for the moral licensing equilibrium.

No accepting, no signaling. As a final contrast, consider the profile in which no sender regardless of type sends the signal in any round, and receivers reject everyone independent of the signal,

$$\sigma^\theta(i,k,c) = 0 \quad \forall i,k,c,\theta \tag{6.13}$$

$$\rho(k) = 0 \quad \forall k \tag{6.14}$$

Such a profile can always arise as a Bayesian Nash equilibrium. Given that receivers reject anyone, senders have no incentive to pay a cost for the signal. Conversely, given that all senders behave in the same way, receivers have no incentive to accept them because of assumption (A3). That is, we have:

Claim 3. *Under the assumptions (A1) – (A4), there is always a Perfect Bayesian Nash equilibrium in which players act according to the no-accepting profile.*

6.4 Discussion

Why do individuals act principled and what kind of behaviour are they principled about? With this model, we sought to explain the ultimate function of principled behaviour. We hypothesised that individuals engage in principled behaviour because it allows them to gain the trust of potential cooperation partners. Being accepted by a partner and entering a cooperative relationship is beneficial, and must outweigh the costs of acting principled. Our theory also suggests that key features of principled behaviour, complementarity and discontinuity, arise from the selectivity of the receivers who only benefit from partnering with someone who is truly principled. Thus, any transgression will irreparably discredit an individual in the eyes of the receiver, who will now know the truth. Similarly, consistently abiding by the principle is the only way to demosntrate one's commitment to it, and with each abiding action, the reputation capital of trustworthiness increase. Individuals have more and more to lose by deviating, leading to complementarity.

Our results show that there is an equilibrium where senders display consistency while receivers only accept senders who sent the signal in all rounds. Moreover, it is the only possible equilibrium with a highly selective receiver. Other sender behaviours that still involve some signalling are not possible in equilibrium with a maximally choosy receiver. Similarly, the licensing equilibrium shows that if receivers relax their criteria, senders will not act consistently in equilibrium, but will license and send the signal in only one round whenever possible. Any strategy that leads the senders to send in both rounds even once will not be in equilibrium. In particular, our conditions on the different profiles highlight this push and pull dynamic between the two players in the game. In the licensing case, conditions on the senders are easier to satisfy, since they only need to signal once to be accepted. However, the conditions on the receiver are harder to satisfy as there are more inferior senders in equilibrium. It is easier to fake one signal than to fake two, hence the importance of q_h^I. High costs are diagnostic of an inferior type since both types will pay the low cost, but only the inferior type will pay the high cost under some circumstances rather than always. If they are unlikely, inferior types have more opportunites to pass as superior types undetected. On the other hand, if they are frequent, they will have to reveal themselves. The strain on this probability is greater in licensing, as there are more inferior types in equilibrium. Figure 6.2 clearly shows that in the zero-tolerance equilibria, inferior types are able to pass as superior types only half of the time (2 cost distributions out of 4), while in the licensing equilibria, they are able to do so three quarter of the time (3 cost distributions out of 4).

For these reasons, the rarity of the superior type and the likelihood of high costs are the key parameters distinguishing the two behavioural effects. Senders are more likely to act consistently when receivers strongly want to avoid inferior receivers, either because b^I is very low or because p^S is small. When p^S is 1, the condition on the receiver is easy to satisfy. Things get trickier when the superior type is rare. The theory relies on extremely rare perfect senders as the superior type and assumes that most people are imposters trying to appear as perfect. When we think about ourselves, this makes sense. Even though many people believe in principles and claim to follow them, and to some extent they do, in reality principles are often betrayed [297]. However, for licensing, the superior type is now the more common, normal individual. It represents individuals who simply do not want to be categorised with a bad reputation, and doing the bare minimum to exculpate themselves. Here, the inferior type does not bother to license when the cost is too high and accept their bad reputation.

The next step would be to test these predictions experimentally. To do this, it would be useful to design an experiment that shows complementarity and discontinuity in the behaviour of the sender, and how this in turn affects the level of trust in an observer (the receiver in our model). We have a design with a deception task where players can lie or tell the truth about which bonus bundle is more profitable for a receiver. We vary the cost of telling the truth by implementing two levels of temptation, one high and one low. This allows us to test whether or not a particular sequence of high and low costs leads to consistency or not. We test complementarity by showing that senders are more likely to cooperate/not deceive in the current round if they have cooperated before, even if the cost is high. The probability of paying the high cost is higher in the second round than the first round, provided that the sender paid the cost in the first round. Since we condition our effect on the cost of the previous round, this means that we are looking at the effect of an effect. So a significant but small effect in the first round may not carry over to the second round, making our design difficult to calibrate. It is important that we look at previous costs, not just previous action, to ensure that we are not simply comparing with participants who are naturally more cooperative.

Our model shows that acting consistent when potential partners have high standards is the best course of action, while licensing is sufficient in situations where expectations are lower, highlighting the importance of trust and partner choice. Evolutionary game theory is a powerful tool for understanding the ultimate causes of behaviour. Using both models and experiments, we can get to the roots of a behaviour and make testable predictions about its ultimate function.

Chapter 7

Discussion

Contents

The title of my thesis “Mechanisms and Benefits of Reciprocal Relationships”, was intended to highlight how central cooperation is to our social lives, even in areas where we might least suspect it. The first part of the thesis focused on the mechanisms of cooperation, most notably direct reciprocity, while the second part used the benefit of such cooperative relationships to explain other kinds of social behaviour. Evolution and game theory are the binding agent of these projects: it is in the evolutionary simulations, the implementation of experimental games, or in the use of a signalling model to provide ultimate explanations for observed behaviour.

In Chapter 2, I reviewed of the theoretical and experimental literature surrounding direct reciprocity, focusing in particular on the repeated prisoner’s dilemma game. This chapter highlighted the importance of contrasting results from theoretical analyses and experimental methods to gain insights into the mechanisms of direct reciprocity and identified where potential future research would be fruitful. The next chapter explored reciprocity in concurrent interactions, acknowledging the complexity of human social life and the limitations of studying isolated interactions. The results show mixed effects on cooperation, with concurrent games suggesting a negative effect, while emphasising the complex interplay of strategic motives and spillovers. Chapter 4 was a collaboration where my skills as an experimentalist were needed to provide evidence for the validity of a new reciprocal strategy of cumulative reciprocity. Unlike traditional strategies, CURE relies on a tally of interaction imbalances rather than exact memories of previous rounds, providing a cognitively realistic account of human reciprocity in everyday life. Chapter 5 reviews how different cues signal different levels of cooperativeness and highlights the reliability of past behaviour in predicting cooperation. Finally, Chapter 6 provides a proof of concept for how principled behaviour can be interpreted by others as a signal of trustworthiness and cooperativeness.

Through this array of projects, I have learned many skills and contributed to varied aspects of reciprocal relationships. In this discussion, I will highlight where I want to take my research and the lessons I learned along my Doctoral journey.

7.1 Future avenues for direct reciprocity

I am particularly interested in the extension of work on reciprocity along multiple social ties. People are embedded in highly dense social networks of simultaneous interactions, yet very little work has addressed this question beyond fixed games in fixed networks. Our project on concurrent games took a different approach to study multiple social interactions by restricting the network to two independent interactions. Comparing two interactions with the same partner or with two different partners are essentially two different dimensions of a network [298]. Our paper examines two things: whether playing two different games with the same person leads to strategy linkage, and whether reciprocity spills over from one independent interaction to another. Looking at a smaller set of linked interactions allows us to answer new questions about the mechanisms of reciprocity in such environments.

In the case of multiple interactions with different partners, it would be interesting to consider scenarios where resources need to be distributed along these multiple social ties, especially when resources are unequal or limited [65, 299]. In such a situation, should efforts be divided equally or unequally? And should contributions be based on past contributions and remain purely reciprocal, or should they also take into account the individual characteristics of others? Is cooperation higher or similar in heterogenous networks? I plan to address these questions with Maria Kleshnina, a collaborator who is mainly working on the modelling part of this question. We want to design and run an experiment where participants interact in multiple unequal interactions with a limited common fund. The results could shed light on how heterogenous individuals reciprocate in networks of unequal partners.

The effect of inequality can also be studied in our same-partner design, where participants interact with the same partner in two different interactions. One possible project would vary the costs and benefits of the two repeated prisoner's dilemma games as in the original project, but in such a way that the costlier game is different for each player. This would create a coordination problem as each player would prefer to prioritise a different game, but still benefit from their partner's cooperation. Another interesting design is to embed players in multiple interactions across two dimensions: a different game played with the same partner and the same game with a different partner. Su et al. have already developed a model in a network using this design [298]. They also apply their model to existing multi-layer social networks from a variety of social settings. Analysing reciprocity in a real network would be an exciting collaboration with anthropologists or sociologists.

Lesson 1: Often it doesn't work. Only one of my experiments worked on the first try and ran smoothly. The others took months to pilot. For some I have had to try a different game, for others I worked on an idea for 8 months only to realise that it doesn't work and try something else. Effects that are known to work fail. Two of my projects rely on different stake sizes as a mechanism to influence the likelihood of cooperation. Participants from different subject pools were not affected by the difference, even when the difference was 10 times larger. Working with human subjects can be very messy, but not unpredictable. Overall, I am happy that most projects have borne fruits. Most importantly, I know that I need to be more patient and have more realistic expectations of what can be achieved in a given time frame [1].

7.2 Mixing methods

All of my projects relied on the use of both computational methods and experimental approaches. Most of my previous training had been in experimental methods and I wanted to learn more about modelling, in which I had only minimal training. Game theory and its biological application, evolutionary game theory, are primarily mathematical theories. If I wanted to

[1] I started my first project thinking it would be drafted after 6 months.

understand, test and learn them, I had to use them. Combining both methods was a great way to learn new skills and perfect old ones.

In a sense, models are experiments in themselves; they seek to understand the dynamics and causal relationships of important social behaviours. For a model to arrive at these answers, it must be fundamentally simple. The price of such simplicity is external validity, as models can sometimes be poor predictors of behaviour in the wild. We have seen that the most common strategies in the prisoner's dilemma do not explain human behaviour in these games very well. Nevertheless, models remain extremely valuable in providing proofs of concept, showing if a strategy is even feasible in the first place. Models show in numerous ways how reciprocity can evolve, and experiments do show that people are mostly reciprocal, even if an exact strategy is hard to pin down. Behaviour can have many different explanations, and some are difficult or even impossible to collect data on. Models can help us work out when a theory is internally consistent, and formalise arguments to find which answers are plausible. The simplicity in turn helps us understand the consequences of a few key assumptions by breaking down complex processes into clear and understandable concepts.

I have often heard my fellow theorists complain that "verbal" theories as opposed to mathematical ones lack precision. Words are too open to misinterpretation, and too many debates boils down to defining things properly. Mathematics cannot be misinterpreted. I did not fully grasp what their point was exactly, until I developed a model myself. The epiphany came when I realised that there can be two ways of defining a strategy for a player that lead to the same behaviour in equilibrium, but different responses in out-of-equilibrium scenarios. The distinction was never made clear to me by mere words, it is only when I did the mathematical analysis that I understood it.

Nevertheless, it was challenging to accept that effects known to be critical in real life often have to be excluded. After all, that is exactly what we experimentalists when we deliberately control for confounding variables! Ultimately, all scientists in the field are trying to explain behaviour, and theorists are just using a more abstract and inexpensive method. But experiments remain fundamental: they inspire and challenge theories. Once models have made predictions, they need be tested. Predictions are usually only qualitative: they indicate the direction of change, rather than the precise magnitude of the effect. We know that increasing the number of players in the group will eventually leads to a breakdown in cooperation, but we do not predict exactly how much cooperation will be lost by increasing the number of group members. Sometimes models are so good that they can predict precise numbers, as in case of the famous early study of sex allocation [56]. Whether qualitative or quantitative, model predictions should lead to data collection, which will often result in some model revision, and leads to more data collection. The cycle repeats as we gain more knowledge about the system and can start considering more complex scenarios.

Lesson 2: Sometimes it hurts the brain. Learning to model with game theory was a real exercise in abstraction. To quote the preface of the book *Mathematical Models of Social Evolution: A Guide for the Perplexed* by Robert Boyd and Richard McElreath [300]: "[My] algebra skills [had] atrophied from disuse, and even factoring a polynomial [was] only an ancient memory, as if from a past life." But it was not the algebra itself that was the hardest, it was the ability to read mathematics. To look at an equation and understand what it means. I used to be amazed when my colleagues could immediately see how the system would behave if some parameter was high or low, while I had to calculate almost every term in my head to come to the same conclusion 10 minutes later. It still takes me some time, but I have come to accept that the most valuable to me is the ability to interpret models to understand their implications and how they can be tested empirically. But the satisfaction of reading an equation and truly understanding what it means is a feeling I will cherish for a long time.

7.3 An evolutionary lens

In writing this thesis, I dug up some old lecture material from my psychology undergraduate and was amused to see some now familiar names on my old reading list. [2] In particular, one of my favourite class was called "Altruism, Cooperation, and Helping" and was my first introduction to the study of human behaviour in light of the theory of evolution. The evolutionary part was not what stuck with me in those days, it was the use of economic games to study human behaviour. It was not until I was deep into my Masters in Behavioural and Economic Science that I realised the breadth of game theory and the methods economists used to answer similar questions to psychologists. That is when I knew I wanted to work on social behaviour using game theory. It is what led me to the South African bush to assist graduate students and postdocs studying the social behaviour of vervet monkeys. It is what led me to apply for a PhD position in the theory department of a biology institute.

The most important lesson I have learnt from interacting with evolutionary scientists is that questions have multiple levels of explanation, and that evolution is really a mechanism for only one of them. I used to say that my degree in psychology made me lose faith in humanity – I had seen so much evidence of how selfish, stupid, and manipulable we can be. In particular, I did not believe in *true* altruism, and was always vehemently oppposed the concept of "social preferences". It seemed obvious to me that altruism always has a hidden benefit somewhere: to make you feel good about yourself or to give you a good reputation. And the word "preference" made it sound as if the individual choosing a fair split in an ultimatum game did so out of the goodness of their heart, rather than out of an understanding of the punitive power of the other player. But here I learned that cooperation is everywhere. So much so that the puzzle of cooperation has kept scientists from many different fields busy for decades. We are not selfish, stupid or manipulable. All these proximate effects (mostly) have a good reason for existing. Psychologists and

[2] I find it particularly funny that in those days I thought that this list was incredibly long. I have now read most of them and many more, and even know some authors personally.

economists have just not looked at the ultimate cause. Now I understand that even if there is no such thing as pure altruism, it doesn't mean that people don't feel it. Social preferences are just the proximate mechanisms, something to add to our utility.

Lesson 3: Dig deeper. Psychology research is full of quirky effects and behaviour. As interesting and important are those first observations and experiments, they often beg the question: but why? Like a child who asks questions after questions, human social behaviour can only really be understood if we investigate the underlying incentives and dynamics at play. We are a social species, as such non of our behaviour happens in a vaccum. Many things (everything?) comes down to being a good partner.

Behavioural scientists in economics and psychology are relatively unfamiliar with the concept of ultimate functions. However, the accomplished evolutionary game theorists Nowak and Sigmund [62] say it themselves: the main themes of the research carried in these fields are about cooperation and communication between individuals, and are therefore inherently game theoretic. Evolutionary psychology and anthropology, fields that are very familiar with these concepts and that look at behaviour with a functional analysis in mind, lack game theory. Economists who are familiar with game theory, lack often lack an evolutionary lens in their research. My hope for the field of human behaviour is for researchers from different disciplines to learn the methods and theories of each other. Economics, psychology, anthropology and biology have so much in common, and yet some key discoveries take decades to reach the researchers who would benefit from them. Some young scientists are crossing the bridge with new models of human social behaviour, combining evolutionary explanations with existing models of behaviour [76, 301, 302]. I would particularly like to see psychology embrace game theory models and not just economic games. Now that I understand how useful mathematical models can be to characterise a theory, I am certain that psychologist should learn to create and interpret their own.

Lesson 4: Cooperation is truly everywhere. The diversity of researchers I have had the privilege of interacting with over the course of this PhD is astounding. Coming from a background in psychology and economics, I already knew that the field of cooperation was interdisciplinary. But the overarching field of behaviour and evolution is immense and much broader than I had realised. I have met some amazing researchers from other fields, sometimes leading to current and future projects with them. It is fascinating how differently other fields approach the same questions. I have shown how theorists and empiricists complement each other, and I am proud to now be an accomplished translator of mathematical models able to transform their insights into testable experimental designs. Yet there are always more techniques and approaches to explore and learn form. The vastness of the field is dizzying, and I wish I had more time to read from all every corner.

Bibliography

[1] N. Raihani, *The Social Instinct: How Cooperation Shaped the World.* Random House, 2021.

[2] S. T. Powers, C. P. Van Schaik, and L. Lehmann, "Cooperation in large-scale human societies — What, if anything, makes it unique, and how did it evolve?," *Evolutionary Anthropology: Issues, News, and Reviews*, vol. 30, no. 4, pp. 280–293, 2021.

[3] M. R. Frank, N. Obradovich, L. Sun, W. L. Woon, B. L. LeVeck, and I. Rahwan, "Detecting reciprocity at a global scale," *Science Advances*, vol. 4, no. 1, p. eaao5348, 2018.

[4] H. S. Kaplan, P. L. Hooper, and M. Gurven, "The evolutionary and ecological roots of human social organization," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1533, pp. 3289–3299, 2009.

[5] C. L. Apicella and J. B. Silk, "The evolution of human cooperation," *Current Biology*, vol. 29, no. 11, pp. R447–R450, 2019.

[6] U. Fischbacher, S. Gachter, and E. Fehr, "Are people conditionally cooperative? Evidence from a public goods experiment," *Economics Letters*, 2001.

[7] E. Fehr and U. Fischbacher, "The nature of human altruism," *Nature*, vol. 425, no. 6960, pp. 785–791, 2003.

[8] P. Hammerstein, *Genetic and Cultural Evolution of Cooperation.* MIT Press, 2003.

[9] R. Axelrod, *The Evolution of Cooperation: Revised Edition.* New York, NY: Basic Books, revised edition ed., 2006.

[10] E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press, 1990.

[11] P. A. Van Lange, J. Joireman, C. D. Parks, and E. Van Dijk, "The psychology of social dilemmas: A review," *Organizational Behavior and Human Decision Processes*, vol. 120, no. 2, pp. 125–141, 2013.

[12] A. M. Colman, "The puzzle of cooperation," *Nature*, vol. 440, no. 7085, pp. 744–745, 2006.

[13] A. Diekmann and S. Lindenberg, "Cooperation: Sociological Aspects," in *International Encyclopedia of the Social & Behavioral Sciences*, pp. 862–866, Elsevier, 2015.

[14] M. A. Nowak, “Five Rules for the Evolution of Cooperation,” *Science*, vol. 314, no. 5805, pp. 1560–1563, 2006.

[15] W. Hamilton, “The genetical evolution of social behaviour ll.,” *Journal of Theoretical Biology*, vol. 7, no. 1, pp. 17–52, 1964.

[16] R. L. Trivers, “The Evolution of Reciprocal Altruism,” *The Quarterly Review of Biology*, vol. 46, no. 1, pp. 35–57, 1971.

[17] E. Fehr and S. Gächter, “Altruistic punishment in humans,” *Nature*, vol. 415, no. 6868, pp. 137–140, 2002.

[18] N. J. Raihani and R. Bshary, “Punishment: One tool, many uses,” *Evolutionary Human Sciences*, vol. 1, p. e12, 2019.

[19] M. Milinski, “Reputation, a universal currency for human social interactions,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1687, p. 20150100, 2016.

[20] G. Roberts, N. Raihani, R. Bshary, H. M. Manrique, A. Farina, F. Samu, and P. Barclay, “The benefits of being seen to help others: Indirect reciprocity and reputation-based partner choice,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, no. 1838, p. 20200290, 2021.

[21] P. Barclay, “Reputational benefits for altruistic punishment,” *Evolution and Human Behavior*, vol. 27, no. 5, pp. 325–344, 2006.

[22] M. dos Santos, D. J. Rankin, and C. Wedekind, “The evolution of punishment through reputation,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 278, no. 1704, pp. 371–377, 2010.

[23] C. Hilbe and A. Traulsen, “Emergence of responsible sanctions without second order free riders, antisocial punishment or spite,” *Scientific Reports*, vol. 2, no. 1, p. 458, 2012.

[24] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior, 2nd Rev. Ed.* Theory of Games and Economic Behavior, 2nd Rev. Ed, Princeton, NJ, US: Princeton University Press, 1947.

[25] D. Fudenberg and J. Tirole, *Game Theory.* MIT Press, 1991.

[26] A. Rapoport and A. M. Chammah, *Prisoner’s Dilemma: A Study in Conflict and Cooperation.* University of Michigan Press, 1965.

[27] P. Dal Bó and G. R. Fréchette, “On the Determinants of Cooperation in Infinitely Repeated Games: A Survey,” *Journal of Economic Literature*, vol. 56, no. 1, pp. 60–114, 2018.

[28] R. Boyd, “Is the repeated prisoner’s dilemma a good model of reciprocal altruism?,” *Ethology and Sociobiology*, vol. 9, no. 2-4, pp. 211–222, 1988.

[29] N. J. Raihani and R. Bshary, "Resolving the iterated prisoner's dilemma: Theory and reality: Resolvings the IPD," *Journal of Evolutionary Biology*, vol. 24, no. 8, pp. 1628–1639, 2011.

[30] J. F. Nash, "Equilibrium Points in n-Person Games," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 36, no. 1, pp. 48–49, 1950.

[31] I. Thielmann, R. Böhm, M. Ott, and B. E. Hilbig, "Economic Games: An Introduction and Guide for Research," *Collabra: Psychology*, vol. 7, no. 1, p. 19004, 2021.

[32] J. K. Goeree and C. A. Holt, "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions," *American Economic Review*, vol. 91, no. 5, pp. 1402–1422, 2001.

[33] J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, N. S. Henrich, K. Hill, F. Gil-White, M. Gurven, F. W. Marlowe, J. Q. Patton, and D. Tracer, "“Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies," *Behavioral and Brain Sciences*, vol. 28, no. 6, pp. 795–815, 2005.

[34] D. Kahneman and A. Tversky, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, vol. 47, no. 2, 1979.

[35] R. H. Thaler, "Anomalies: The January Effect," *Journal of Economic Perspectives*, vol. 1, no. 1, pp. 197–201, 1987.

[36] M. Hoffman and E. Yoeli, *Hidden Games: The Surprising Power of Game Theory to Explain Irrational Human Behavior.* New York: Basic Books, first edition ed., 2022.

[37] L. Bruni and R. Sugden, "The Road not Taken: How Psychology was Removed from Economics, and How it Might be Brought Back," *The Economic Journal*, vol. 117, no. 516, pp. 146–173, 2007.

[38] R. H. Thaler, *Misbehaving: The Making of Behavioral Economics.* Misbehaving: The Making of Behavioral Economics, New York, NY, US: W W Norton & Co, 2015.

[39] E. Hoffman, K. A. Mccabe, and V. L. Smith, "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry*, vol. 36, no. 3, pp. 335–352, 1998.

[40] M. K. Schweinfurth and M. Taborsky, "Reciprocal Trading of Different Commodities in Norway Rats," *Current Biology*, vol. 28, no. 4, pp. 594–599.e3, 2018.

[41] G. S. Wilkinson, "Reciprocal food sharing in the vampire bat," *Nature*, vol. 308, no. 5955, pp. 181–184, 1984.

[42] M. Milinski, "TIT FOR TAT in sticklebacks and the evolution of cooperation," *Nature*, vol. 325, no. 6103, pp. 433–435, 1987.

[43] T. Clutton-Brock, "Cooperation between non-kin in animal societies," *Nature*, vol. 462, no. 7269, pp. 51–57, 2009.

[44] R. J. Aumann, "Backward Induction and Common Knowledge of Rationality," *Games and Economic Behavior*, vol. 8, 1995.

[45] D. Fudenberg and E. Maskin, "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica*, vol. 54, no. 3, pp. 533–554, 1986.

[46] P. Dal Bó, "Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games," *THE AMERICAN ECONOMIC REVIEW*, vol. 95, no. 5, 2005.

[47] P. R. Blake, D. G. Rand, D. Tingley, and F. Warneken, "The shadow of the future promotes cooperation in a repeated prisoner's dilemma for children," *Scientific Reports*, vol. 5, no. 1, p. 14559, 2015.

[48] S. K. Baek, H.-C. Jeong, C. Hilbe, and M. A. Nowak, "Comparing reactive and memory-one strategies of direct reciprocity," *Scientific Reports*, vol. 6, no. 1, p. 25676, 2016.

[49] N. E. Glynatsi and V. A. Knight, "Using a theory of mind to find best responses to memory-one strategies," *Scientific Reports*, vol. 10, no. 1, p. 17287, 2020.

[50] R. Axelrod and W. D. Hamilton, "The Evolution of Cooperation," *Science*, vol. 211, no. 4489, pp. 1390–1396, 1981.

[51] L. A. Martinez-Vaquero, J. A. Cuesta, and A. Sánchez, "Generosity Pays in the Presence of Direct Reciprocity: A Comprehensive Study of 2×2 Repeated Games," *PLOS ONE*, vol. 7, no. 4, p. e35135, 2012.

[52] J. M. Smith, *Evolution & the Theory of Games*. Cambridge New York Port-Chester[etc]: Cambridge university press, 1989.

[53] J. M. McNamara and O. Leimar, *Game Theory in Biology: Concepts and Frontiers*. Oxford University Press, 2020.

[54] J. M. Smith and G. R. Price, "The Logic of Animal Conflict," *Nature*, vol. 246, no. 5427, pp. 15–18, 1973.

[55] N. Davies, "Territorial defence in the speckled wood butterfly (Pararge aegeria): The resident always wins," *Animal Behaviour*, vol. 26, pp. 138–147, 1978.

[56] S. R. A. Fisher, *The Genetical Theory of Natural Selection: A Complete Variorum Edition*. OUP Oxford, 1930.

[57] N. B. Davies, "Games Animals Play," in *Games: Conflict, Competition, and Cooperation* (D. Blagden and M. de Rond, eds.), Darwin College Lectures, pp. 120–144, Cambridge: Cambridge University Press, 2018.

[58] H. Kokko, "Dyadic contests: Modelling fights between two individuals," in *Animal Contests* (I. C. W. Hardy and M. Briffa, eds.), pp. 5–32, Cambridge University Press, 1 ed., 2013.

[59] D. S. Falster and M. Westoby, "Plant height and evolutionary games," *Trends in Ecology & Evolution*, vol. 18, no. 7, pp. 337–343, 2003.

[60] K. Sigmund, *The Calculus of Selfishness:*. Princeton University Press, 2010.

[61] P. Hammerstein and R. Selten, "Chapter 28 Game theory and evolutionary biology," in *Handbook of Game Theory with Economic Applications*, vol. 2, pp. 929–993, Elsevier, 1994.

[62] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg, "Emergence of cooperation and evolutionary stability in finite populations," *Nature*, vol. 428, no. 6983, pp. 646–650, 2004.

[63] J. Grujić, B. Eke, A. Cabrales, J. A. Cuesta, and A. Sánchez, "Three is a crowd in iterated prisoner's dilemmas: Experimental evidence on reciprocal behavior," *Scientific Reports*, vol. 2, no. 1, p. 638, 2012.

[64] M. C. Couto, S. Giaimo, and C. Hilbe, "Introspection dynamics: A simple model of counterfactual learning in asymmetric games," *New Journal of Physics*, vol. 24, no. 6, p. 063010, 2022.

[65] X. Wang, M. C. Couto, N. Wang, X. An, B. Chen, Y. Dong, C. Hilbe, and B. Zhang, "Cooperation and coordination in heterogeneous populations," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 378, no. 1876, p. 20210504, 2023.

[66] S. F. Brosnan and F. B. M. de Waal, "A proximate perspective on reciprocal altruism," *Human Nature*, vol. 13, no. 1, pp. 129–152, 2002.

[67] G. Schino and F. Aureli, "Grooming reciprocation among female primates: A meta-analysis," *Biology Letters*, vol. 4, no. 1, pp. 9–11, 2008.

[68] J. B. Silk, "Cooperation without counting," in *Genetic and Cultural Evolution of Cooperation*, pp. 37–54, MIT Press Cambridge, UK, 2003.

[69] P. Todd, J. Ortega, J. Davis, G. Gigerenzer, D. Goldstein, A. Goodie, R. Hertwig, U. Hoffrage, K. Laskey, L. Martignon, and G. Miller, *Simple Heuristics That Make Us Smart.* Oxford University Press, USA, 1999.

[70] G. Gigerenzer and R. Selten, eds., *Bounded Rationality: The Adaptive Toolbox.* Dahlem Workshop Reports, Cambridge, Mass.: MIT Press, 1. mit press paperback ed ed., 2002.

[71] E. Fehr and K. M. Schmidt, "A Theory of Fairness, Competition, and Cooperation*," *The Quarterly Journal of Economics*, vol. 114, no. 3, pp. 817–868, 1999.

[72] U. Fischbacher and S. Gächter, "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments," *American Economic Review*, vol. 100, no. 1, pp. 541–556, 2010.

[73] J. Castro Santa, F. Exadaktylos, and S. Soto-Faraco, “Beliefs about others’ intentions determine whether cooperation is the faster choice,” *Scientific Reports*, vol. 8, no. 1, p. 7509, 2018.

[74] M. Aoyagi, G. R. Fréchette, and S. Yuksel, “Beliefs in repeated games,” *ISER DP*, no. 1119RR, 2022.

[75] N. Tinbergen, “On aims and methods of Ethology,” *Zeitschrift für Tierpsychologie*, vol. 20, no. 4, pp. 410–433, 1963.

[76] L. Page, *Optimally Irrational: The Good Reasons We Behave the Way We Do.* Cambridge: Cambridge University Press, 2022.

[77] J. H. Barkow, L. Cosmides, and J. Tooby, *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* Oxford University Press, 1995.

[78] J. Henrich and R. McElreath, “The evolution of cultural evolution,” *Evolutionary Anthropology: Issues, News, and Reviews*, vol. 12, no. 3, pp. 123–135, 2003.

[79] R. Noë and P. Hammerstein, “Biological markets: Supply and demand determine the effect of partner choice in cooperation, mutualism and mating,” *Behavioral Ecology and Sociobiology*, vol. 35, no. 1, pp. 1–11, 1994.

[80] P. Barclay, “Strategies for cooperation in biological markets, especially for humans,” *Evolution and Human Behavior*, vol. 34, no. 3, pp. 164–175, 2013.

[81] K. Sylwester and G. Roberts, “Cooperators benefit through reputation-based partner choice in economic games,” *Biology Letters*, vol. 6, no. 5, pp. 659–662, 2010.

[82] P. Barclay, “Harnessing the Power of Reputation: Strengths and Limits for Promoting Cooperative Behaviors,” *Evolutionary Psychology*, vol. 10, no. 5, p. 147470491201000509, 2012.

[83] M. Milinski, D. Semmann, T. C. M. Bakker, and H.-J. Krambeck, “Cooperation through indirect reciprocity: Image scoring or standing strategy?,” *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1484, pp. 2495–2501, 2001.

[84] D. Semmann, H.-J. Krambeck, and M. Milinski, “Strategic investment in reputation,” *Behavioral Ecology and Sociobiology*, vol. 56, no. 3, pp. 248–252, 2004.

[85] C. S. L. Rossetti and C. Hilbe, “Direct reciprocity among humans,” *Ethology*, p. eth.13407, 2023.

[86] C. S. Rossetti, C. Hilbe, and O. P. Hauser, “(Mis)perceiving cooperativeness,” *Current Opinion in Psychology*, vol. 43, pp. 151–155, 2022.

[87] M. Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups, With a New Preface and Appendix.* Harvard University Press, 1971.

[88] G. Hardin, “The Tragedy of the Commons,” *Science*, vol. 162, no. 3859, pp. 1243–1248, 1968.

[89] N. E. Glynatsi and V. A. Knight, "A bibliometric study of research topics, collaboration, and centrality in the iterated prisoner's dilemma," *Humanities and Social Sciences Communications*, vol. 8, no. 1, pp. 1–12, 2021.

[90] C. Hilbe, Š. Šimsa, K. Chatterjee, and M. A. Nowak, "Evolution of cooperation in stochastic games," *Nature*, vol. 559, no. 7713, pp. 246–249, 2018.

[91] M. Wörle and M. Paulus, "Normative foundations of reciprocity in preschoolers," *Journal of Experimental Child Psychology*, vol. 188, p. 104693, 2019.

[92] J. Grujić, C. Gracia-Lázaro, M. Milinski, D. Semmann, A. Traulsen, J. A. Cuesta, Y. Moreno, and A. Sánchez, "A comparative analysis of spatial Prisoner's Dilemma experiments: Conditional cooperation and payoff irrelevance," *Scientific Reports*, vol. 4, no. 1, p. 4615, 2014.

[93] J. W. Crandall, M. Oudah, Tennom, F. Ishowo-Oloko, S. Abdallah, J.-F. Bonnefon, M. Cebrian, A. Shariff, M. A. Goodrich, and I. Rahwan, "Cooperating with machines," *Nature Communications*, vol. 9, no. 1, p. 233, 2018.

[94] C. Hilbe, B. Wu, A. Traulsen, and M. A. Nowak, "Cooperation and control in multiplayer social dilemmas," *Proceedings of the National Academy of Sciences*, vol. 111, no. 46, pp. 16425–16430, 2014.

[95] Z. Wang, Y. Zhou, J. W. Lien, J. Zheng, and B. Xu, "Extortion can outperform generosity in the iterated prisoner's dilemma," *Nature Communications*, vol. 7, no. 1, p. 11125, 2016.

[96] C. Hauert and H. G. Schuster, "Effects of increasing the number of players and memory size in the iterated Prisoner's Dilemma: A numerical approach," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 264, no. 1381, pp. 513–519, 1997.

[97] C. Hilbe, L. A. Martinez-Vaquero, K. Chatterjee, and M. A. Nowak, "Memory-n strategies of direct reciprocity," *Proceedings of the National Academy of Sciences*, vol. 114, no. 18, pp. 4715–4720, 2017.

[98] Y. Murase and S. K. Baek, "Five rules for friendly rivalry in direct reciprocity," *Scientific Reports*, vol. 10, no. 1, p. 16904, 2020.

[99] M. Nowak and K. Sigmund, "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game," *Nature*, vol. 364, no. 6432, pp. 56–58, 1993.

[100] M. van Veelen, J. García, D. G. Rand, and M. A. Nowak, "Direct reciprocity in structured populations," *Proceedings of the National Academy of Sciences*, vol. 109, no. 25, pp. 9929–9934, 2012.

[101] R. Selten, "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 1975.

[102] J. R. Stevens, J. Volstorf, L. J. Schooler, and J. Rieskamp, "Forgetting Constrains the Emergence of Cooperative Decision Strategies," *Frontiers in Psychology*, vol. 1, 2011.

[103] M. J. A. Tazelaar, P. A. M. Van Lange, and J. W. Ouwerkerk, “How to Cope With "Noise" in Social Dilemmas: The Benefits of Communication,” *Journal of Personality and Social Psychology*, vol. 87, no. 6, pp. 845–859, 2004.

[104] P. A. M. Van Lange, J. W. Ouwerkerk, and M. J. A. Tazelaar, “How to overcome the detrimental effects of noise in social interaction: The benefits of generosity,” *Journal of Personality and Social Psychology*, vol. 82, no. 5, pp. 768–780, 2002.

[105] D. M. Kreps, P. Milgrom, J. Roberts, and R. Wilson, “Rational cooperation in the finitely repeated prisoners’ dilemma,” *Journal of Economic Theory*, vol. 27, no. 2, pp. 245–252, 1982.

[106] J. M. McNamara, Z. Barta, and A. I. Houston, “Variation in behaviour promotes cooperation in the Prisoner’s Dilemma game,” *Nature*, vol. 428, no. 6984, pp. 745–748, 2004.

[107] E. Akin, “The Iterated Prisoner’s Dilemma: Good Strategies and Their Dynamics,” in *Ergodic Theory: Advances in Dynamical Systems*, Walter de Gruyter GmbH & Co KG, 2017.

[108] A. J. Stewart and J. B. Plotkin, “The Evolvability of Cooperation under Local and Non-Local Mutations,” *Games*, vol. 6, no. 3, pp. 231–250, 2015.

[109] C. Hilbe, A. Traulsen, and K. Sigmund, “Partners or rivals? Strategies for the iterated prisoner’s dilemma,” *Games and Economic Behavior*, vol. 92, pp. 41–52, 2015.

[110] H. Zhang, “Errors can increase cooperation in finite populations,” *Games and Economic Behavior*, vol. 107, pp. 203–219, 2018.

[111] L. Schmid, K. Chatterjee, C. Hilbe, and M. A. Nowak, “A unified framework of direct and indirect reciprocity,” *Nature Human Behaviour*, vol. 5, no. 10, pp. 1292–1302, 2021.

[112] W. H. Press and F. J. Dyson, “Iterated Prisoner’s Dilemma contains strategies that dominate any evolutionary opponent,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 26, pp. 10409–10413, 2012.

[113] G. J. Mailath and L. Samuelson, *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, USA, 2006.

[114] V. Knight, O. Campbell, M. Harper, K. Langner, J. Campbell, T. Campbell, A. Carney, M. Chorley, C. Davidson-Pilon, K. Glass, N. Glynatsi, T. Ehrlich, M. Jones, G. Koutsovoulos, H. Tibble, J. Müller, G. Palmer, P. Petunov, P. Slavin, T. Standen, L. Visintini, and K. Molden, “An Open Framework for the Reproducible Study of the Iterated Prisoner’s Dilemma,” *Journal of Open Research Software*, vol. 4, no. 1, p. 35, 2016.

[115] A. Rapoport, D. A. Seale, and A. M. Colman, “Is Tit-for-Tat the Answer? On the Conclusions Drawn from Axelrod’s Tournaments,” *PLOS ONE*, vol. 10, no. 7, p. e0134128, 2015.

[116] A. Szolnoki and M. Perc, “Defection and extortion as unexpected catalysts of unconditional cooperation in structured populations,” *Scientific Reports*, vol. 4, no. 1, p. 5496, 2014.

[117] A. J. Stewart and J. B. Plotkin, "Collapse of cooperation in evolving games," *Proceedings of the National Academy of Sciences*, vol. 111, no. 49, pp. 17558–17563, 2014.

[118] J. Tkadlec, C. Hilbe, and M. A. Nowak, "Mutation enhances cooperation in direct reciprocity," *Proceedings of the National Academy of Sciences*, vol. 120, no. 20, p. e2221080120, 2023.

[119] J. J. Horton, D. G. Rand, and R. J. Zeckhauser, "The online laboratory: Conducting experiments in a real labor market," *Experimental Economics*, vol. 14, no. 3, pp. 399–425, 2011.

[120] R. Cooper, D. V. DeJong, R. Forsythe, and T. W. Ross, "Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games," *Games and Economic Behavior*, vol. 12, no. 2, pp. 187–218, 1996.

[121] M. Embrey, G. R. Fréchette, and S. Yuksel, "Cooperation in the Finitely Repeated Prisoner's Dilemma*," *The Quarterly Journal of Economics*, vol. 133, no. 1, pp. 509–551, 2018.

[122] A. Mao, L. Dworkin, S. Suri, and D. J. Watts, "Resilient cooperators stabilize long-run cooperation in the finitely repeated Prisoner's Dilemma," *Nature Communications*, vol. 8, no. 1, p. 13800, 2017.

[123] J. Andreoni and J. H. Miller, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *The Economic Journal*, vol. 103, no. 418, p. 570, 1993.

[124] A. E. Roth and J. Murnighan, "Equilibrium behavior and repeated play of the prisoner's dilemma," *Journal of Mathematical Psychology*, vol. 17, no. 2, pp. 189–198, 1978.

[125] M. Blonski, P. Ockenfels, and G. Spagnolo, "Equilibrium Selection in the Repeated Prisoner's Dilemma: Axiomatic Approach and Experimental Evidence," *American Economic Journal: Microeconomics*, vol. 3, no. 3, pp. 164–192, 2011.

[126] P. Dal Bó and G. R. Fréchette, "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence," *American Economic Review*, vol. 101, no. 1, pp. 411–429, 2011.

[127] D. Gill and Y. Rosokha, "Beliefs, Learning, and Personality in the Indefinitely Repeated Prisoner's Dilemma," 2022.

[128] P. Dal Bó and G. R. Fréchette, "Strategy Choice in the Infinitely Repeated Prisoner's Dilemma," *American Economic Review*, vol. 109, no. 11, pp. 3929–3952, 2019.

[129] W. Raub and C. Snijders, "Gains, losses, and cooperation in social dilemmas and collective action: The effects of risk preferences," *The Journal of Mathematical Sociology*, vol. 22, no. 3, pp. 263–302, 1997.

[130] M. Aoyagi and G. R. Fréchette, "Collusion in Repeated Games with Imperfect Public Monitoring," 2003.

[131] D. Fudenberg, D. G. Rand, and A. Dreber, “Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World,” *American Economic Review*, vol. 102, no. 2, pp. 720–749, 2012.

[132] J. Li, X. Zhao, B. Li, C. S. L. Rossetti, C. Hilbe, and H. Xia, “Evolution of cooperation through cumulative reciprocity,” *Nature Computational Science*, vol. 2, no. 10, pp. 677–686, 2022.

[133] R. Selten, “Die strategiemethode zur erforschung des eingeschr nkt rationale verhaltens im rahmen eines oligopolexperiments,” *Beitr ge zur experimentellen Wirtschaftsforschung*, 1967.

[134] D. Friedman and R. Oprea, “A Continuous Dilemma,” *American Economic Review*, vol. 102, no. 1, pp. 337–363, 2012.

[135] M. Milinski and C. Wedekind, “Working memory constrains human cooperation in the Prisoner’s Dilemma,” *Proceedings of the National Academy of Sciences*, vol. 95, no. 23, pp. 13755–13758, 1998.

[136] E. Montero-Porras, J. Grujić, E. Fernández Domingos, and T. Lenaerts, “Inferring strategies from observations in long iterated Prisoner’s dilemma experiments,” *Scientific Reports*, vol. 12, no. 1, p. 7589, 2022.

[137] T. R. Palfrey and H. Rosenthal, “Repeated Play, Cooperation and Coordination: An Experimental Study,” *The Review of Economic Studies*, vol. 61, no. 3, pp. 545–565, 1994.

[138] C. Wedekind and M. Milinski, “Human cooperation in the simultaneous and the alternating Prisoner’s Dilemma: Pavlov versus Generous Tit-for-Tat.,” *Proceedings of the National Academy of Sciences*, vol. 93, no. 7, pp. 2686–2689, 1996.

[139] Y. Breitmoser, “Cooperation, but No Reciprocity: Individual Strategies in the Repeated Prisoner’s Dilemma,” *American Economic Review*, vol. 105, no. 9, pp. 2882–2910, 2015.

[140] T. Winke and J. R. Stevens, “Is Cooperative Memory Special? The Role of Costly Errors, Context, and Social Network Size When Remembering Cooperative Actions,” *Frontiers in Robotics and AI*, vol. 4, p. 52, 2017.

[141] J. Volstorf, J. Rieskamp, and J. R. Stevens, “The Good, the Bad, and the Rare: Memory for Partners in Social Interactions,” *PLoS ONE*, vol. 6, no. 4, p. e18945, 2011.

[142] R. Bell, N. Koranyi, A. Buchner, and K. Rothermund, “The implicit cognition of reciprocal exchange: Automatic retrieval of positive and negative experiences with partners in a prisoner’s dilemma game,” *Cognition and Emotion*, vol. 31, no. 4, pp. 657–670, 2017.

[143] S. Duffy and J. Smith, “Cognitive load in the multi-player prisoner’s dilemma game: Are there brains in games?,” *Journal of Behavioral and Experimental Economics*, vol. 51, pp. 47–56, 2014.

[144] B. Allen, G. Lippner, Y.-T. Chen, B. Fotouhi, N. Momeni, S.-T. Yau, and M. A. Nowak, "Evolutionary dynamics on any population structure," *Nature*, vol. 544, no. 7649, pp. 227–230, 2017.

[145] M. A. Nowak and K. Sigmund, "Tit for tat in heterogeneous populations," *Nature*, vol. 355, no. 6357, pp. 250–253, 1992.

[146] C. Gracia-Lázaro, J. A. Cuesta, A. Sánchez, and Y. Moreno, "Human behavior in Prisoner's Dilemma experiments suppresses network reciprocity," *Scientific Reports*, vol. 2, no. 1, p. 325, 2012.

[147] J. Grujić, C. Fosco, L. Araujo, J. A. Cuesta, and A. Sánchez, "Social Experiments in the Mesoscale: Humans Playing a Spatial Prisoner's Dilemma," *PLoS ONE*, vol. 5, no. 11, p. e13749, 2010.

[148] A. Traulsen, D. Semmann, R. D. Sommerfeld, H.-J. Krambeck, and M. Milinski, "Human strategy updating in evolutionary games," *Proceedings of the National Academy of Sciences*, vol. 107, no. 7, pp. 2962–2966, 2010.

[149] D. G. Rand, M. A. Nowak, J. H. Fowler, and N. A. Christakis, "Static network structure can stabilize human cooperation," *Proceedings of the National Academy of Sciences*, vol. 111, no. 48, pp. 17093–17098, 2014.

[150] I. Martinez-Martinez and H.-T. Normann, "Cooperation in Multiplayer Dilemmas," 2022.

[151] J. Grujić, J. A. Cuesta, and A. Sánchez, "On the coexistence of cooperators, defectors and conditional cooperators in the multiplayer iterated Prisoner's Dilemma," *Journal of Theoretical Biology*, vol. 300, pp. 299–308, 2012.

[152] K. Fehl, D. J. Van Der Post, and D. Semmann, "Co-evolution of behaviour and social network structure promotes human cooperation: Cooperation and self-organization in networks," *Ecology Letters*, vol. 14, no. 6, pp. 546–551, 2011.

[153] D. G. Rand, S. Arbesman, and N. A. Christakis, "Dynamic social networks promote cooperation in experiments with humans," *Proceedings of the National Academy of Sciences*, vol. 108, no. 48, pp. 19193–19198, 2011.

[154] P. Bednarik, K. Fehl, and D. Semmann, "Costs for switching partners reduce network dynamics but not cooperative behaviour," *Proceedings of the Royal Society B: Biological Sciences*, vol. 281, no. 1792, p. 20141661, 2014.

[155] J. Wang, S. Suri, and D. J. Watts, "Cooperation and assortativity with dynamic partner updating," *Proceedings of the National Academy of Sciences*, vol. 109, no. 36, pp. 14363–14368, 2012.

[156] A. Antonioni, M. Tomassini, and A. Sánchez, "Short-Range Mobility and the Evolution of Cooperation: An Experimental Study," *Scientific Reports*, vol. 5, no. 1, p. 10282, 2015.

[157] G. Spadaro, I. Tiddi, S. Columbus, S. Jin, A. ten Teije, and D. Balliet, "The Cooperation Databank: Machine-Readable Science Accelerates Research Synthesis," *Perspectives on Psychological Science*, vol. 17, no. 5, pp. 1472–1489, 2022.

[158] A. Romano, A. S. Saral, and J. Wu, "Direct and indirect reciprocity among individuals and groups," *Current Opinion in Psychology*, vol. 43, pp. 254–259, 2022.

[159] D. Kraines and V. Kraines, "Learning to cooperate with Pavlov an adaptive strategy for the iterated Prisoner's Dilemma with noise," *Theory and Decision*, vol. 35, no. 2, pp. 107–150, 1993.

[160] B. Wu, C. S. Gokhale, L. Wang, and A. Traulsen, "How small are small mutation rates?," *Journal of Mathematical Biology*, vol. 64, no. 5, pp. 803–827, 2012.

[161] J. Romero and Y. Rosokha, "Constructing strategies in the indefinitely repeated prisoner's dilemma game," *European Economic Review*, vol. 104, pp. 185–219, 2018.

[162] J. W. Friedman, "A Non-cooperative Equilibrium for Supergames12," *The Review of Economic Studies*, vol. 38, no. 1, pp. 1–12, 1971.

[163] A. Szolnoki, M. Perc, and G. Szabó, "Phase diagrams for three-strategy evolutionary prisoner's dilemma games on regular graphs," *Physical Review E*, vol. 80, no. 5, p. 056104, 2009.

[164] P. Duersch, J. Oechssler, and B. C. Schipper, "When is tit-for-tat unbeatable?," *International Journal of Game Theory*, vol. 43, no. 1, pp. 25–36, 2014.

[165] P. Molander, "The Optimal Level of Generosity in a Selfish, Uncertain Environment," *Journal of Conflict Resolution*, vol. 29, no. 4, pp. 611–618, 1985.

[166] M. A. Nowak and R. M. May, "Evolutionary games and spatial chaos," *Nature*, vol. 359, no. 6398, pp. 826–829, 1992.

[167] C. Hauert and O. Stenull, "Simple Adaptive Strategy Wins the Prisoner's Dilemma," *Journal of Theoretical Biology*, vol. 218, no. 3, pp. 261–272, 2002.

[168] T. Killingback and M. Doebeli, "The Continuous Prisoner's Dilemma and the Evolution of Cooperation through Reciprocal Altruism with Variable Investment.," *The American Naturalist*, vol. 160, no. 4, pp. 421–438, 2002.

[169] S. Kurokawa, J. Y. Wakano, and Y. Ihara, "Generous cooperators can outperform non-generous cooperators when replacing a population of defectors," *Theoretical Population Biology*, vol. 77, no. 4, pp. 257–262, 2010.

[170] A. J. Stewart and J. B. Plotkin, "Small groups and long memories promote cooperation," *Scientific Reports*, vol. 6, no. 1, p. 26889, 2016.

[171] A. McAvoy and M. A. Nowak, "Reactive learning strategies for iterated games," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 475, no. 2223, p. 20180819, 2019.

[172] R. Boyd and J. P. Lorberbaum, “No pure strategy is evolutionarily stable in the repeated Prisoner’s Dilemma game,” *Nature*, vol. 327, no. 6117, pp. 58–59, 1987.

[173] S. Van Segbroeck, J. M. Pacheco, T. Lenaerts, and F. C. Santos, “Emergence of Fairness in Repeated Group Interactions,” *Physical Review Letters*, vol. 108, no. 15, p. 158104, 2012.

[174] V. Knight, M. Harper, N. E. Glynatsi, and O. Campbell, “Evolution reinforces cooperation with the emergence of self-recognition mechanisms: An empirical study of strategies in the Moran process for the iterated prisoner’s dilemma,” *PLOS ONE*, vol. 13, no. 10, p. e0204981, 2018.

[175] X. Chen and F. Fu, “Outlearning extortioners: Unbending strategies can foster reciprocal fairness and cooperation,” *PNAS Nexus*, vol. 2, no. 6, p. pgad176, 2023.

[176] K. Donahue, O. P. Hauser, M. A. Nowak, and C. Hilbe, “Evolving cooperation in multichannel games,” *Nature Communications*, vol. 11, no. 1, p. 3885, 2020.

[177] J. G. Reiter, C. Hilbe, D. G. Rand, K. Chatterjee, and M. A. Nowak, “Crosstalk in concurrent repeated games impedes direct reciprocity and requires stronger levels of forgiveness,” *Nature Communications*, vol. 9, no. 1, p. 555, 2018.

[178] B. D. Bernheim and M. D. Whinston, “Multimarket Contact and Collusive Behavior,” *The RAND Journal of Economics*, vol. 21, no. 1, p. 1, 1990.

[179] H. Matsushima, “Multimarket Contact, Imperfect Monitoring, and Implicit Collusion,” *Journal of Economic Theory*, vol. 98, no. 1, pp. 158–178, 2001.

[180] J. H. Fowler and N. A. Christakis, “Cooperative behavior cascades in human social networks,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 12, pp. 5334–5338, 2010.

[181] J. Yang, T. Kawamura, and K. Ogawa, “Experimental Multimarket Contact Inhibits Cooperation,” *Metroeconomica*, vol. 67, no. 1, pp. 21–43, 2016.

[182] V. Laferrière, J. Montez, C. Roux, and C. Thöni, “Multigame Contact: A Double-Edged Sword for Cooperation,” 2023.

[183] W. Raub and J. Weesie, “Reputation and Efficiency in Social Interactions: An Example of Network Effects,” | *American Journal of Sociology*, vol. 96, 1990.

[184] M. Kandori, “Social Norms and Community Enforcement,” *The Review of Economic Studies*, vol. 59, no. 1, pp. 63–80, 1992.

[185] M. C. Couto and S. Pal, “Introspection Dynamics in Asymmetric Multiplayer Games,” *Dynamic Games and Applications*, vol. 13, no. 4, pp. 1256–1285, 2023.

[186] P. Dolan and M. M. Galizzi, “Like ripples on a pond: Behavioral spillovers and their implications for research and policy,” *Journal of Economic Psychology*, vol. 47, pp. 1–16, 2015.

[187] A. Maki, A. R. Carrico, K. T. Raimi, H. B. Truelove, B. Araujo, and K. L. Yeung, "Meta-analysis of pro-environmental behaviour spillover," *Nature Sustainability*, vol. 2, no. 4, pp. 307–315, 2019.

[188] M. Rabin and G. Weizsäcker, "Narrow Bracketing and Dominated Choices," *American Economic Review*, vol. 99, no. 4, pp. 1508–1543, 2009.

[189] R. A. Hill and R. I. M. Dunbar, "Social network size in humans," *Human Nature*, vol. 14, no. 1, pp. 53–72, 2003.

[190] G. Ichinose and N. Masuda, "Zero-determinant strategies in finitely repeated games," *Journal of Theoretical Biology*, vol. 438, pp. 61–77, 2018.

[191] A. Traulsen, M. A. Nowak, and J. M. Pacheco, "Stochastic dynamics of invasion and fixation," *Physical Review E*, vol. 74, no. 1, p. 011909, 2006.

[192] L. E. Blume, "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, vol. 5, no. 3, pp. 387–424, 1993.

[193] G. Szabó and C. Tőke, "Evolutionary prisoner's dilemma game on a square lattice," *Physical Review E*, vol. 58, no. 1, pp. 69–73, 1998.

[194] D. L. Chen, M. Schonger, and C. Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, vol. 9, pp. 88–97, 2016.

[195] W. S. Neilson, "The economics of favors," *Journal of Economic Behavior & Organization*, vol. 39, no. 4, pp. 387–397, 1999.

[196] M. K. Schweinfurth and J. Call, "Reciprocity: Different behavioural strategies, cognitive mechanisms and psychological processes," *Learning & Behavior*, vol. 47, no. 4, pp. 284–301, 2019.

[197] R. O. Keohane, "Reciprocity in international relations," *International Organization*, vol. 40, no. 1, pp. 1–27, 1986.

[198] C. Hilbe, K. Chatterjee, and M. A. Nowak, "Partners and rivals in direct reciprocity," *Nature Human Behaviour*, vol. 2, no. 7, pp. 469–477, 2018.

[199] D. Kraines and V. Kraines, "Pavlov and the prisoner's dilemma," *Theory and Decision*, vol. 26, no. 1, pp. 47–79, 1989.

[200] E. Akin, "What You Gotta Know to Play Good in the Iterated Prisoner's Dilemma," *Games*, vol. 6, no. 3, pp. 175–190, 2015.

[201] H. Brandt and K. Sigmund, "The good, the bad and the discriminator—Errors in direct and indirect reciprocity," *Journal of Theoretical Biology*, vol. 239, no. 2, pp. 183–194, 2006.

[202] Q. Su, A. McAvoy, L. Wang, and M. A. Nowak, "Evolutionary dynamics with game transitions," *Proceedings of the National Academy of Sciences*, vol. 116, no. 51, pp. 25398–25404, 2019.

[203] S. D. Yi, S. K. Baek, and J.-K. Choi, "Combination with anti-tit-for-tat remedies problems of tit-for-tat," *Journal of Theoretical Biology*, vol. 412, pp. 1–7, 2017.

[204] D. Hao, Z. Rong, and T. Zhou, "Extortion under uncertainty: Zero-determinant strategies in noisy games," *Physical Review E*, vol. 91, no. 5, p. 052803, 2015.

[205] A. McAvoy and C. Hauert, "Autocratic strategies for iterated games with arbitrary action spaces," *Proceedings of the National Academy of Sciences*, vol. 113, no. 13, pp. 3573–3578, 2016.

[206] R. Boyd, "Mistakes allow evolutionary stability in the repeated prisoner's dilemma game," *Journal of Theoretical Biology*, vol. 136, no. 1, pp. 47–56, 1989.

[207] C. Xia, C. Gracia-Lázaro, and Y. Moreno, "Effect of memory, intolerance, and second-order reputation on cooperation," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 6, p. 063122, 2020.

[208] F. L. Pinheiro, V. V. Vasconcelos, F. C. Santos, and J. M. Pacheco, "Evolution of All-or-None Strategies in Repeated Public Goods Dilemmas," *PLOS Computational Biology*, vol. 10, no. 11, p. e1003945, 2014.

[209] J. García and M. Van Veelen, "In and out of equilibrium I: Evolution of strategies in repeated games with discounting," *Journal of Economic Theory*, vol. 161, pp. 161–189, 2016.

[210] P. Duersch, J. Oechssler, and B. C. Schipper, "Unbeatable imitation," *Games and Economic Behavior*, vol. 76, no. 1, pp. 88–96, 2012.

[211] T. K. Shackelford and D. M. Buss, "Betrayal in Mateships, Friendships, and Coalitions," *Personality and Social Psychology Bulletin*, vol. 22, no. 11, pp. 1151–1164, 1996.

[212] J. Bendor and P. Swistak, "Types of evolutionary stability and the problem of cooperation.," *Proceedings of the National Academy of Sciences*, vol. 92, no. 8, pp. 3596–3600, 1995.

[213] J. García and M. van Veelen, "No Strategy Can Win in the Repeated Prisoner's Dilemma: Linking Game Theory and Computer Simulations," *Frontiers in Robotics and AI*, vol. 5, 2018.

[214] E. Akin, "Good Strategies for the Iterated Prisoner's Dilemma : Smale vs. Markov," 2017.

[215] D. G. Rand and M. A. Nowak, "Human cooperation," *Trends in cognitive sciences*, vol. 17, no. 8, pp. 413–425, 2013.

[216] O. P. Hauser, C. Hilbe, K. Chatterjee, and M. A. Nowak, "Social dilemmas among unequals," *Nature*, vol. 572, no. 7770, pp. 524–527, 2019.

[217] J. L. Pletzer, D. Balliet, J. Joireman, D. M. Kuhlman, S. C. Voelpel, and P. A. Van Lange, "Social Value Orientation, Expectations, and Cooperation in Social Dilemmas: A Meta–Analysis," *European Journal of Personality*, vol. 32, no. 1, pp. 62–83, 2018.

[218] C. Alós-Ferrer and F. Farolfi, "Trust Games and Beyond," *Frontiers in Neuroscience*, vol. 13, p. 887, 2019.

[219] D. L. Ferrin, M. C. Bligh, and J. C. Kohles, "It takes two to tango: An interdependence analysis of the spiraling of perceived trustworthiness and cooperation in interpersonal and intergroup relationships," *Organizational Behavior and Human Decision Processes*, vol. 107, no. 2, pp. 161–178, 2008.

[220] A. Peysakhovich, M. A. Nowak, and D. G. Rand, "Humans display a 'cooperative phenotype' that is domain general and temporally stable," *Nature Communications*, vol. 5, no. 1, p. 4939, 2014.

[221] A. G. Reigstad, E. A. Strømland, and G. Tinghög, "Extending the Cooperative Phenotype: Assessing the Stability of Cooperation across Countries," *Frontiers in Psychology*, vol. 8, p. 1990, 2017.

[222] P. Barclay, "Biological markets and the effects of partner choice on cooperation and friendship," *Current Opinion in Psychology*, vol. 7, pp. 33–38, 2016.

[223] S. Fehrler and W. Przepiorka, "Charitable giving as a signal of trustworthiness: Disentangling the signaling benefits of altruistic acts," *Evolution and Human Behavior*, vol. 34, no. 2, pp. 139–145, 2013.

[224] P. Barclay, "Reciprocity creates a stake in one's partner, or why you should cooperate even when anonymous," *Proceedings of the Royal Society B: Biological Sciences*, vol. 287, no. 1929, p. 20200819, 2020.

[225] S. Vesely, C. A. Klöckner, and C. Brick, "Pro-environmental behavior as a signal of cooperativeness: Evidence from a social dilemma experiment," *Journal of Environmental Psychology*, vol. 67, p. 101362, 2020.

[226] J. J. Jordan, M. Hoffman, M. A. Nowak, and D. G. Rand, "Uncalculating cooperation is used to signal trustworthiness," *Proceedings of the National Academy of Sciences*, vol. 113, no. 31, pp. 8658–8663, 2016.

[227] B. Rockenbach and M. Milinski, "To qualify as a social partner, humans hide severe punishment, although their observed cooperativeness is decisive," *Proceedings of the National Academy of Sciences*, vol. 108, no. 45, pp. 18307–18312, 2011.

[228] K. Eriksson, P. Strimling, P. A. Andersson, M. Aveyard, M. Brauer, V. Gritskov, T. Kiyonari, D. M. Kuhlman, A. T. Maitner, Z. Manesi, C. Molho, L. S. Peperkoorn, M. Rizwan, A. W.

Stivers, Q. Tian, P. A. M. Van Lange, I. Vartanova, J. Wu, and T. Yamagishi, “Cultural Universals and Cultural Differences in Meta-Norms about Peer Punishment,” *Management and Organization Review*, vol. 13, no. 4, pp. 851–870, 2017.

[229] M. Hoffman, E. Yoeli, and M. A. Nowak, “Cooperate without looking: Why we care what people think and not just what they do,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 6, pp. 1727–1732, 2015.

[230] E. E. Levine, A. Barasch, D. Rand, and J. Z. Berman, “Signaling Emotion and Reason in Cooperation,” *Journal of Experimental Psychology: General*, vol. 147, no. 5, 2018.

[231] A. Gardner, “The greenbeard effect,” *Current Biology*, vol. 29, no. 11, pp. R430–R431, 2019.

[232] J. Andreoni and L. Vesterlund, “Which is the Fair Sex? Gender Differences in Altruism,” *The Quarterly Journal of Economics*, vol. 116, no. 1, pp. 293–312, 2001.

[233] A. Ortmann and L. K. Tichy, “Gender differences in the laboratory: Evidence from prisoner’s dilemma games,” *Journal of Economic Behavior & Organization*, vol. 39, no. 3, pp. 327–339, 1999.

[234] R. Croson and U. Gneezy, “Gender Differences in Preferences,” *Journal of Economic Literature*, vol. 47, no. 2, pp. 448–474, 2009.

[235] D. Balliet, N. P. Li, S. J. Macfarlan, and M. Van Vugt, “Sex differences in cooperation: A meta-analytic review of social dilemmas.,” *Psychological Bulletin*, vol. 137, no. 6, pp. 881–909, 2011.

[236] A. R. Dorrough and A. Glöckner, “A cross-national analysis of sex differences in prisoner’s dilemma games,” *British Journal of Social Psychology*, vol. 58, no. 1, pp. 225–240, 2019.

[237] C. L. Exley, O. P. Hauser, M. Moore, and J.-H. Pezzuto, “Beliefs about gender differences in social preferences,” *Discussion Papers*, no. 2204, 2022.

[238] F. Aguiar, P. Brañas-Garza, R. Cobo-Reyes, N. Jimenez, and L. M. Miller, “Are women expected to be more generous?,” *Experimental Economics*, vol. 12, no. 1, pp. 93–98, 2009.

[239] P. Brañas-Garza, V. Capraro, and E. Rascón-Ramírez, “Gender differences in altruism on Mechanical Turk: Expectations and actual behaviour,” *Economics Letters*, vol. 170, pp. 19–23, 2018.

[240] W. De Neys, A. Hopfensitz, and J.-F. Bonnefon, “Split-Second Trustworthiness Detection From Faces in an Economic Game,” *Experimental Psychology*, vol. 64, no. 4, pp. 231–239, 2017.

[241] A. F. Danvers and M. N. Shiota, “Dynamically engaged smiling predicts cooperation above and beyond average smiling levels,” *Evolution and Human Behavior*, vol. 39, no. 1, pp. 112–119, 2018.

[242] B. Jaeger, B. Oud, T. Williams, E. G. Krumhuber, E. Fehr, and J. B. Engelmann, “Can people detect the trustworthiness of strangers based on their facial appearance?,” *Evolution and Human Behavior*, vol. 43, no. 4, pp. 296–303, 2022.

[243] K. K. Knowles and A. C. Little, “Vocal fundamental and formant frequencies affect perceptions of speaker cooperativeness,” *Quarterly Journal of Experimental Psychology*, vol. 69, no. 9, pp. 1657–1675, 2016.

[244] J. Andreoni and R. Petrie, “Beauty, gender and stereotypes: Evidence from laboratory experiments,” *Journal of Economic Psychology*, vol. 29, no. 1, pp. 73–93, 2008.

[245] R. K. Wilson and C. C. Eckel, “Judging a Book by its Cover: Beauty and Expectations in the Trust Game,” *Political Research Quarterly*, vol. 59, no. 2, pp. 189–202, 2006.

[246] P. K. Piff, M. W. Kraus, S. Côté, B. H. Cheng, and D. Keltner, “Having less, giving more: The influence of social class on prosocial behavior.,” *Journal of Personality and Social Psychology*, vol. 99, no. 5, pp. 771–784, 2010.

[247] P. Smeets, R. Bauer, and U. Gneezy, “Giving behavior of millionaires,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 34, pp. 10641–10644, 2015.

[248] Y. Qi, Q. Li, and F. Du, “Are Rich People Perceived as More Trustworthy? Perceived Socioeconomic Status Modulates Judgments of Trustworthiness and Trust Behavior Based on Facial Appearance,” *Frontiers in Psychology*, vol. 9, p. 512, 2018.

[249] A. F. Martinangeli, “Do what (you think) the rich will do: Inequality and belief heterogeneity in public good provision,” *Journal of Economic Psychology*, vol. 83, p. 102364, 2021.

[250] S. T. Trautmann, X. Wang, Y. Wang, and Y. Xu, “Holding High-Status Individuals to Higher Standards.” 2021.

[251] R. Helénsdotter, “Experimental Evidence on Cooperation, Political Affiliation, and Group Size,” *SSRN Electronic Journal*, 2019.

[252] T. Grünhage and M. Reuter, “Political Orientation is Associated with Behavior in Public-Goods- and Trust-Games,” *Political Behavior*, vol. 44, no. 1, pp. 23–48, 2022.

[253] A. Romano, M. Sutter, J. H. Liu, and D. Balliet, “Political ideology, cooperation and national parochialism across 42 nations,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, 2020.

[254] D. Balliet, J. M. Tybur, J. Wu, C. Antonellis, and P. A. M. Van Lange, “Political Ideology, Trust, and Cooperation: In-group Favoritism among Republicans *and* Democrats during a US National Election,” *Journal of Conflict Resolution*, vol. 62, no. 4, pp. 797–818, 2018.

[255] A. Norenzayan and A. F. Shariff, “The Origin and Evolution of Religious Prosociality,” *Science*, vol. 322, no. 5898, pp. 58–62, 2008.

[256] J. Billingsley, C. M. Gomes, and M. E. McCullough, “Implicit and explicit influences of religious cognition on Dictator Game transfers,” *Royal Society Open Science*, vol. 5, no. 8, p. 170238, 2018.

[257] M. E. McCullough, P. Swartwout, J. H. Shaver, E. C. Carter, and R. Sosis, “Christian religious badges instill trust in Christian and non-Christian perceivers.,” *Psychology of Religion and Spirituality*, vol. 8, no. 2, pp. 149–163, 2016.

[258] J. A. C. Everett, O. S. Haque, and D. G. Rand, “How Good Is the Samaritan, and Why?: An Experimental Investigation of the Extent and Nature of Religious Prosociality Using Economic Games,” *Social Psychological and Personality Science*, vol. 7, no. 3, pp. 248–255, 2016.

[259] D. H. Bostyn and A. Roets, “Trust, trolleys and social dilemmas: A replication study.,” *Journal of Experimental Psychology: General*, vol. 146, no. 5, pp. e1–e7, 2017.

[260] V. Capraro, J. Sippel, B. Zhao, L. Hornischer, M. Savary, Z. Terzopoulou, P. Faucher, and S. F. Griffioen, “People making deontological judgments in the Trapdoor dilemma are perceived to be more prosocial in economic games than they actually are,” *PLOS ONE*, vol. 13, no. 10, p. e0205066, 2018.

[261] Y. Chen and S. X. Li, “Group Identity and Social Preferences,” *American Economic Review*, vol. 99, no. 1, pp. 431–457, 2009.

[262] L. Goette, D. Huffman, and S. Meier, “The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups,” *American Economic Review*, vol. 96, no. 2, pp. 212–216, 2006.

[263] M. M. Kumar, L. Tsoi, M. S. Lee, J. Cone, and K. McAuliffe, “Nationality dominates gender in decision-making in the Dictator and Prisoner’s Dilemma Games,” *PLOS ONE*, vol. 16, no. 1, p. e0244568, 2021.

[264] D. Balliet, J. Wu, and C. K. W. De Dreu, “Ingroup favoritism in cooperation: A meta-analysis.,” *Psychological Bulletin*, vol. 140, no. 6, pp. 1556–1581, 2014.

[265] G. Charness, L. Rigotti, and A. Rustichini, “Social surplus determines cooperation rates in the one-shot Prisoner’s Dilemma,” *Games and Economic Behavior*, vol. 100, pp. 113–124, 2016.

[266] D. G. Rand, J. D. Greene, and M. A. Nowak, “Spontaneous giving and calculated greed,” *Nature*, vol. 489, no. 7416, pp. 427–430, 2012.

[267] S. Bouwmeester, P. P. J. L. Verkoeijen, B. Aczel, F. Barbosa, L. Bègue, P. Brañas-Garza, T. G. H. Chmura, G. Cornelissen, F. S. Døssing, A. M. Espín, A. M. Evans, F. Ferreira-Santos, S. Fiedler, J. Flegr, M. Ghaffari, A. Glöckner, T. Goeschl, L. Guo, O. P. Hauser, R. Hernan-Gonzalez, A. Herrero, Z. Horne, P. Houdek, M. Johannesson, L. Koppel, P. Kujal, T. Laine, J. Lohse, E. C. Martins, C. Mauro, D. Mischkowski, S. Mukherjee, K. O. R.

Myrseth, D. Navarro-Martínez, T. M. S. Neal, J. Novakova, R. Pagà, T. O. Paiva, B. Palfi, M. Piovesan, R.-M. Rahal, E. Salomon, N. Srinivasan, A. Srivastava, B. Szaszi, A. Szollosi, K. Ø. Thor, G. Tinghög, J. S. Trueblood, J. J. Van Bavel, A. E. van 't Veer, D. Västfjäll, M. Warner, E. Wengström, J. Wills, and C. E. Wollbrant, "Registered Replication Report: Rand, Greene, and Nowak (2012)," *Perspectives on Psychological Science*, vol. 12, no. 3, pp. 527–542, 2017.

[268] A. M. Evans, K. D. Dillon, and D. G. Rand, "Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas," *Journal of Experimental Psychology: General*, vol. 144, no. 5, pp. 951–966, 2015.

[269] A. M. Evans and P. P. van de Calseyde, "The effects of observed decision time on expectations of extremity and cooperation," *Journal of Experimental Social Psychology*, vol. 68, pp. 50–59, 2017.

[270] D. Balliet, "Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review," *Journal of Conflict Resolution*, vol. 54, no. 1, pp. 39–57, 2010.

[271] S. He, T. Offerman, and J. van de Ven, "The Sources of the Communication Gap," *Management Science*, vol. 63, no. 9, pp. 2832–2846, 2017.

[272] A. Sparks, T. Burleigh, and P. Barclay, "We can see inside: Accurate prediction of Prisoner's Dilemma decisions in announced games following a face-to-face interaction," *Evolution and Human Behavior*, vol. 37, no. 3, pp. 210–216, 2016.

[273] J. J. Jordan, M. Hoffman, P. Bloom, and D. G. Rand, "Third-party punishment as a costly signal of trustworthiness," *Nature*, vol. 530, no. 7591, pp. 473–476, 2016.

[274] B. Jaeger, A. M. Evans, M. Stel, and I. van Beest, "Explaining the persistent influence of facial cues in social decision-making.," *Journal of Experimental Psychology: General*, vol. 148, no. 6, pp. 1008–1021, 2019.

[275] A. Zylbersztejn, Z. Babutsidze, and N. Hanaki, "Preferences for observable information in a strategic setting: An experiment," *Journal of Economic Behavior & Organization*, vol. 170, pp. 268–285, 2020.

[276] C. Darwin, *The Descent of Man*. 1871.

[277] J. Graham, P. Meindl, E. Beall, K. M. Johnson, and L. Zhang, "Cultural differences in moral judgment and behavior, across and within societies," *Current Opinion in Psychology*, vol. 8, pp. 125–130, 2016.

[278] J. Graham, J. Haidt, and B. A. Nosek, "Liberals and conservatives rely on different sets of moral foundations.," *Journal of Personality and Social Psychology*, vol. 96, no. 5, pp. 1029–1046, 2009.

[279] D. Fudenberg and D. K. Levine, "Reputation and equilibrium selection in games with a patient player," in *A Long-Run Collaboration on Long-Run Games*, pp. 123–142, WORLD SCIENTIFIC, 2008.

[280] M. Singh and M. Hoffman, "Commitment and impersonation: A reputation-based theory of principled behavior," preprint, PsyArXiv, 2021.

[281] N. Baumard, J.-B. André, and D. Sperber, "A mutualistic approach to morality: The evolution of fairness by partner choice," *Behavioral and Brain Sciences*, vol. 36, no. 1, pp. 59–78, 2013.

[282] M. E. McCullough, R. Kurzban, and B. A. Tabak, "Cognitive systems for revenge and forgiveness," *Behavioral and Brain Sciences*, vol. 36, no. 1, pp. 1–15, 2013.

[283] H. Mercier and D. Sperber, "Why do humans reason? Arguments for an argumentative theory," *Behavioral and Brain Sciences*, vol. 34, no. 2, pp. 57–74, 2011.

[284] M. Hoffman, C. Hilbe, and M. A. Nowak, "The signal-burying game can explain why we obscure positive traits and good deeds," *Nature Human Behaviour*, vol. 2, no. 6, pp. 397–404, 2018.

[285] J. Lie-Panis and J.-B. André, "Cooperation as a signal of time preferences," *Proceedings of the Royal Society B: Biological Sciences*, vol. 289, no. 1973, p. 20212266, 2022.

[286] J. Lie-Panis and J.-L. Dessalles, "Runaway signals: Exaggerated displays of commitment may result from second-order signaling," *Journal of Theoretical Biology*, vol. 572, p. 111586, 2023.

[287] A. Zahavi, "Mate selection—A selection for a handicap," *Journal of Theoretical Biology*, vol. 53, no. 1, pp. 205–214, 1975.

[288] A. Grafen, "Biological signals as handicaps," *Journal of Theoretical Biology*, vol. 144, no. 4, pp. 517–546, 1990.

[289] M. Spence, "Competitive and optimal responses to signals: An analysis of efficiency and distribution," *Journal of Economic Theory*, vol. 7, no. 3, pp. 296–332, 1974.

[290] V. Capraro and J. Kuilder, "To know or not to know? Looking at payoffs signals selfish behavior, but it does not actually mean so," *Journal of Behavioral and Experimental Economics*, vol. 65, pp. 79–84, 2016.

[291] L. Festinger, "Cognitive Dissonance," *SCIENTIFIC AMERICAN*, 1962.

[292] J. L. Freedman and S. C. Fraser, "Compliance without pressure: The foot-in-the-door technique.," *Journal of Personality and Social Psychology*, vol. 4, no. 2, pp. 195–202, 1966.

[293] B. Monin and D. T. Miller, "Moral credentials and the expression of prejudice," *Journal of Personality and Social Psychology*, vol. 81, no. 1, pp. 33–43, 2001.

[294] N. Kuper and A. Bott, "Has the evidence for moral licensing been inflated by publication bias?," *Meta-psychology*, 2019.

[295] A. Rotella and P. Barclay, “Failure to replicate moral licensing and moral cleansing in an online experiment,” *Personality and Individual Differences*, vol. 161, p. 109967, 2020.

[296] E. Mullen and B. Monin, “Consistency Versus Licensing Effects of Past Moral Behavior,” *Annual Review of Psychology*, vol. 67, no. 1, pp. 363–385, 2016.

[297] E. A. Posner, “The Strategic Basis of Principled Behavior: A Critique of the Incommensurability Thesis,” *University of Pennsylvania Law Review*, vol. 146, no. 5, pp. 1185–1214, 1998.

[298] Q. Su, A. McAvoy, Y. Mori, and J. B. Plotkin, “Evolution of prosocial behaviours in multilayer populations,” *Nature Human Behaviour*, vol. 6, no. 3, pp. 338–348, 2022.

[299] L. C. Hoenig, R. Pliskin, and C. De Dreu, “Equality and Efficiency shape Cooperation in Multiple-Public Goods Provision Problems.” 2023.

[300] R. McElreath and R. Boyd, *Mathematical Models of Social Evolution: A Guide for the Perplexed.* Chicago ; London: University of Chicago Press, 2007.

[301] J. Newton, “Evolutionary Game Theory: A Renaissance,” *Games*, vol. 9, no. 2, p. 31, 2018.

[302] J.-B. André, M. Cozic, S. De Monte, J. Gayon, P. Huneman, J. Martens, and B. Walliser, *From Evolutionary Biology to Economics and Back: Parallels and Crossings between Economics and Evolution*, vol. 28 of *History, Philosophy and Theory of the Life Sciences*. Cham: Springer International Publishing, 2022.

Appendices

Chapter 3 Supplementary information

Supporting Information

Dynamics of cooperation in concurrent games

Charlotte S. L. Rossetti[1], Oliver P. Hauser[2,†], Christian Hilbe[1,†]

[1]Max Planck Research Group on the Dynamics of Social Behavior,
Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany
[2]Department of Economics, University of Exeter, Exeter, EX4 4PU, UK
[†] These authors contributed equally.

Contents

1 Related literature

To a large extent, evolutionary game theory takes a reductionist approach. It aims to understand social behavior by exploring the dynamics of individual games. This is also the approach of most text books (1–4) and reviews (5–10). Instead, here we are interested in describing how people act if they are involved in several games concurrently (possibly with different interaction partners). In the following, we summarize the previous theoretical and experimental literature on this topic, and we describe how our approach differs from this literature.

1.1 Theoretical work

We study direct reciprocity in concurrently ongoing repeated games. This work is naturally related to several strands of the evolutionary game theory literature.

Evolutionary dynamics of multi-games. Several researchers have explored the evolutionary dynamics in concurrent non-repeated games (11–14). Here, players engage in two or more one-shot games concurrently. This literature asks to which extent the dynamics and the equilibria of such 'multi-games' can be directly inferred from the constituent games. While the equilibria of multi-games are typically directly related to the equilibria of the individual one-shot games (11), the overall evolutionary dynamics may differ. In particular, multi-games may lead to evolutionary oscillations that have no analogue in any of the constituent one-shot games (12). While this work provides important insights into the evolutionary dynamics of one-shot games, it does not extend to reciprocal interactions. In these models, individuals have no possibility (or interest) to strategically link their behavior across different games. Similarly, these models do not study the effect of spillovers from one game to another.

Social dilemmas with punishment or rewards. There is also abundant work in which a social dilemma is coupled with a subsequent game in which individuals can reward or punish each other (15–21). Such social dilemmas with incentives share some similarities with our same-partner treatment, to the extent that individuals can use one game (the rewarding or the punishment game) to incentivize additional cooperation in the other. A crucial difference, however, is that the second punishment/reward stage is explicitly designed to strategically influence co-players to cooperate. In contrast, we are interested in how people approach concurrently ongoing games when there is no exogenous indication that behavior in one game might or even should be tied to the behavior in the other.

Direct reciprocity. While there is a vast literature on the evolution of reciprocal cooperation, most studies assume that individuals only interact in one repeated game at a time, or that they treat all their games as independent (22–40). The setup of these studies thus corresponds to the setup of our control treatment. Compared to these studies, we ask: does the presence of other ongoing games enhance or suppress cooperation? How can we model the different ways in which behavior in one game may consciously or subconsciously affect behavior in another game?

Some of these questions have been taken up by the industrial economics literature, when asking how firms compete in several distinct markets (41, 42). These studies conceptualize the competition across different markets as formally independent repeated games, similar to our same-partner treatment. Assuming that markets and firms are symmetric, this literature finds an irrelevance result (41): multi-market contact does not allow firms to collude more effectively, compared to the possibilities given by each isolated game. The respective papers, however, take an equilibrium perspective. They characterize optimal behavior among rational players. In contrast, we are interested in how people learn to play strategically over time (far from equilibrium). To this end, we also formulate learning processes that aim to capture some of the cognitive constraints and biases people may be susceptible to.

Two studies from the evolutionary game theory literature are most closely related to what we do. First, Reiter et al (43) consider players arranged on a network, who interact in a repeated prisoner's dilemma with each of their neighbors. The players' strategies treat each game as independent (as in our control treatment). That is, the strategies only depend on the previous action of the respective neighbor. However, the model allows for 'crosstalk' between different games. When a crosstalk event occurs, a player who intends to defect against one neighbor may mistakenly also defect against another neighbor (similarly for cooperation). In line with the results of our model extension on exogenous spillovers, the study finds that higher crosstalk rates undermine the evolution of cooperation. Importantly, however, their model does not study the endogenous evolution of linkage between independent games. Instead crosstalk between games is included by an exogenous parameter.

Second, Donahue et al (44) describe the equilibria of our same-partner treatment. As in our case, there are two players who engage in two repeated games in parallel. Players can condition their behavior in either game on the previous outcome of both games. As in our simulation of the baseline scenario (**Fig. 2**), they find that concurrent games can lead to strategic linkage. This linkage in turn can promote cooperation compared to the control treatment. Our model generalizes this work substantially: (*i*) We introduce a consistent framework that allows us to compare strategic effects of linking different games with non-strategic effects (by comparing the same-partner and the different-partners treatment). (*ii*) We explore the impact of several plausible biases and psychological constraints (see also Section 2.2). (*iii*) By designing and analyzing a behavioral experiment based on our model, we compare the theoretical predictions to empirical evidence.

1.2 Experimental work

Direct reciprocity. Just as in the theoretical papers mentioned above, most experimental studies look at individual games in isolation (45). One notable exception is spatial games played on a network where players interact with their neighbors (46–48). In these spatial prisoner's dilemma, players have been found to reciprocate cooperation based on the number of their neighbours who cooperated on the previous round and their own previous action (49). Nonetheless, as the number of co-players increases, cooperation decays (50–52). However, in these experiments players' actions are the same for all neighbors, meaning they cannot tailor their choice of action to each individual neighbor. Further experiments on dynamic networks find that players are eager to cut ties to non-cooperative neighbors when they are offered the possibility (53, 54), even when doing so is costly (55).

Spillover in multi-games. There are a few studies that look into the effect of spillovers between games when played simultaneously (56–59). They look into behavioral spillovers between games, but not reciprocity and strategic linkage in particular. Angelovski *et al* (60) consider a setup similar to our different-partners treatment. They investigate how people interact in two simultaneous 2-players public goods games. Players are arranged on a circle, meaning that each player interacts with two neighbors. The two games differ in their incentives: one has a high incentive, the other a low one. In addition, they include three controls where (1) both games have the high incentive, (2) both games have the low incentive, and (3) both games have the average of the two incentives. The results show that cooperation in the asymmetric public goods game matches the level found in the control with high incentives in both games, meaning that the game with the higher incentive is able to lift cooperation in the game with lower incentive. They also check whether contributions in one game depend on how much the neighbor in the other game contributed on average. Average contributions to each neighbor correlate with each other and are behaviorally interdependent.

Finally, one paper contemporaneous to ours, on the repeated prisoner's dilemma, uses a similar design to us. Laferrière *et al* (61) look at cooperation in repeated games with different payoffs played simultaneously with either the same or two different co-players. The games only differ in their temptation and sucker payoff while the reward and punishment payoffs stay the same. The paper is focused on experiments and has no evolutionary game theory component. In line with our results, they find that average cooperation is the same whether the interactions are with the same or different co-players. However, they do not include a single game control to understand the effect of playing two games simultaneously. Our experiment includes such a control where players interact in a standard repeated prisoner's dilemma with either a high or low benefit of cooperation.

2 Theoretical methods

2.1 Baseline model

We consider two stage games, the high-benefit game (*high game*, $\mathtt{H}$) and the low-benefit game (*low game*, $\mathtt{L}$). In either game, players can either cooperate (C) or defect (D). A cooperator pays a cost $c > 0$ in order for the co-player to get a benefit $b_k > c$. This benefit b_k depends on the type of game, $k \in \{\mathtt{H}, \mathtt{L}\}$. As implied by their names, the high game has a larger benefit $b_{\mathtt{H}} \geq b_{\mathtt{L}}$. Overall, each stage game has the payoff matrix

$$\begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} b_k - c & -c \\ b_k & 0 \end{pmatrix} \end{array} \qquad [1]$$

In particular, both games have the characteristics of a prisoner's dilemma. Mutual cooperation yields the highest total payoff for the two players, but for each individual player, cooperation is dominated. Players engage in these donation games for infinitely many rounds. We refer to each individual iterated game as a *repeated game*. Overall, we analyze three treatments that differ in two aspects:

1. Players either engage in one repeated game, or in both repeated games.
2. When they engage in both games, they either face the same co-player in both games, or a different co-player in each game.

In all cases, players use reactive strategies to make their decisions (62, 63). This means a player's action in any game only depends on the co-players' actions in the previous round (independent of all earlier actions). When players engage in both games, we refer to the resulting supergame as a *concurrent game*. Here players need to make two decisions each round, one for each stage game. Decisions that happen in one repeated game may influence decisions in the other. In the following, we introduce the three treatments in detail and we explain how to compute the players' payoffs in each case.

2.1.1 Control treatment

Setup and strategies. In the control treatment, players only engage in one repeated game at a time. This is the case that is usually considered in the literature. To be consistent with the other treatments, we assume the control treatment involves four players. Players 1 and 2 interact in the repeated game with high benefits; players 3 and 4 engage in the low-benefit game. For either repeated game $k \in \{\mathtt{H}, \mathtt{L}\}$, strategies take the form $\mathbf{p^k} = (p^k_C, p^k_D) \in [0,1]^2$. Here, p^k_a refers to the player's probability to cooperate depending on the co-player's previous action $a \in \{C, D\}$. For example, $\mathbf{p^k} = (0,0)$ refers to the strategy ALLD. As another example, $\mathbf{p^k} = (1,0)$ implements the strategy TFT (Tit-for-Tat).

While we assume for the control treatment that each player only engages in one game at a time, we obtain equivalent results if they play both games (either with the same partner or with different partners), provided players can be assumed to treat both games as independent (see **Fig. S1**). That is, we obtain an equivalent formulation of our control treatment if we assume players engage in both games simultaneously, but strategies are restricted to take the form $\mathbf{p} = (p_C^{\mathrm{H}}, p_D^{\mathrm{H}};\ p_C^{\mathrm{L}}, p_D^{\mathrm{L}})$. Here, p_a^k is the player's cooperation probability in game k, depending on the co-player's previous action in the respective game only. The theoretical results of this alternative formulation might be somewhat easier to compare to the other treatments (because then all treatments would involve two repeated games for each player). However, this alternative formulation is more difficult to implement experimentally. Therefore, we use the first formulation throughout, as depicted in our **Fig. 1b**.

Calculation of payoffs. We describe how to compute payoffs for players 1 and 2. Payoffs for the other two players are computed equivalently. Suppose player 1 uses strategy $\mathbf{p^1} = (p_C^1, p_D^1)$, whereas player 2 uses strategy $\mathbf{p^2} = (p_C^2, p_D^2)$. To compute payoffs, we represent the repeated game as a Markov chain (see, for example, Ref. 3). The possible states of the Markov chain are the possible outcomes of a single stage game, (C,C), (C,D), (D,C), (D,D). Here, the first and the second letter refer to the actions of player 1 and player 2, respectively. Given the players' strategies, the transition matrix takes the following form (3),

$$M(\mathbf{p^1}, \mathbf{p^2}) = \begin{pmatrix} p_C^1\, p_C^2 & p_C^1\,(1-p_C^2) & (1-p_C^1)\, p_C^2 & (1-p_C^1)\,(1-p_C^2) \\ p_C^1\, p_D^2 & p_C^1\,(1-p_D^2) & (1-p_C^1)\, p_D^2 & (1-p_C^1)\,(1-p_D^2) \\ p_D^1\, p_C^2 & p_D^1\,(1-p_C^2) & (1-p_D^1)\, p_C^2 & (1-p_D^1)\,(1-p_C^2) \\ p_D^1\, p_D^2 & p_D^1\,(1-p_D^2) & (1-p_D^1)\, p_D^2 & (1-p_D^1)\,(1-p_D^2) \end{pmatrix}. \qquad [2]$$

As the two players interact in this game for many rounds, play converges to an invariant distribution $\mathbf{v} = (v_{CC}, v_{CD}, v_{DC}, v_{DD})$. Each entry v_{a^1,a^2} gives the long-run probability to observe rounds in which the first player chooses a^1 whereas the second player chooses a^2. For $\mathbf{p^1}, \mathbf{p^2} \in (0,1)^2$, this invariant distribution is unique. By the theorem of Perron-Frobenius, it is given by the solution of the eigenvector problem

$$\mathbf{v} = \mathbf{v}M(\mathbf{p^1}, \mathbf{p^2}). \qquad [3]$$

Based on this invariant distribution, we can compute the players' average cooperation rates as

$$\gamma^1 = v_{CC} + v_{CD} \quad \text{and} \quad \gamma^2 = v_{CC} + v_{DC}. \qquad [4]$$

In the control treatment, these cooperation rates can be computed explicitly (3), as a function of the players' reactive strategies $\mathbf{p^{1k}}$ and $\mathbf{p^{2k}}$,

$$\gamma^1 = \frac{p_D^1 + (p_C^1 - p_D^1)\cdot p_D^2}{1 - (p_C^1 - p_D^1)(p_C^2 - p_D^2)} \quad \text{and} \quad \gamma^2 = \frac{p_D^2 + (p_C^2 - p_D^2)\cdot p_D^1}{1 - (p_C^1 - p_D^1)(p_C^2 - p_D^2)}. \qquad [5]$$

As a result, we can compute the players' average payoffs for the repeated game as

$$\pi^1 = b_{\text{H}} \cdot \gamma^2 - c \cdot \gamma^1 \quad \text{and} \quad \pi^2 = b_{\text{H}} \cdot \gamma^1 - c \cdot \gamma^2. \tag{6}$$

2.1.2 Same-partner treatment

Setup and strategies. Here, again we consider two pairs of players, players 1 and 2, and players 3 and 4. However, this time each pair engages in both repeated games in parallel. It follows that one player's action in one game may affect the co-player's next action in the other game. Such a situation has been previously referred to as a 'multichannel game' (44). A player's strategy for repeated game k is a 4-tuple $\mathbf{p^k} = (p^k_{CC}, p^k_{CD}, p^k_{DC}, p^k_{DD}) \in [0,1]^4$. An entry $p^k_{a^{\text{H}},a^{\text{L}}}$ refers to the player's cooperation probability in game k, depending on the co-player's previous actions in both the high and the low game. We note that this strategy space contains the strategy space of the control treatment as a special case. For example, a strategy $\mathbf{p}^{\text{H}} = (p^{\text{H}}_C, p^{\text{H}}_D)$ of the control treatment can be represented as $\mathbf{p}^{\text{H}} = (p^{\text{H}}_C, p^{\text{H}}_C, p^{\text{H}}_D, p^{\text{H}}_D)$ within the same-partner treatment. This reflects a case in which a player could, in principle, react to the co-player's previous actions in both games but decides to only react to the actions in the respective game. A strategy for the concurrent game is a pair $\mathbf{p} = (\mathbf{p}^{\text{H}}, \mathbf{p}^{\text{L}}) \in [0,1]^8$ that tells the player what to do in each repeated game.

Calculation of payoffs. As in the first treatment, we compute payoffs with a Markov chain approach. Without loss of generality, consider the first two players, with strategies $\mathbf{p^{1k}}$ and $\mathbf{p^{2k}}$, respectively. Because players no longer treat each game as independent, a state is now a 4-tuple $\mathbf{a} = (a^{1\text{H}}, a^{2\text{H}}, a^{1\text{L}}, a^{2\text{L}})$ with $a^{ik} \in \{C, D\}$. Each state corresponds to a possible outcomes of a single round, with the entries describing the players' actions in each game. For example (C, C, D, C) refers to a state in which both players cooperated in the high game, but only player 2 cooperated in the low game. There are $2^4 = 16$ states in total. For two possible states $\mathbf{a} = (a^{1\text{H}}, a^{2\text{H}}, a^{1\text{L}}, a^{2\text{L}})$ and $\tilde{\mathbf{a}} = (\tilde{a}^{1\text{H}}, \tilde{a}^{2\text{H}}, \tilde{a}^{1\text{L}}, \tilde{a}^{2\text{L}})$, we can compute the transition probability that players move from state $\mathbf{a}$ to $\tilde{\mathbf{a}}$ within one round. The respective transition probability is a product of four factors,

$$m_{\mathbf{a},\tilde{\mathbf{a}}} = y^{1\text{H}}_{a^{2\text{H}}a^{2\text{L}}} \cdot y^{2\text{H}}_{a^{1\text{H}}a^{1\text{L}}} \cdot y^{1\text{L}}_{a^{2\text{H}}a^{2\text{L}}} \cdot y^{2\text{L}}_{a^{1\text{H}}a^{1\text{L}}}. \tag{7}$$

These four factors correspond to the decisions that the two players make in each of the two games. The first factor is the probability that player 1 chooses the action required by $\tilde{\mathbf{a}}$ in the high game,

$$y^{1\text{H}}_{a^{2\text{H}}a^{2\text{L}}} := \begin{cases} p^{1\text{H}}_{a^{2\text{H}}a^{2\text{L}}} & \text{if } \tilde{a}^{1\text{H}} = C \\ 1 - p^{1\text{H}}_{a^{2\text{H}}a^{2\text{L}}} & \text{if } \tilde{a}^{1\text{H}} = D. \end{cases} \tag{8}$$

The three other factors in Eq. [7] are defined analogously. Again, we collect these transition probabilities in a 16×16 matrix $M(\mathbf{p^1}, \mathbf{p^2}) = (m_{\mathbf{a},\tilde{\mathbf{a}}})$. By computing the invariant distribution $\mathbf{v} = (v_{\mathbf{a}})$ of this matrix, we infer how often the two players visit each state $\mathbf{a} = (a^{1\text{H}}, a^{2\text{H}}, a^{1\text{L}}, a^{2\text{L}}) \in \{C, D\}^4$ over the course of the game. Based on this invariant distribution, we compute the average cooperation rate of player i in game k.

To this end, we sum up over all possible states in which player i cooperates in game k,

$$\gamma^{ik} = \sum_{\mathbf{a}\in\{C,D\}^4} v_{\mathbf{a}} \cdot 1_{\{a^{ik}=C\}}. \qquad [9]$$

Here, 1_P is an indicator function. Its value is one if the statement P is true, and it is zero otherwise. Based on these cooperation rates we define the player's payoffs in repeated game $k\in\{\mathtt{H},\mathtt{L}$ as

$$\pi^{1k}=b_k\cdot\gamma^{2k}-c\cdot\gamma^{1k} \quad \text{and} \quad \pi^{2k}=b_k\cdot\gamma^{1k}-c\cdot\gamma^{2k}. \qquad [10]$$

We define the payoffs for the concurrent game by adding up the payoffs for each repeated game,

$$\pi^1=\pi^{1\mathtt{H}}+\pi^{1\mathtt{L}} \quad \text{and} \quad \pi^2=\pi^{2\mathtt{H}}+\pi^{2\mathtt{L}}. \qquad [11]$$

The payoffs for players 3 and 4 can be computed with the same algorithm.

2.1.3 Different-partners treatment

Setup and strategies. In this treatment, players have different interaction partners for the high and the low game. Specifically, we consider four players, with interactions as displayed in **Fig. 1** of the main text. For example, player 1's interaction partner in the high game is player 2, whereas the interaction partner in the low game is player 3. Similar to the same-partner treatment, a player's action in one game may depend on the previous outcome of the other game. That is, a strategy for game k again takes the form $\mathbf{p^k}=(p^k_{CC},p^k_{CD},p^k_{DC},p^k_{DD}) \in [0,1]^4$. Each entry $p^k_{a^{\mathtt{H}},a^{\mathtt{L}}}$ indicates the player's probability to cooperate in game k, depending on the previous action $a^{\mathtt{H}}$ of the co-player in the high game, and on the action $a^{\mathtt{L}}$ of the co-player in the low game. Again, a strategy for the concurrent game is a pair $\mathbf{p}=(\mathbf{p}^{\mathtt{H}},\mathbf{p}^{\mathtt{L}})\in[0,1]^8$. It tells the player what to do in each of the two games.

Calculation of payoffs. As before, payoffs are computed with a Markov chain approach. However, because we no longer have two distinct pairs of players, the state space is yet again bigger. This time, the set of states consist of all 8-tuples $\mathbf{a}=(a^{1\mathtt{H}},a^{2\mathtt{H}},a^{3\mathtt{H}},a^{4\mathtt{H}},a^{1\mathtt{L}},a^{2\mathtt{L}},a^{3\mathtt{L}},a^{4\mathtt{L}})\in\{C,D\}^8$. For example, a tuple $\mathbf{a}$ with $a^{3\mathtt{L}}=C$ refers to a round in which player 3 cooperates in the low game with the respective interaction partner (player 1). Because there are now eight (independent) decisions made each round, the transition probability from state $\mathbf{a}$ to $\tilde{\mathbf{a}}$ is a product of eight factors,

$$m_{\mathbf{a},\tilde{\mathbf{a}}} = y^{1\mathtt{H}}_{a^{2\mathtt{H}}a^{3\mathtt{L}}} \cdot y^{2\mathtt{H}}_{a^{1\mathtt{H}}a^{4\mathtt{L}}} \cdot y^{3\mathtt{H}}_{a^{4\mathtt{H}}a^{1\mathtt{L}}} \cdot y^{4\mathtt{H}}_{a^{3\mathtt{H}}a^{2\mathtt{L}}} \cdot y^{1\mathtt{L}}_{a^{2\mathtt{H}}a^{3\mathtt{L}}} \cdot y^{2\mathtt{L}}_{a^{1\mathtt{H}}a^{4\mathtt{L}}} \cdot y^{3\mathtt{L}}_{a^{4\mathtt{H}}a^{1\mathtt{L}}} \cdot y^{4\mathtt{L}}_{a^{3\mathtt{H}}a^{2\mathtt{L}}}. \qquad [12]$$

The entries $y^{ik}_{a^{\mathtt{H}}a^{\mathtt{L}}}$ are defined the same way as before, see Eq. [8]. The respective transition matrix $M(\mathbf{p^1},\mathbf{p^2},\mathbf{p^3},\mathbf{p^4})=(m_{\mathbf{a},\tilde{\mathbf{a}}})$ that summarizes all these transition probabilities is now of size 256×256 (in particular, the players' payoffs are now more computationally expensive to derive). However, the remaining steps are analogous to the previous treatments. Again, we first compute the players' average cooperation rates in each game by summing up over all relevant states $\mathbf{a}=(a^{1\mathrm{H}}, a^{2\mathrm{H}}, a^{3\mathrm{H}}, a^{4\mathrm{H}}, a^{1\mathrm{L}}, a^{2\mathrm{L}}, a^{3\mathrm{L}}, a^{4\mathrm{L}})$,

$$\gamma^{ik} = \sum_{\mathbf{a}\in\{C,D\}^8} v_{\mathbf{a}} \cdot 1_{\{a^{ik}=C\}}. \qquad [13]$$

Based on these eight cooperation rates, we can compute the players' payoffs in each game as follows,

$$\begin{aligned}
\pi^{1\mathrm{H}} &= b_{\mathrm{H}}\cdot\gamma^{2\mathrm{H}} - c\cdot\gamma^{1\mathrm{H}}, & \pi^{1\mathrm{L}} &= b_{\mathrm{L}}\cdot\gamma^{3\mathrm{L}} - c\cdot\gamma^{1\mathrm{L}},\\
\pi^{2\mathrm{H}} &= b_{\mathrm{H}}\cdot\gamma^{1\mathrm{H}} - c\cdot\gamma^{2\mathrm{H}}, & \pi^{2\mathrm{L}} &= b_{\mathrm{L}}\cdot\gamma^{4\mathrm{L}} - c\cdot\gamma^{2\mathrm{L}},\\
\pi^{3\mathrm{H}} &= b_{\mathrm{H}}\cdot\gamma^{4\mathrm{H}} - c\cdot\gamma^{3\mathrm{H}}, & \pi^{3\mathrm{L}} &= b_{\mathrm{L}}\cdot\gamma^{1\mathrm{L}} - c\cdot\gamma^{3\mathrm{L}},\\
\pi^{4\mathrm{H}} &= b_{\mathrm{H}}\cdot\gamma^{3\mathrm{H}} - c\cdot\gamma^{4\mathrm{H}}, & \pi^{4\mathrm{L}} &= b_{\mathrm{L}}\cdot\gamma^{2\mathrm{L}} - c\cdot\gamma^{4\mathrm{L}}.
\end{aligned} \qquad [14]$$

The overall payoff of player i in the concurrent game is the sum $\pi^i = \pi^{i\mathrm{H}} + \pi^{i\mathrm{L}}$.

2.1.4 Trembling-hand errors

The strategies introduced above describe which actions the players *wish* to take. However, actions may be subject to errors. For example, players may intend to cooperate, but they may fail because of a lack of attention. Similarly, players may misimplement their intended action because of a *trembling hand* (64). The assumption of trembling-hand errors is fairly common in the evolutionary literature on repeated games (9). In addition to making the model more realistic, errors have two useful mathematical implications. First, errors ensure that all finite histories of a repeated game are visited with a positive probability. As a consequence, each entry of a player's strategy has an effect on the player's payoff (there are no entries that are neutral merely because the respective history is never visited). Second, errors ensure that the Markov chains described in the previous sections are ergodic, irrespective of the players' strategies. This implies that payoffs are well-defined even when the unperturbed Markov chain has multiple absorbing states.

We assume trembling-hand errors occur with a constant probability $\varepsilon_{\mathrm{TH}}$, and they affect cooperative and defective actions alike. As a result, a player's *nominal strategy* $\mathbf{p^k}$ for game k translates into an *effective strategy* $\mathbf{p}^{\mathbf{k}}_{\varepsilon}$, with

$$\mathbf{p}^{\mathbf{k}}_{\varepsilon} := (1-\varepsilon_{\mathrm{TH}})\mathbf{p^k} + \varepsilon_{\mathrm{TH}}(\mathbf{1}-\mathbf{p^k}). \qquad [15]$$

Here, $\mathbf{1}$ is a vector that has the same size as $\mathbf{p^k}$, but with all entries being equal to one. As an example, under this transformation, the repeated-game strategy Tit-for-Tat (1,0) in the control treatment is mapped to the effective strategy $(1-\varepsilon_{\mathrm{TH}}, \varepsilon_{\mathrm{TH}})$. More generally, this transformation takes a strategy $\mathbf{p^k} \in [0,1]^n$, and maps it into the interior of this strategy space, $\mathbf{p}^{\mathbf{k}}_{\varepsilon} \in [\varepsilon_{\mathrm{TH}}, 1-\varepsilon_{\mathrm{TH}}]^n$. The above strategy transformation works for all three considered treatments. We can compute the players' (effective) payoffs by simply taking the effective strategies (instead of the nominal strategies) as the input in the respective payoff algorithms.

2.2 Modelling cognitive constraints and alternative behavioral processes

Once players interact in several games in parallel, their behavior may be subject to cognitive constraints that cannot be studied within the classical framework of (independently) repeated games. In the following, we consider some of these constraints, and we discuss how they can be incorporated into our framework for the same-partner treatment and the different-partners treatment, respectively.

2.2.1 Imperfect recall

Motivation. When engaging in several interactions in parallel, people may confuse the outcome of one interaction with the outcome of another. This effect is best documented in studies where participants need to memorize the outcome of different social dilemmas with changing interaction partners (65, 66). As one may expect, these studies find that it is more difficult to correctly recollect one interaction partner's past decision when several other interactions (with different outcomes) occurred in the meanwhile.

Imperfect recall can undermine reciprocal cooperation because it restricts the players' ability to give targeted responses. This may also affect the predictions of our baseline model. As an example, consider a player whose co-player in the high game defected in the previous round, and whose co-player in the low game cooperated. With perfect recall, such a player would react by cooperating with probability $p^{\mathtt{H}}_{DC}$ and $p^{\mathtt{L}}_{DC}$, respectively. With imperfect recall, this player may confuse the outcomes of the last round, and choose to cooperate with probability $p^{\mathtt{H}}_{CD}$ and $p^{\mathtt{L}}_{CD}$ instead. The impact of this cognitive constraint depends on the exact values of the cooperation probabilities. The impact is small when $p^k_{CD} \approx p^k_{DC}$ for both $k \in \{\mathtt{H}, \mathtt{L}\}$. It can be substantial when the cooperation probabilities differ considerably.

Incorporating this kind of imperfect recall into our model is not as straightforward as incorporating trembling-hand errors in Section 2.1.4. There we assumed that trembling-hand errors affect each game of a player independently. An error in one game does not increase or decrease the likelihood that a similar error occurs in the other game. As a result, we can model the effect of errors by simply replacing nominal strategies by effective strategies, see Eq. [15]. In contrast, imperfect recall affects both games simultaneously. In the following, we describe how confusion can be integrated into our framework, both for the same-partner and the different-partners treatment.

Same-partner treatment. We suppose that in any given round, each player may be subject to imperfect recall with probability $\varepsilon_{\text{IR}} \geq 0$. Let $\mathbf{a} = (a^{1\mathtt{H}}, a^{2\mathtt{H}}, a^{1\mathtt{L}}, a^{2\mathtt{L}})$ and $\tilde{\mathbf{a}} = (\tilde{a}^{1\mathtt{H}}, \tilde{a}^{2\mathtt{H}}, \tilde{a}^{1\mathtt{L}}, \tilde{a}^{2\mathtt{L}})$ be the current and the next state, respectively, with $\mathbf{a}, \tilde{\mathbf{a}} \in \{C, D\}^4$. We distinguish four cases.

1. With probability $(1-\varepsilon_{\text{IR}})^2$ no player commits an error. In that case, the transition probability from $\mathbf{a}$ to $\tilde{\mathbf{a}}$ is the same as in the baseline case, $y^{1\mathtt{H}}_{a^{2\mathtt{H}}a^{2\mathtt{L}}} \cdot y^{2\mathtt{H}}_{a^{1\mathtt{H}}a^{1\mathtt{L}}} \cdot y^{1\mathtt{L}}_{a^{2\mathtt{H}}a^{2\mathtt{L}}} \cdot y^{2\mathtt{L}}_{a^{1\mathtt{H}}a^{1\mathtt{L}}}$, with each factor $y^{ik}_{a^{\mathtt{H}}a^{\mathtt{L}}}$ being defined as in Eq. [8].

2. With probability $\varepsilon_{\text{IR}}(1-\varepsilon_{\text{IR}})$ player 1 commits an error but player 2 does not. In that case, the relevant

transition probability is $y^{1\mathtt{H}}_{a^{2\mathtt{L}}a^{2\mathtt{H}}} \cdot y^{2\mathtt{H}}_{a^{1\mathtt{H}}a^{1\mathtt{L}}} \cdot y^{1\mathtt{L}}_{a^{2\mathtt{L}}a^{2\mathtt{H}}} \cdot y^{2\mathtt{L}}_{a^{1\mathtt{H}}a^{1\mathtt{L}}}$ (i.e., the first and the third factor are modified).

3. With the same probability $\varepsilon_{\text{IR}}(1-\varepsilon_{\text{IR}})$ player 2 commits an error but player 1 does not. The respective transition probability is $y^{1\mathtt{H}}_{a^{2\mathtt{H}}a^{2\mathtt{L}}} \cdot y^{2\mathtt{H}}_{a^{1\mathtt{L}}a^{1\mathtt{H}}} \cdot y^{1\mathtt{L}}_{a^{2\mathtt{H}}a^{2\mathtt{L}}} \cdot y^{2\mathtt{L}}_{a^{1\mathtt{L}}a^{1\mathtt{H}}}$ (the second and fourth factor are modified).

4. With probability $\varepsilon^2_{\text{IR}}$ both players commit an error simultaneously. In that case, the transition probability becomes $y^{1\mathtt{H}}_{a^{2\mathtt{L}}a^{2\mathtt{H}}} \cdot y^{2\mathtt{H}}_{a^{1\mathtt{L}}a^{1\mathtt{H}}} \cdot y^{1\mathtt{L}}_{a^{2\mathtt{L}}a^{2\mathtt{H}}} \cdot y^{2\mathtt{L}}_{a^{1\mathtt{L}}a^{1\mathtt{H}}}$ (all factors are modified compared to the baseline).

The overall transition probability from $\mathbf{a}$ to $\tilde{\mathbf{a}}$ is the weighted sum of these four conditional transition probabilities.

Equivalently, we can also write the respective transition matrix more explicitly. For a given repeated-game strategy $\mathbf{p^k} = (p^k_{CC}, p^k_{CD}, p^k_{DC}, p^k_{DD})$, define a perturbed strategy by $\tilde{\mathbf{p}}^\mathbf{k} := (p^k_{CC}, p^k_{DC}, p^k_{CD}, p^k_{DD})$ (i.e. the second and third entry change their position). For a concurrent-game strategy $\mathbf{p} = (\mathbf{p}^{\mathtt{H}}, \mathbf{p}^{\mathtt{L}})$, define $\tilde{\mathbf{p}} := (\tilde{\mathbf{p}}^{\mathtt{H}}, \tilde{\mathbf{p}}^{\mathtt{L}})$ as the strategy with perturbed components. Based on this notation, let $\mathbf{p^1}$ and $\mathbf{p^2}$ be the strategies of the two players in the same-partner treatment. Then the respective transition matrix for the case of imperfect recall can be written as

$$M_{\text{IR}} = (1-\varepsilon_{\text{IR}})^2 \cdot M(\mathbf{p^1}, \mathbf{p^2}) + \varepsilon_{\text{IR}}(1-\varepsilon_{\text{IR}})\Big(M(\tilde{\mathbf{p}}^\mathbf{1}, \mathbf{p^2}) + M(\mathbf{p^1}, \tilde{\mathbf{p}}^\mathbf{2})\Big) + \varepsilon^2_{\text{IR}} \cdot M(\tilde{\mathbf{p}}^\mathbf{1}, \tilde{\mathbf{p}}^\mathbf{2}). \qquad [16]$$

The matrices $M(\mathbf{x}, \mathbf{y})$ on the right hand side are defined as in the baseline model, Eq. [7]. In particular, as errors vanish, $\varepsilon_{\text{IR}} \to 0$, the matrix M_{IR} recovers the baseline transition matrix. For positive error rates, we can use the matrix M_{IR} in Eq. [16] to compute the invariant distribution, the players' average cooperation rates, and their expected payoffs as before.

Different-partners treatment. The logic of integrating imperfect recall into the different-partners treatment is analogous to the same-partner treatment. However, because there are now four players, there are more cases to consider. In a given round, the number of players who are subject to imperfect recall may be any number between zero and four.

Using the same notation as before, the respective transition matrix becomes

$$\begin{aligned}
M_{\text{IR}} = \;& (1-\varepsilon_{\text{IR}})^4 \quad M(\mathbf{p^1}; \mathbf{p^2}; \mathbf{p^3}; \mathbf{p^4}) \\
&+ \varepsilon_{\text{IR}}(1-\varepsilon_{\text{IR}})^3 \Big(M(\tilde{\mathbf{p}}^\mathbf{1}; \mathbf{p^2}; \mathbf{p^3}; \mathbf{p^4}) + M(\mathbf{p^1}; \tilde{\mathbf{p}}^\mathbf{2}; \mathbf{p^3}; \mathbf{p^4}) + M(\mathbf{p^1}; \mathbf{p^2}; \tilde{\mathbf{p}}^\mathbf{3}; \mathbf{p^4}) + M(\mathbf{p^1}; \mathbf{p^2}; \mathbf{p^3}; \tilde{\mathbf{p}}^\mathbf{4})\Big) \\
&+ \varepsilon^2_{\text{IR}}(1-\varepsilon_{\text{IR}})^2 \Big(M(\tilde{\mathbf{p}}^\mathbf{1}; \tilde{\mathbf{p}}^\mathbf{2}; \mathbf{p^3}; \mathbf{p^4}) + M(\tilde{\mathbf{p}}^\mathbf{1}; \mathbf{p^2}; \tilde{\mathbf{p}}^\mathbf{3}; \mathbf{p^4}) + M(\tilde{\mathbf{p}}^\mathbf{1}; \mathbf{p^2}; \mathbf{p^3}; \tilde{\mathbf{p}}^\mathbf{4}) \\
&\qquad + M(\mathbf{p^1}; \tilde{\mathbf{p}}^\mathbf{2}; \tilde{\mathbf{p}}^\mathbf{3}; \mathbf{p^4}) + M(\mathbf{p^1}; \tilde{\mathbf{p}}^\mathbf{2}; \mathbf{p^3}; \tilde{\mathbf{p}}^\mathbf{4}) + M(\mathbf{p^1}; \mathbf{p^2}; \tilde{\mathbf{p}}^\mathbf{3}; \tilde{\mathbf{p}}^\mathbf{4})\Big) \\
&+ \varepsilon^3_{\text{IR}}(1-\varepsilon_{\text{IR}}) \Big(M(\tilde{\mathbf{p}}^\mathbf{1}; \tilde{\mathbf{p}}^\mathbf{2}; \tilde{\mathbf{p}}^\mathbf{3}; \mathbf{p^4}) + M(\tilde{\mathbf{p}}^\mathbf{1}; \tilde{\mathbf{p}}^\mathbf{2}; \mathbf{p^3}; \tilde{\mathbf{p}}^\mathbf{4}) + M(\tilde{\mathbf{p}}^\mathbf{1}; \mathbf{p^2}; \tilde{\mathbf{p}}^\mathbf{3}; \tilde{\mathbf{p}}^\mathbf{4}) + M(\mathbf{p^1}; \tilde{\mathbf{p}}^\mathbf{2}; \tilde{\mathbf{p}}^\mathbf{3}; \tilde{\mathbf{p}}^\mathbf{4})\Big) \\
&+ \quad \varepsilon^4_{\text{IR}} \quad M(\tilde{\mathbf{p}}^\mathbf{1}; \tilde{\mathbf{p}}^\mathbf{2}; \tilde{\mathbf{p}}^\mathbf{3}; \tilde{\mathbf{p}}^\mathbf{4})
\end{aligned} \qquad [17]$$

Again, given this transition matrix, it is straightforward to compute the players' cooperation rates and their payoffs, as described by Eqs. [13] and [14].

2.2.2 Behavioral spillovers

Motivation. Behavioral spillovers occur when an individual's action in one domain leads that individual to take a similar action in a different domain. Such spillovers may occur consciously or subconsciously, and they can have important policy implications (67, 68). In the context of our framework, spillovers induce additional correlations in a player's behavior across the two games. For any given history, spillovers increase the chance that a player cooperates in both games (rather than cooperating in one game and defecting in the other). Similarly, they also increase the chance that players defect in both games. In the following we describe how (exogenous) behavioral spillovers can be integrated into our framework.

Same-partner treatment. We assume spillovers occur with a constant probability ε_{SP} and they affect each player independently. Moreover, we assume that both cooperative and defective actions are equally likely to spill over to a different context. To describe the effects of such spillovers formally, we consider two players with strategies $\mathbf{p^1}$ and $\mathbf{p^2}$. We aim to describe the probability that they make the transition from state $\mathbf{a}=(a^{1\text{H}}, a^{2\text{H}}, a^{1\text{L}}, a^{2\text{L}}) \in \{C, D\}^4$ in one round to state $\tilde{\mathbf{a}}=(\tilde{a}^{1\text{H}}, \tilde{a}^{2\text{H}}, \tilde{a}^{1\text{L}}, \tilde{a}^{2\text{L}}) \in \{C, D\}^4$ in the next round. Again we need to distinguish several cases. For example, there are different ways that would lead player 1 to cooperate in both games in the next round:

1. Player 1 may decide to cooperate in both games from the outset. Given the previous actions of the second player, this happens with probability $p^{1\text{H}}_{a^{2\text{H}}a^{2\text{L}}} \cdot p^{1\text{L}}_{a^{2\text{H}}a^{2\text{L}}}$.
2. Player 1 initially decides to cooperate in the high game but to defect in the other; but due to a spillover, the player ends up cooperating in both games. The respective probability is $\frac{\varepsilon_{\text{SP}}}{2} \cdot p^{1\text{H}}_{a^{2\text{H}}a^{2\text{L}}}(1-p^{1\text{L}}_{a^{2\text{H}}a^{2\text{L}}})$. The factor of one half indicates that spillovers could equally lead the player to defect in both games.
3. Similarly, player 1 may wish to defect in the high game, to cooperate in the low game, but ends up cooperating on both games due to a spillover. This probability is $\frac{\varepsilon_{\text{SP}}}{2} \cdot (1-p^{1\text{H}}_{a^{2\text{H}}a^{2\text{L}}})p^{1\text{L}}_{a^{2\text{H}}a^{2\text{L}}}$.
4. If player 1 wishes to defect in both games, spillovers cannot affect the player's decision.

Overall, in the presence of spillovers, the transition probability from state $\mathbf{a}$ to $\tilde{\mathbf{a}}$ takes the form

$$m_{\mathbf{a},\tilde{\mathbf{a}}} = z^1_{a^{2\text{H}}a^{2\text{L}}} \cdot z^2_{a^{1\text{H}}a^{1\text{L}}}. \qquad [18]$$

Here the two variable $z^i_{a^{\text{H}}a^{\text{L}}}$ for $i \in \{1, 2\}$ describe the decisions of the two players in both the high and the low game. For example, for the first variable,

$$z^{1}_{a^{2H},a^{2L}} := \begin{cases} p^{1H}_{a^{2H}a^{2L}} p^{1L}_{a^{2H}a^{2L}} + \frac{\varepsilon_{SP}}{2}\Big[p^{1H}_{a^{2H}a^{2L}}(1-p^{1L}_{a^{2H}a^{2L}}) + (1-p^{1H}_{a^{2H}a^{2L}})p^{1L}_{a^{2H}a^{2L}}\Big] & \text{if } \tilde{a}^{1H}=C,\ \tilde{a}^{2H}=C \\ (1-\varepsilon_{SP})p^{1H}_{a^{2H}a^{2L}}(1-p^{1L}_{a^{2H}a^{2L}}) & \text{if } \tilde{a}^{1H}=C,\ \tilde{a}^{2H}=D \\ (1-\varepsilon_{SP})(1-p^{1H}_{a^{2H}a^{2L}})p^{1L}_{a^{2H}a^{2L}} & \text{if } \tilde{a}^{1H}=D,\ \tilde{a}^{2H}=C \\ (1-p^{1H}_{a^{2H}a^{2L}})(1-p^{1L}_{a^{2H}a^{2L}}) + \frac{\varepsilon_{SP}}{2}\Big[p^{1H}_{a^{2H}a^{2L}}(1-p^{1L}_{a^{2H}a^{2L}})+(1-p^{1H}_{a^{2H}a^{2L}})p^{1L}_{a^{2H}a^{2L}}\Big] & \text{if } \tilde{a}^{1H}=D,\ \tilde{a}^{2H}=D \end{cases} \quad [19]$$

The variable $z^{2}_{a^{1H},a^{1L}}$ for the second player is defined analogously. In the limiting case that exogenous spillovers are rare, $\varepsilon_{SP} = 0$, the transition probability [18] simplifies to the formula [7] of the baseline model, which is reassuring. However, as ε_{SP} increases, each player becomes increasingly unlikely to choose different actions in the two games. In particular, in the limiting case $\varepsilon_{SP} = 1$, there is a perfect correlation between a player's actions. Each player either cooperates in both games, or defects in both games.

By collecting the transition probabilities [18] and writing them as a matrix $M_{SP} = (m_{\mathbf{a},\tilde{\mathbf{a}}})$, we can again use the methods from the baseline model to compute expected cooperation rates and payoffs.

Different-partners treatment. Spillovers can be incorporated in the same way as in the same-partners treatment. If $\mathbf{p^1}$, $\mathbf{p^2}$, $\mathbf{p^3}$, $\mathbf{p^4}$ are the strategies of the four players, and $\mathbf{a}$ and $\tilde{\mathbf{a}}$ are the current and the next state, respectively, the transition probability takes the form

$$m_{\mathbf{a},\tilde{\mathbf{a}}} = z^{1}_{a^{2H}a^{3L}} \cdot z^{2}_{a^{1H}a^{4L}} \cdot z^{3}_{a^{4H}a^{1L}} \cdot z^{4}_{a^{3H}a^{2L}}. \quad [20]$$

The individual factors $z^{i}_{a^{H}a^{L}}$ are defined analogously as in the previous section, see Eq. [19]. Based on the resulting transition matrix $M_{SP} = (m_{\mathbf{a},\tilde{\mathbf{a}}})$, we can again compute cooperation rates and payoffs.

2.2.3 Preferential updating

Motivation. The two previous subsections dealt with two plausible kinds of errors. These errors affect how people with given strategies act in a concurrent game. The next two subsections describe two behaviorally plausible modifications of the learning process. These modifications affect how people choose their strategies in the first place. The baseline model assumes that in each step of the learning process, players are equally likely to update their strategy for the high game and for the low game. In the following, we discuss a model extension that allows for preferential updating. Here, players are more likely to update the strategy in the game that currently yields the lower relative payoff.

Implementation of the learning process. We incorporate preferential updating as follows. As in the baseline model, we assume that in any given time step t, one individual, say player j, is picked at random. This player is then given a chance to generate an alternative strategy, by either modifying their strategy in the high or the low game. However, in this case, the high-game strategy is modified with a probability given by the Fermi-function,

$$\xi^j = \frac{1}{1+\exp[\kappa(\frac{\pi^{j\mathrm{H}}}{b_\mathrm{H}} - \frac{\pi^{j\mathrm{L}}}{b_\mathrm{L}})]}. \qquad [21]$$

The parameter $\kappa \geq 0$ measures to which extent updating is biased towards one game or the other. For $\kappa = 0$, we obtain $\xi^j = 1/2$, irrespective of the value of the other variables. In that case, we recover the baseline model where both games are equally likely to be modified. For positive κ, the player is more likely to modify the strategy that currently yields a lower payoff, relative to the maximum achievable payoff b_k in the respective game. Once an alternative strategy is generated, the further process is the same as described in the main text. As described there, the player is more likely to adopt the generated strategy $\tilde{\mathbf{p}}^{\mathbf{j}}$, the higher the hypothetical payoff $\tilde{\pi}^j$ is compared to the player's current payoff π^j. If $\tilde{\mathbf{p}}^{\mathbf{j}}$ yields a high payoff, it is likely to be adopted. If $\tilde{\mathbf{p}}^{\mathbf{j}}$ yields a low payoff, it is likely to be discarded and the player continues to use their current strategy $\mathbf{p}^{\mathbf{j}}$.

2.2.4 Updating under narrow bracketing

Motivation. Narrow bracketing refers to situations in which people make decisions in one domain without fully taking into account its consequences on decisions in other domains (69). Such a bias may also unfold in concurrent games. Players who update their strategies for a given repeated game $k \in \{\mathtt{H}, \mathtt{L}\}$ may neglect the impact of this strategy change on the other game. As a result, a strategy change that is beneficial in the respective game may be detrimental overall. In the following, we describe a simple model of narrow bracketing in the context of concurrent games.

Implementation of the learning process. The learning process under narrow bracketing follows along the same lines as the baseline learning process described in the main text. As before, a randomly chosen focal player updates its strategy in a randomly chosen game $k \in \{\mathtt{H}, \mathtt{L}\}$. When deciding whether to adopt the new concurrent strategy $\tilde{\mathbf{p}}^{\mathbf{j}}$, with probability $1 - \lambda$, the player compares the total payoffs $\tilde{\pi}^j$ and π^j. With the converse probability λ, however, the player only compares the payoffs $\tilde{\pi}^{jk}$ and π^{jk} of the particular game k in which the strategy $\tilde{\mathbf{p}}^j$ got updated. In particular, a player may adopt the alternative strategy because $\tilde{\pi}^{jk} > \pi^{jk}$ even though the total payoff effect is negative, $\tilde{\pi}^j = \tilde{\pi}^{j\mathrm{H}} + \tilde{\pi}^{j\mathrm{L}} < \pi^{j\mathrm{H}} + \pi^{j\mathrm{L}} = \pi^j$. The total effect may be negative because in concurrent games, a player's strategy in one game can affect the co-player's response in the other game. We refer to λ as the probability of narrow bracketing. When $\lambda = 0$, decisions are made as in the baseline model. As λ increases, updating decisions are increasingly based on the payoff effects in repeated game k only.

3 Experimental methods

3.1 Experimental procedures

Experimental setup. Our experiments has three different treatments, matching those of the theoretical model **Fig. 1**. The treatments differ in the number of games played simultaneously and in the number of co-players. The control treatment is a standard 2-player game. The same-partner treatment is also a 2-player game, but now both players interact through two games simultaneously. The different-partners treatment also involves two games, but now each participant interacts with two different co-players, one for each game.

Treatment	**Number of games**	**Number of co-players**
Control	1	1
Same-Partner	2	1
Different-Partner	2	2

Each stage game is a donation game. Players choose between paying a cost to send a benefit to the co-player or do nothing. The games differ only in the benefit b sent to the other player; the cost c remains constant. In the high game, $b = 4$ points and $c = 2$ points, whereas in the low game, $b = 3$ points and $c = 2$ points.

Participants interacted for a minimum of 20 rounds with their respective co-player(s). To avoid end-game effects, after the 20th round, each subsequent round had a 50% chance of occurring. Participants in the multi-game treatments (same-partner and different-partners) made decisions for both games simultaneously (i.e. on the same page) on every round. The decisions for each game were separately elicited from each other. Participants were told what the other player(s) had chosen after every round, and they were reminded of their co-players' previous decisions when making their next one. Once all rounds had been played, each participant was informed of their total payoff across the whole game in points as well as the converted amount (in GBP). They were asked to fill a demographics form, were thanked for their time and informed that the money would be paid to them by the lab manager at a future date. The games are built with the python package oTree (70) and ran online. oTree was developed specifically to run real-time interactive sessions with multiple participants.

Participant recruitment. We recruited 363 participants between April and October 2021 from the University of Exeter pool of subjects. Participants received a baseline payment of £3 for participating in our experiment. In addition, they were able to receive a bonus payment based on their performance during the experiment. To earn this bonus, participant accumulated points from one or two games that were converted to British pounds (£) at the end. One point was worth £1/75 (or 20 points = £0.26) and the average bonus earned was £1.66, £1.62, and £0.88 in the Different-Partners, Same-partner and Control treatments, respectively. The sample size was estimated from a rough power analysis and based on previous similar research, given constraints on the size of the FEELE subject pool. The student pool showed lower dropout rates than most online recruiting platforms. The experiment was approved by the faculty of Medicine of the Christian-Albrechts-Univeristy in Kiel (ID number D 571/20)

Experimental procedures. The design of the experiment was between-subjects, meaning that participants only took part in one treatment. Fifteen sessions of up to 32 participants were run sequentially. Each session was for one treatment only. Participants signed up for an open slot on the FEELE platform of the University of Exeter without prior knowledge of the experiment or treatment. Ten minutes prior to the start of the session, the experimenter opened a room in oTree so that participants could connect and wait until everyone had arrived. Once the room was full, the experimenter started the study. Once participants entered the study, the first page was a consent form describing their rights. Clicking on the button to continue was explained to mean they gave their consent to take part in the study. Then the instructions were provided along with a series of comprehension questions. If they gave a wrong answer, a message appeared on the screen asking them to reread the instructions and change their answer. They could only continue once they gave the correct answer to each question. After these steps were completed, participants entered a virtual waiting room.

Once at least four participants were in the virtual waiting room, they would be paired together to form a group and proceed to the main task. Players were always grouped together in fours, but they only play as an interactive group of four in the different-partners treatment. This constraint comes from the design of the different-partners treatment, but to ensure consistency in waiting times throughout the experiment, it was implemented for all treatments. In the other two treatments, two pairs go through the game at the same time but are unaware of the other pair's existence. Participants waited for a maximum of seven minutes. If not enough players showed up in that timeframe, participants were considered unmatched and instructed to leave the study and receive their participation fee, but no bonus. If after pairing one player in a group dropped out, all others were informed and asked to leave the study. They were paid a fixed bonus of £1.

3.2 Statistical methods

For our data analysis, we only considered groups where all participants completed their total number of rounds, resulting in 316 subjects in 79 groups. To increase power, in the same-partner treatment, we separate the two independent pairs into subgroups. We use these smaller denominations for our group level analysis. This means we have 36 groups of 2 for the same-partner treatment, 32 groups of 4 for the different partners treatment, and 29 groups of 4 for the control. Since all considered groups played at least 20 rounds but differed in the total number of rounds, we used only rounds 1-20 for the main analysis. All averages are aggregated at the group level for all our analysis so that we get the average cooperation across all rounds of interest and all group participants.

3.3 Additional results

We looked at cooperative decision on the first round only. Given that cooperation levels in the two multi-game treatment do not differ significantly despite difference in conditional behavior, we wanted to see if this difference was already significant on the first round. Cooperation rates are already significantly lower in the first round for the same-partner treatment compared to the control ($p = 0.021$) and close to significant for the different-partners treatment compared to control ($p = 0.052$). There is no difference between the same-partner and different-partners treatments ($p = 0.902$) (see **Fig. 4**.)

4 Supplementary Figures

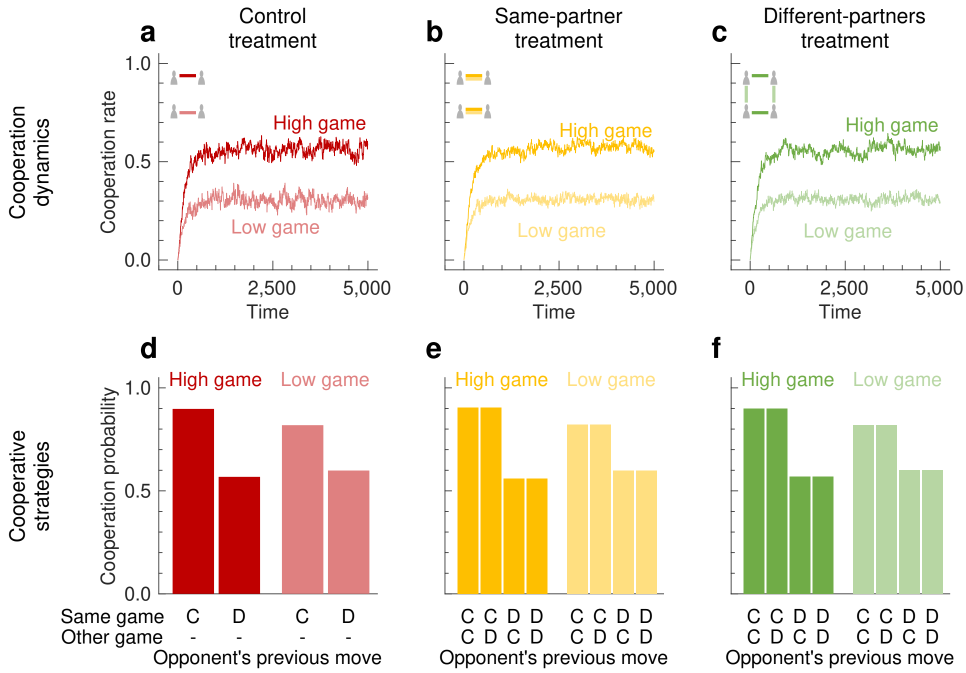


Figure S1: Learning dynamics among players who treat each game as independent. This figure uses the same setup and the same parameters as in **Fig. 2**. However, this time, players in the same-partner and the different-partners treatment are artificially restricted to treat each repeated game as independent. That is, for these simulations, we only allow those mutant strategies that satisfy Eq. [3] in the main text. In this case, all three treatments are equivalent. They lead to the same cooperation dynamics, and they generate the same cooperative strategies.

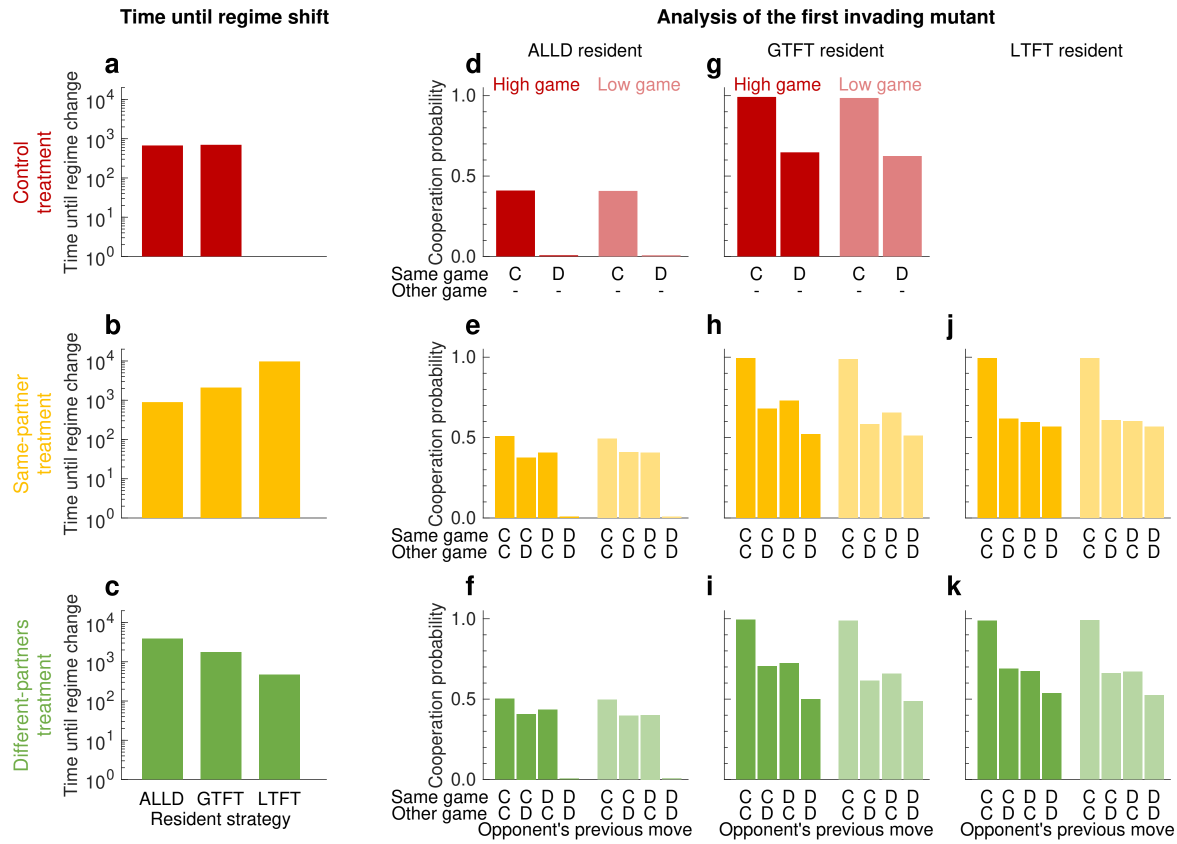


Figure S2: Invasion dynamics across the three treatments. To gain further insights into the learning dynamics, we have run simulations assuming that players start out with predefined strategies. Here, we allow for three such strategies. `ALLD` cooperates with probability 0.01 each round, corresponding to a noisy variant of always defect. `GTFT` treats each game as independent and cooperates with probability 0.99 after cooperation, and 0.50 after defection. `LTFT` corresponds to a version of `GTFT` with linkage. It cooperates with probability 0.99 if the co-player previously cooperated in both games; otherwise it cooperates with with probability 0.50. Note that this strategy only exists in the same-partner and the different-partners treatment. Initially, all players are assumed to adopt the same strategy (we refer to the strategy as the 'resident'). Then we simulate the learning dynamics until there has been a 'regime shift' (meaning that the average cooperation rate among the four players exceeds 50% after starting from `ALLD`, or that it falls below 50% after starting from `GTFT` or `LTFT`). We record the time it takes for a regime shift to occur (panels **a–c**). In addition, we record the first strategy that a player adopts instead of the resident strategy (to which we refer as the 'first invading mutant', panels **d–k**). In the same-partner treatment, `LTFT` is the most robust resident strategy with respect to regime shifts. In the different-partners treatment, the most robust strategy is `ALLD`. In general, we find that `ALLD` is typically invaded by conditionally cooperative strategies (which show a very small cooperation probability after full defection). In contrast, both `GTFT` and `LTFT` are typically invaded by strategies similar to `LTFT` in the last two treatments. All bars represent averages of 1,000 independent simulations, using the parameters of **Fig. 2**.

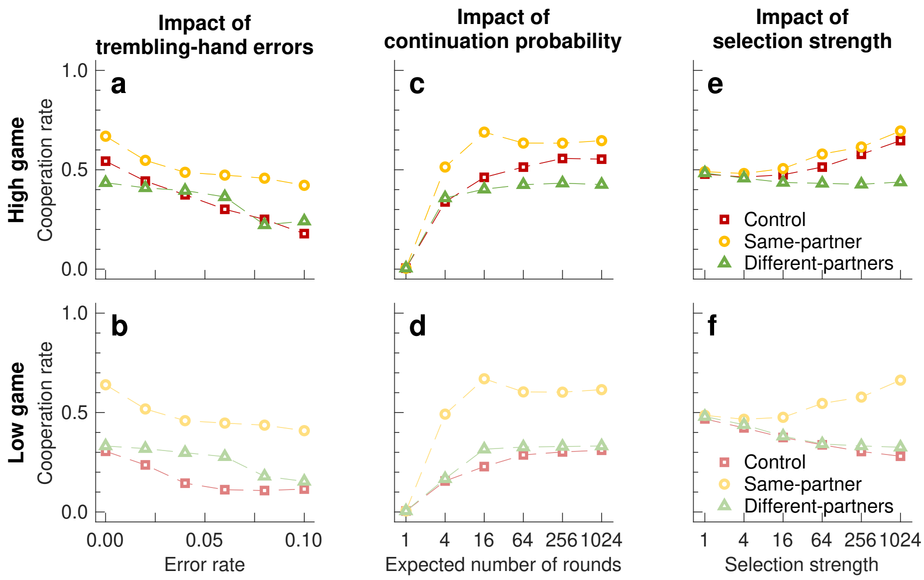


Figure S3: Robustness of evolutionary results. To explore the robustness of our results, we systematically vary three key parameters. **a,** In previous figures, we have assumed that players always perfectly implement their intended actions. Here, we explore the impact of trembling-hand errors, see Section 2.1.4. Overall, such errors make it more difficult to sustain cooperation. **b,** In our previous theoretical analysis, we have assumed for simplicity that the game is infinitely repeated. However, it is straightforward to analyze repeated games in which after each round, the game only continues with a probability $\delta < 1$, see for example Ref. (71). Here, we report the corresponding simulation results, as a function of the expected number of rounds, $1/(1-\delta)$. As one may expect, players become more cooperative when there are more rounds. **c,** Finally, we have also varied the selection strength of the evolutionary process (in the main text we use $\beta = 200$). The stronger selection, the more we see a difference between the three treatments. Across all conditions, our results suggest that one should expect most cooperation in the same-partner treatment, in line with Ref. (44). Parameters are the same as in **Fig. 2**.

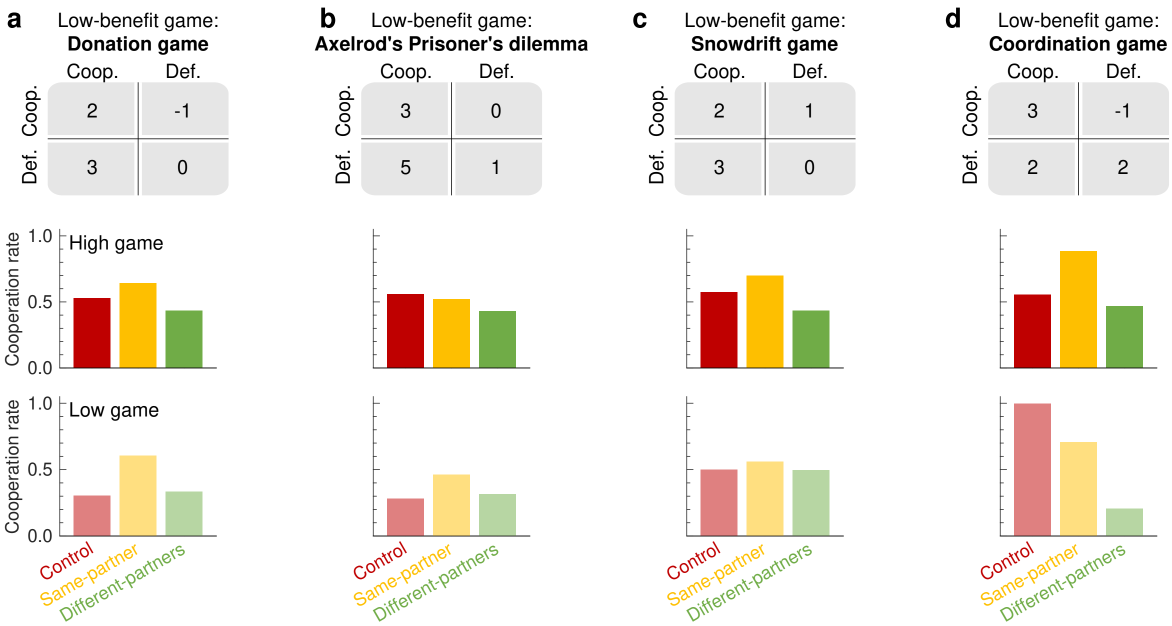


Figure S4: Concurrent games beyond the prisoner's dilemma. a, In our previous analysis, we assumed players engage in different donation games. The respective result of **Fig. 2** is reproduced here. In addition, we consider three cases in which the low-benefit game is replaced by **b,** the payoffs used by Axelrod (23), **c,** a snowdrift game, and **d,** a coordination game. Across all games, we observe that the same partner-treatment should result in more cooperation than the different-partners treatment. Yet in the case of a coordination game, the control treatment seems to be best able to single out the most efficient equilibrium. Except for the low-game payoffs, all parameters are as in **Fig. 2**.

Instructions

You are **Participant 1** in this study. The other participants are **Participant 2** and **Participant 3**.

In each round, you and Participant 2 will simultaneously interact with each other in Task A, and you and Participant 3 will simultaneously interact with each other in Task B. In each task, you and your co-participant on that task must choose between two options. The points you earn in each round will depend on what you, Participant 2 and Participant 3 independently decide for your respective tasks. The two tasks are independent: The points you earn in either task do not depend on the decisions made in the other. In the beginning of each round, you and the other two participants receive an endowment of 2 points in task A and 2 points in task B.

Below you can see the two options that **you and Participant 2** can choose from in Task A in each round.

Task A:

- You **lose 2** pts for Participant 2 to **receive 4** pts
- You **lose 0** pts for Participant 2 to **receive 0** pts

Participant 2 faces exactly the same Task A with **you** as the recipient. You and Participant 2 have to choose simultaneously between these two options.

Below you can see the two options that that **you and Participant 3** can choose from in Task B in each round.

Task B:

- You **lose 2** pts for Participant 3 to **receive 3** pts
- You **lose 0** pts for Participant 3 to **receive 0** pts

Participant 3 faces exactly the same Task B with **you** as the recipient. You and Participant 3 have to choose simultaneously between these two options.

Note that, while Participant 2 and Participant 3 interact with you in Tasks A and B, they both also simultaneously interact with another Participant (but not with each other) in a similar task. That is, each of you is engaged in two tasks simultaneously at all times.

The session will last for multiple rounds. There will be at least 20 rounds. After the 20th round, there will be a 50% chance of another round. After that round, there will again be a 50% chance of another round, and after that round, there will again be another 50% chance of another round, and so on until the tasks end.

You and the other participants are both endowed with 80 points at the start of the game (2 points for Task A and 2 points for Task B, for each of the 20 rounds). For any additional round that may occur after the 20th round, you receive an additional endowment of 2 points for Task A and 2 points for Task B.

Please answer the following questions to continue:

In Task A, what amount will you receive from Participant 2 if they choose to pay 2 pts:

- ○ You will earn 0 pts.
- ○ You will earn 2 pts.
- ● You will earn 4 pts.

What are the chances that there will be another round after the 20th round?

- ○ 10%
- ● 50%
- ○ 100%

What are the chances that there will be another round after the 21th round?

- ○ 10%
- ● 50%
- ○ 100%

Next

Example

Recall the options for Task A and Task B:

Task A:

(played with Participant 2)

- You **lose 2 pts** for Participant 2 to **receive 4 pts**.
- You **lose 0 pts** for Participant 2 to **receive 0 pts**.

Task B:

(played with Participant 3)

- You **lose 2 pts** for Participant 3 to **receive 3 pts**.
- You **lose 0 pts** for Participant 3 to **receive 0 pts**.

Example

Consider a scenario where Participant 1 and Participant 2 made the following decisions:

Task A:

(played with Participant 2)

Participant 1 chose:

- to **lose 2 pts** for Participant 2 to **receive 4 pts**

Participant 2 chose:

- to **lose 0 pts** for Participant 1 to **receive 0 pts**

Participant 1's points in this round: -2 + 0 = **-2**

Participant 2's points in this round: 0 + 4 = **4**

Consider a scenario where Participant 1 and Participant 3 made the following decisions:

Task B:

(played with Participant 3)

Participant 1 chose:

- to **lose 2 pts** for Participant 3 to **receive 3 pts**

Participant 3 chose:

- to **lose 2 pts** for Participant 1 to **receive 3 pts**

Participant 1's points in this round: -2 + 3 = **1**

Participant 3's points in this round: -2 + 3 = **1**

In total across both tasks, Participant 1 earned **-1 pts**.

Please answer the following questions to continue:

Across both tasks, how many points did Participant 1 earn in total?

- (●) -1 points
- ( ) 1 points
- ( ) 5 points

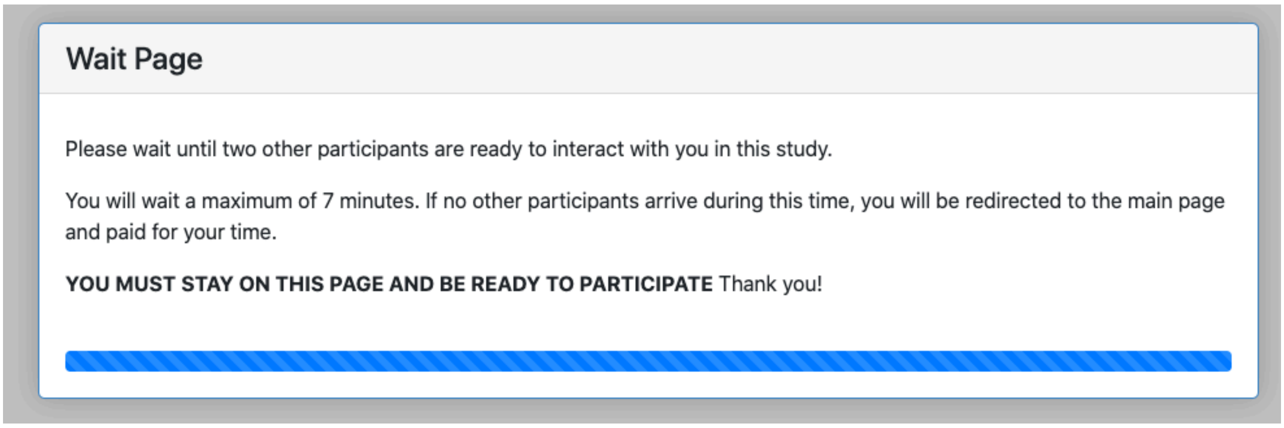
In Task A, how many points did Participant 2 earn?

- ( ) 0 points
- ( ) 2 points
- (●) 4 points

In Task B, how many points did Participant 3 earn?

- ( ) -2 points
- (●) 1 points
- ( ) 3 points

Next



Wait Page

Please wait until two other participants are ready to interact with you in this study.

You will wait a maximum of 7 minutes. If no other participants arrive during this time, you will be redirected to the main page and paid for your time.

YOU MUST STAY ON THIS PAGE AND BE READY TO PARTICIPATE Thank you!

Round Decision

In the previous round:

Task A:
(played with Participant 2)

You chose:

to **pay 2 pts** for Participant 2 to **receive 4 pts**.

Participant 2 chose:

to **pay 2 pts** for you to **receive 4 pts**.

Task B:
(played with Participant 3)

You chose:

to **pay 2 pts** for Participant 3 to **receive 3 pts**.

Participant 3 chose:

to **pay 0 pts** for you to **receive 0 pts**

In this round:

Below are your two options for **Task A** and **Task B**, respectively. You need to make a choice in each task to continue.

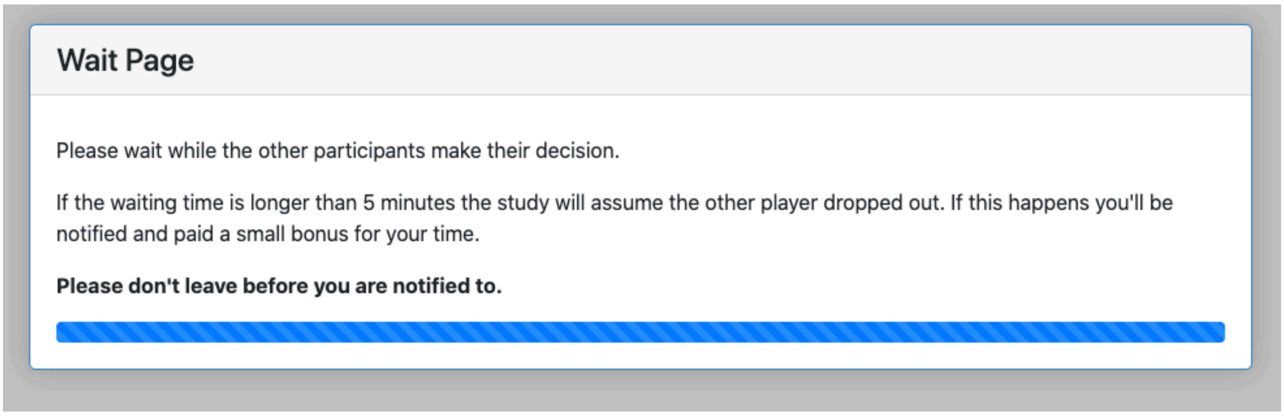
Task A:
(played with Participant 2)

- (●) You lose 2 pts for Participant 2 to receive 4 pts.
- ( ) You pay 0 pts for Participant 2 to receive 0 pts.

Task B:
(played with Participant 3)

- ( ) You lose 2 pts for Participant 3 to receive 3 pts.
- (●) You pay 0 pts for Participant 3 to receive 0 pts.

Next


Wait Page
Please wait while the other participants make their decision.
If the waiting time is longer than 5 minutes the study will assume the other player dropped out. If this happens you'll be notified and paid a small bonus for your time.
Please don't leave before you are notified to.

Round Results

Task A:

(played with Participant 2)

You chose:

- to **lose 2 pts** for Participant 2 to **receive 4 pts**.

Participant 2 chose:

- to **lose 2 pts** for you to **receive 4 pts**.

Your points in this round: -2 + 4 = **2**

Task B:

(played with Participant 3)

You chose:

- to **lose 0 pts** for Participant 3 to **receive 0 pts**.

Participant 3 chose:

- to **lose 0 pts** for you to **receive 0 pts**.

Your points in this round: 0 + 0 = **0**

Next

Final results

Task A:

Round	My outcome	Opponent's outcome
1	4 pts	4 pts
2	4 pts	4 pts
3	4 pts	4 pts
4	4 pts	4 pts
5	4 pts	4 pts
6	4 pts	4 pts
7	4 pts	4 pts
8	4 pts	4 pts
9	4 pts	4 pts
10	4 pts	4 pts
11	4 pts	4 pts
12	4 pts	4 pts
13	4 pts	4 pts
14	4 pts	4 pts
15	4 pts	4 pts
16	4 pts	4 pts
17	4 pts	4 pts
18	4 pts	4 pts
19	4 pts	4 pts
20	4 pts	4 pts
21	4 pts	4 pts
22	4 pts	4 pts

In Task A, across all rounds, you earned a total of **88** points.

Task B:

Round	My outcome	Opponent's outcome
1	0 pts	5 pts
2	3 pts	3 pts
3	3 pts	3 pts
4	3 pts	3 pts
5	3 pts	3 pts
6	3 pts	3 pts
7	3 pts	3 pts
8	3 pts	3 pts
9	3 pts	3 pts
10	3 pts	3 pts
11	3 pts	3 pts
12	3 pts	3 pts
13	3 pts	3 pts
14	3 pts	3 pts
15	3 pts	3 pts
16	3 pts	3 pts
17	3 pts	3 pts
18	3 pts	3 pts
19	3 pts	3 pts
20	3 pts	3 pts
21	3 pts	3 pts
22	3 pts	3 pts

In Task B, across all rounds, you earned a total of **63** points.

Next

Demographic Questions

Please answer the following questions. These are used for demographic purposes only. This information will not be associated with your name. You remain anonymous to all the other players and the experimenters.

What is your age?

What gender do you identify as?

- ○ Female
- ○ Male
- ○ Other

What is the total combined income of your household?

- ○ £9.999 or below
- ○ £10.000 - £29.999
- ○ £30.000 - £49.999
- ○ £50.000 - £69.999
- ○ £70.000 - £89.999
- ○ £90.000 or over
- ○ Prefer not to say

What is the highest level of education you have completed?

- ○ No formal education
- ○ GCSE or equivalent
- ○ A-Levels or equivalent
- ○ Vocational training
- ○ Undergraduate degree
- ○ Postgraduate degree
- ○ Prefer not to say

What is your ethnicity?

- ○ Asian/Asian British
- ○ Black/African/Caribbean/Black British
- ○ Mixed/Multiple Ethnic groups
- ○ White
- ○ Other

Next

Quick Survey

We appreciate feedback. If you have any comments on the implementation of this study, or on the user interface, please write it down below.

Next

The End of the Study

The study is now over. **Thank you for participating!** Below is your total score and the money you earned.

Total points earned in all rounds across both tasks	88 + 68 = 156
Conversion rate	15pts = £0.20
Total of the bonus in GBP	£2.08
Participation payment	£3.00

Final Payment to you (participation payment + bonus) £5.08

You can now close your browser.

You will be paid directly into your student bank account by the University of Exeter Finance Department. Please ensure your bank account details are up to date in your SRS account. Payment usually takes 1 to 2 weeks to come through.

	Dependent variable:	
	Cooperation	
	High game	Low game
Partner's previous decision in the high game	0.673*** (0.043)	
Partner's previous decision in the low game		0.651*** (0.034)
Constant	0.222*** (0.034)	0.216*** (0.025)
Observations	2,470	2,470
R^2	0.452	0.424
Adjusted R^2	0.452	0.424
Residual Std. Error (df = 2468)	0.344	0.369
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table S1: Conditional cooperation in the control treatment. Similar to **Table 1**, here we report the results of a linear regression for the players' conditional cooperation probability in the control treatment.

5 Supplementary References

[1] Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, UK, 1998).

[2] Nowak, M. A. *Evolutionary dynamics* (Harvard University Press, Cambridge MA, 2006).

[3] Sigmund, K. The calculus of selfishness. In *The Calculus of Selfishness* (Princeton University Press, 2010).

[4] Broom, M. & Rychtář, J. *Game-Theoretical Models in Biology* (Chapman and Hall/CRC, 2013).

[5] Traulsen, A. & Hauert, C. Stochastic evolutionary game dynamics. In Schuster, H. G. (ed.) *Reviews of Nonlinear Dynamics and Complexity*, 25–61 (Wiley-VCH, Weinheim, 2009).

[6] Rand, D. G. & Nowak, M. A. Human cooperation. *Trends in cognitive sciences* **17**, 413–425 (2013).

[7] Perc, M., Gómez-Gardeñes, J., Szolnoki, A., Floría, L. M. & Moreno, Y. Evolutionary dynamics of group interactions on structured populations: A review. *Journal of The Royal Society Interface* **10**, 20120997 (2013).

[8] Perc, M. *et al.* Statistical physics of human cooperation. *Physics Reports* **687**, 1–51 (2017).

[9] Hilbe, C., Chatterjee, K. & Nowak, M. A. Partners and rivals in direct reciprocity. *Nature Human Behaviour* **2**, 469–477 (2018).

[10] García, J. & van Veelen, M. No strategy can win in the repeated prisoner's dilemma: Linking game theory and computer simulations. *Frontiers in Robotics and AI* **5**, 102 (2018).

[11] Cressman, R., Gaunersdorfer, A. & Wen, J. F. Evolutionary and dynamic stability in symmetric evolutionary games with two independent decisions. *International Game Theory Review* **2** (2000).

[12] Chamberland, M. & Cressman, R. An example of dynamic (in)consistency in symmetric extensive form evolutionary games. *Games and Economic Behavior* **30**, 319–326 (2000).

[13] Hashimoto, K. Unpredictability induced by unfocused games in evolutionary game dynamics. *Journal of Theoretical Biology* **241**, 669–675 (2006).

[14] Venkateswaran, V. R. & Gokhale, C. S. Evolutionary dynamics of complex multiple games. *Proceedings of the Royal Society B* **286**, 20190900 (2019).

[15] Sigmund, K., Hauert, C. & Nowak, M. A. Reward and punishment. *Proceedings of the National Academy of Sciences USA* **98**, 10757–10762 (2001).

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

[17] Panchanathan, K. & Boyd, R. Indirect reciprocity can stabilize cooperation without the second-order free-rider problem. *Nature* **432**, 499–502 (2004).

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

[19] dos Santos, M., Rankin, D. J. & Wedekind, C. The evolution of punishment through reputation. *Proceedings of the Royal Society B* **278**, 371–377 (2011).

[20] Hauser, O. P., Hendriks, A., Rand, D. G. & Nowak, M. A. Think global, act local: Preserving the global commons. *Scientific Reports* **6**, 36079 (2016).

[21] Pal, S. & Hilbe, C. Reputation effects drie the joint evolution of cooperation and social rewarding. *Nature Communications* **13**, 5928 (2022).

[22] Trivers, R. L. The evolution of reciprocal altruism. *The Quarterly Review of Biology* **46**, 35–57 (1971).

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

[24] Boyd, R. & Lorberbaum, J. P. No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature* **327**, 58–59 (1987).

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

[26] Frean, M. R. The prisoner's dilemma without synchrony. *Proceedings of the Royal Society B* **257**, 75–79 (1994).

[27] Hauert, C. & Schuster, H. G. Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. *Proceedings of the Royal Society B* **264**, 513–519 (1997).

[28] Szabó, G., Antal, T., Szabó, P. & Droz, M. Spatial evolutionary prisoner's dilemma game with three strategies and external constraints. *Physical Review E* **62**, 1095–1103 (2000).

[29] Killingback, T. & Doebeli, M. The continuous Prisoner's Dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *The American Naturalist* **160**, 421–438 (2002).

[30] Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).

[31] Kurokawa, S., Wakano, J. & Ihara, Y. Generous cooperators can outperform non-generous cooperators when replacing a population of defectors. *Theoretical Population Biology* **77** (2010).

[32] Martinez-Vaquero, L. A., Cuesta, J. A. & Sanchez, A. Generosity pays in the presence of direct reciprocity: A comprehensive study of 2x2 repeated games. *PLoS ONE* **7**, E35135 (2012).

[33] Grujic, J., Cuesta, J. A. & Sanchez, A. On the coexistence of cooperators, defectors and conditional cooperators in the multiplayer iterated prisoner's dilemma. *Journal of Theoretical Biology* **300**, 299–308 (2012).

[34] van Veelen, M., García, J., Rand, D. G. & Nowak, M. A. Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences USA* **109**, 9929–9934 (2012).

[35] van Segbroeck, S., Pacheco, J. M., Lenaerts, T. & Santos, F. C. Emergence of fairness in repeated group interactions. *Physical Review Letters* **108**, 158104 (2012).

[36] Stewart, A. J. & Plotkin, J. B. From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proceedings of the National Academy of Sciences USA* **110**, 15348–15353 (2013).

[37] Stewart, A. J. & Plotkin, J. B. Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences USA* **111**, 17558 – 17563 (2014).

[38] Knight, V., Harper, M., Glynatsi, N. & Campbell, O. Evolution reinforces cooperation with the

emergence of self-recognition mechanisms: An empirical study of strategies in the Moran process for the iterated prisoner's dilemma. *PLoS One* **13**, e0204981 (2018).

[39] Park, P. S., Nowak, M. A. & Hilbe, C. Cooperation in alternating interactions with memory constraints. *Nature Communications* **13**, 737 (2022).

[40] Chen, X. & Fu, F. Outlearning extortioners: Unbending strategies can foster reciprocal fairness and cooperation. *PNAS Nexus* **2**, pgad176 (2023).

[41] Bernheim, D. & Whinston, M. D. Multimarket contact and collusive behavior. *The RAND Journal of Economics* **21**, 1–26 (1990).

[42] Matsushima, H. Multimarket contact, imperfect monitoring, and implicit collusion. *Journal of Economic Theory* **98**, 158–178 (2001).

[43] Reiter, J. G., Hilbe, C., Rand, D. G., Chatterjee, K. & Nowak, M. A. Crosstalk in concurrent repeated games impedes direct reciprocity and requires stronger levels of forgiveness. *Nature Communications* **9**, 1–8 (2018).

[44] Donahue, K., Hauser, O. P., Nowak, M. A. & Hilbe, C. Evolving cooperation in multichannel games. *Nature Communications* **11**, 1–9 (2020).

[45] Dal Bó, P. & Fréchette, G. R. On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature* **56**, 60–114 (2018).

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

[47] Grujic, J., Fosco, C., Araujo, L., Cuesta, J. & Sanchez, A. Social experiments in the mesoscale: Humans playing a spatial prisoner's dilemma. *PLoS One* **5**, e13749 (2010).

[48] Gracia-Lázaro, C. *et al.* Heterogeneous networks do not promote cooperation when humans play a prisoner's dilemma. *Proceedings of the National Academy of Sciences USA* **109**, 12922–12926 (2012).

[49] Grujic, J. *et al.* A comparative analysis of spatial prisoner's dilemma experiments: Conditional cooperation and payoff irrelevance. *Scientific Reports* **4**, 4615 (2014).

[50] Fischbacher, U. & Gächter, S. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review* **100**, 541–556 (2010).

[51] Martinez-Martinez, I. & Normann, H. Cooperation in multiplayer dilemmas. *SSRN* http://dx.doi.org/10.2139/ssrn.4294660 (2022).

[52] Grujić, J., Eke, B., Cabrales, A., Cuesta, J. A. & Sánchez, A. Three is a crowd in iterated prisoner's dilemmas: experimental evidence on reciprocal behavior. *Scientific reports* **2**, 1–7 (2012).

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

[54] Rand, D. G., Arbesman, S. & Christakis, N. A. Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences USA* **108**, 19193–19198 (2011).

[55] Bednarik, P., Fehl, K. & Semmann, D. Costs for switching partners reduce network dynamics but not cooperative behaviour. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20141661 (2014).

[56] Falk, A., Fischbacher, U. & Gächter, S. Living in two neighborhoods? Social interaction effects in the laboratory. *Economic Inquiry* **51**, 563–578 (2013).

[57] McCarter, M. W., Samek, A. & Sheremeta, R. M. Divided loyalists or conditional cooperators? creating consensus about cooperation in multiple simultaneous social dilemmas. *Group & Organization Management* **39**, 744–771 (2014).

[58] Krieg, J. & Samek, A. When charities compete: A laboratory experiment with simultaneous public goods. *Journal of behavioral and experimental economics* **66**, 40–57 (2017).

[59] Yang, J., Kawamura, T. & Ogawa, K. Experimental multimarket contact inhibits cooperation. *Metroeconomica* **67**, 21–43 (2016).

[60] Angelovski, A., Di Cagno, D., Güth, W., Marazzi, F. & Panaccione, L. Behavioral spillovers in local public good provision: an experimental study. *Journal of Economic Psychology* **67**, 116–134 (2018).

[61] Laferriere, V., Montez, J., Roux, C. & Thöni, C. Multigame contact: A double-edged sword for cooperation. *AEJ: Microeconomics* (forthcoming) (2023).

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

[63] Baek, S. K., Jeong, H. C., Hilbe, C. & Nowak, M. A. Comparing reactive and memory-one strategies of direct reciprocity. *Scientific Reports* **6**, 25676 (2016).

[64] Selten, R. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* **4**, 25–55 (1975).

[65] Stevens, J. R., Volstorf, J., Schooler, L. J. & Rieskamp, J. Forgetting constrains the emergence of cooperative decision strategies. *Frontiers in Psychology* **1**, 235 (2011).

[66] Volstorf, J., Rieskamp, J. & Stevens, J. R. The good, the bad, and the rare: Memory for partners in social interactions. *PloS one* **6**, e18945 (2011).

[67] Dolan, P. & Galizzi, M. M. Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology* **47**, 1–16 (2015).

[68] Maki, A. *et al.* Meta-analysis of pro-environmental behaviour spillover. *Nature Sustainability* **2**, 307–315 (2019).

[69] Rabin, M. & Weizsäcker, G. Narrow bracketing and dominated choices. *American Economic Review* **99**, 1508–1543 (2009).

[70] Chen, D. L., Schonger, M. & Wickens, C. oTree – An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* **9**, 88–97 (2016).

[71] Ichinose, G. & Masuda, N. Zero-determinant strategies in finitely repeated games. *Journal of Theoretical Biology* **438**, 61–77 (2018).

Chapter 4 Supplementary information

Supplementary Information

Evolution of cooperation through cumulative reciprocity

Li et al.

Contents

Supplementary Section 1. Mathematical analysis of cumulative reciprocity

In this section we report a detailed analysis of the basic properties of the CURE strategy to support the main text sections "Payoffs against selected strategies" and "Fairness and stability of cumulative reciprocity".

a. CURE vs. CURE

Consider an infinitely repeated prisoner's dilemma between two players, 'Alice' and 'Bob'. For a given round t, let n_A and n_B denote how often each player has defected up to round t. Let $d = n_B - n_A$ denote the difference, to which we refer as the *defection difference statistic*. As in the main text, if Alice adopts CURE with some predefined threshold $\Delta_A \in \mathbb{N}$, she cooperates if and only if $d \leq \Delta_A$. Sometimes it will prove useful to highlight Alice's threshold, in which case we refer to her strategy as $CURE_{\Delta_A}$. Similarly, when Bob adopts CURE with threshold $\Delta_B \in \mathbb{N}$, he cooperates if and only if $-d \leq \Delta_B$. We get the following result when both players use the same threshold.

Proposition 1. *Consider an infinitely repeated prisoner's dilemma with possible payoffs R, S, T, P, respectively. Let* $\varepsilon > 0$ *denote the probability that a given player makes an error in any given round. Suppose the two players use CURE with an identical threshold* $\Delta \geq 1$. *Then, each player's average cooperation rate is*

$$\rho_{CURE} = 1 - \frac{2 - 3\varepsilon + 2\Delta(1-2\varepsilon)}{1 - 2\varepsilon^2 + 2\Delta(1-2\varepsilon)}\varepsilon. \tag{1.1}$$

The payoff of each player becomes

$$\begin{aligned}\pi(\text{CURE}, \text{CURE}) = &\frac{(1-\varepsilon)^2 \cdot \big(1 - 2(1-\varepsilon)\varepsilon + 2\Delta(1-2\varepsilon)\big)}{1 - 2\varepsilon^2 + 2\Delta(1-2\varepsilon)} \cdot R \\ &+ \frac{2\varepsilon(1-\varepsilon) \cdot (\Delta(1-2\varepsilon) + (1-\varepsilon)^2)}{1 - 2\varepsilon^2 + 2\Delta(1-2\varepsilon)} \cdot (T+S) \\ &+ \frac{\varepsilon^2 \cdot \big(3 - 2(3-\varepsilon)\varepsilon + 2\Delta(1-2\varepsilon)\big)}{1 - 2\varepsilon^2 + 2\Delta(1-2\varepsilon)} \cdot P.\end{aligned} \tag{1.2}$$

As an important special case of a prisoner's dilemma, we apply these results to the donation game[1]. Here, cooperation means to pay a cost $c > 0$ to provide a benefit $b > c$ to the co-player. As a result, $R = b - c$, $S = -c$, $T = b$, $P = 0$, and the payoffs simplify to

$$\pi(\text{CURE}, \text{CURE}) = \frac{1 - \varepsilon + 2\Delta(1-2\varepsilon)}{1 - 2\varepsilon^2 + 2\Delta(1-2\varepsilon)}(1-\varepsilon)(b-c). \tag{1.3}$$

In particular, for any positive error rate, the self-payoff of CURE is arbitrarily close to the theoretical maximum $(1-\varepsilon)(b-c)$, provided the strategy's tolerance Δ is sufficiently large.

Proof of Proposition 1. For a given round t, denote $x_d(t)$ as the probability that by round t, the difference in how often the players have cooperated is $d = n_B - n_A$. Assuming we know $x_d(t)$ for all $d \in \mathbb{Z}$, we can recursively compute $x_d(t+1)$,

$$\begin{array}{rcll}
\vdots & & \vdots & \\
x_{-\Delta-k}(t+1) & = & 2\varepsilon(1-\varepsilon)x_{-\Delta-k}(t) + \varepsilon^2 x_{-\Delta-k+1}(t) + (1-\varepsilon)^2 x_{-\Delta-k-1}(t) & \text{for } k \geq 2 \\
x_{-\Delta-1}(t+1) & = & 2\varepsilon(1-\varepsilon)x_{-\Delta-1}(t) + \varepsilon(1-\varepsilon)x_{-\Delta}(t) + (1-\varepsilon)^2 x_{-\Delta-2}(t) & \\
x_{-\Delta}(t+1) & = & ((1-\varepsilon)^2+\varepsilon^2)x_{-\Delta}(t) + \varepsilon(1-\varepsilon)x_{-\Delta+1}(t) + (1-\varepsilon)^2 x_{-\Delta-1}(t) & \\
\vdots & & \vdots & \\
x_k(t+1) & = & ((1-\varepsilon)^2+\varepsilon^2)x_k(t) + \varepsilon(1-\varepsilon)x_{k-1}(t) + \varepsilon(1-\varepsilon)x_{k+1}(t) & \text{for } -\Delta < k < \Delta \\
\vdots & & \vdots & \\
x_{\Delta}(t+1) & = & ((1-\varepsilon)^2+\varepsilon^2)x_{\Delta}(t) + \varepsilon(1-\varepsilon)x_{\Delta-1}(t) + (1-\varepsilon)^2 x_{\Delta+1}(t) & \\
x_{\Delta+1}(t+1) & = & 2\varepsilon(1-\varepsilon)x_{\Delta+1}(t) + \varepsilon(1-\varepsilon)x_{\Delta}(t) + (1-\varepsilon)^2 x_{\Delta+2}(t) & \\
x_{\Delta+k}(t+1) & = & 2\varepsilon(1-\varepsilon)x_{\Delta+k}(t) + \varepsilon^2 x_{\Delta+k-1}(t) + (1-\varepsilon)^2 x_{\Delta+k+1}(t) & \text{for } k \geq 2 \\
\vdots & & \vdots &
\end{array} \tag{1.4}$$

This recursion holds for all $t > 0$. By summing up the respective equations for $t \in \{1, \dots, \tau\}$, dividing both sides by τ, and taking the limit $\tau \to \infty$, we obtain

$$\begin{array}{rcll}
\vdots & & \vdots & \\
x_{-\Delta-k} & = & 2\varepsilon(1-\varepsilon)x_{-\Delta-k} + \varepsilon^2 x_{-\Delta-k+1} + (1-\varepsilon)^2 x_{-\Delta-k-1} & \text{for } k \geq 2 \\
x_{-\Delta-1} & = & 2\varepsilon(1-\varepsilon)x_{-\Delta-1} + \varepsilon(1-\varepsilon)x_{-\Delta} + (1-\varepsilon)^2 x_{-\Delta-2} & \\
x_{-\Delta} & = & ((1-\varepsilon)^2+\varepsilon^2)x_{-\Delta} + \varepsilon(1-\varepsilon)x_{-\Delta+1} + (1-\varepsilon)^2 x_{-\Delta-1} & \\
\vdots & & \vdots & \\
x_k & = & ((1-\varepsilon)^2+\varepsilon^2)x_k + \varepsilon(1-\varepsilon)x_{k-1} + \varepsilon(1-\varepsilon)x_{k+1} & \text{for } -\Delta < k < \Delta \\
\vdots & & \vdots & \\
x_{\Delta} & = & ((1-\varepsilon)^2+\varepsilon^2)x_{\Delta} + \varepsilon(1-\varepsilon)x_{\Delta-1} + (1-\varepsilon)^2 x_{\Delta+1} & \\
x_{\Delta+1} & = & 2\varepsilon(1-\varepsilon)x_{\Delta+1} + \varepsilon(1-\varepsilon)x_{\Delta} + (1-\varepsilon)^2 x_{\Delta+2} & \\
x_{\Delta+k} & = & 2\varepsilon(1-\varepsilon)x_{\Delta+k} + \varepsilon^2 x_{\Delta+k-1} + (1-\varepsilon)^2 x_{\Delta+k+1} & \text{for } k \geq 2 \\
\vdots & & \vdots &
\end{array} \tag{1.5}$$

where $x_d \coloneqq \lim_{\tau\to\infty} \frac{1}{\tau}\sum_{t=0}^{\tau} x_d(t)$ is the average time to observe difference d over all rounds t. Now, because both players apply the same strategy, it follows by symmetry that $x_{-d} = x_d$ for all $d \in \mathbb{N}$. Therefore, the above (double infinite) recursion simplifies to the following (single infinite) recursion

$$\begin{array}{rcll}
x_0 & = & ((1-\varepsilon)^2+\varepsilon^2)x_0 + 2\varepsilon(1-\varepsilon)x_1 & \\
\vdots & & \vdots & \\
x_k & = & ((1-\varepsilon)^2+\varepsilon^2)x_k + \varepsilon(1-\varepsilon)x_{k-1} + \varepsilon(1-\varepsilon)x_{k+1} & \text{for } 0 < k < \Delta \\
\vdots & & \vdots & \\
x_{\Delta} & = & ((1-\varepsilon)^2+\varepsilon^2)x_{\Delta} + \varepsilon(1-\varepsilon)x_{\Delta-1} + (1-\varepsilon)^2 x_{\Delta+1} & \\
x_{\Delta+1} & = & 2\varepsilon(1-\varepsilon)x_{\Delta+1} + \varepsilon(1-\varepsilon)x_{\Delta} + (1-\varepsilon)^2 x_{\Delta+2} & \\
x_{\Delta+k} & = & 2\varepsilon(1-\varepsilon)x_{\Delta+k} + \varepsilon^2 x_{\Delta+k-1} + (1-\varepsilon^2) x_{\Delta+k+1} & \text{for } k \geq 2
\end{array} \tag{1.6}$$

$\vdots \qquad\qquad\qquad\qquad \vdots$

By consecutively solving each equation for x_0 starting from the top, we obtain

$$x_d = \begin{cases} x_0 & if\ |d| \le \Delta \\ x_0 \left(\dfrac{\varepsilon}{1-\varepsilon}\right)^{2(|d|-\Delta)-1} & if\ |d| > \Delta \end{cases} \tag{1.7}$$

In addition, because all probabilities must end up to one,

$$1 = \sum_{d \in \mathbb{Z}} x_d = x_0 + 2\sum_{d=1}^{\infty} x_d = x_0 + 2\sum_{d=1}^{\Delta} x_0 + 2\sum_{d=\Delta+1}^{\infty} x_0 \left(\frac{\varepsilon}{1-\varepsilon}\right)^{2(d-\Delta)-1}. \tag{1.8}$$

By solving this equation for x_0 and plugging the solution into equation (1.7), we get

$$x_d = \begin{cases} \dfrac{1-2\varepsilon}{1+2\Delta-4\Delta\varepsilon-2\varepsilon^2} & if\ |d| \le \Delta \\ \dfrac{1-2\varepsilon}{1+2\Delta-4\Delta\varepsilon-2\varepsilon^2}\left(\dfrac{\varepsilon}{1-\varepsilon}\right)^{2(|d|-\Delta)-1} & if\ |d| > \Delta \end{cases} \tag{1.9}$$

Now, to calculate the players' average cooperation rates, we note the following: When $|d| < \Delta$, the probability that both players cooperate is $(1-\varepsilon)^2$, whereas the probability that exactly one player cooperates is $2\varepsilon(1-\varepsilon)$. Similarly, when $|d| > \Delta$, the probability that both players cooperate is $\varepsilon(1-\varepsilon)$, whereas the probability that exactly one player cooperates is $(1-\varepsilon)^2+\varepsilon^2$. As a result, we obtain

$$\rho_{CURE} = \sum_{d=-\Delta}^{\Delta} x_d \left[(1-\varepsilon)^2 + \frac{1}{2}2\varepsilon(1-\varepsilon)\right] + 2\sum_{d=\Delta+1}^{\infty} x_d \left[\varepsilon(1-\varepsilon) + \frac{1}{2}((1-\varepsilon)^2+\varepsilon^2)\right]. \tag{1.10}$$

Using the expressions for x_d in (1.9), we can explicitly compute ρ_{CURE}, yielding Eq. (1.1). To calculate the players' average payoffs, we proceed similarly, yielding the equation

$$\begin{aligned} \pi(\text{CURE}, \text{CURE}) = & \sum_{d=-\Delta}^{\Delta} x_d \left[(1-\varepsilon)^2 R + 2\varepsilon(1-\varepsilon)\frac{T+S}{2} + \varepsilon^2 P\right] \\ & + 2\sum_{d=\Delta+1}^{\infty} x_d \left[\varepsilon(1-\varepsilon)R + ((1-\varepsilon)^2+\varepsilon^2)\frac{T+S}{2} + \varepsilon(1-\varepsilon)P\right]. \end{aligned} \tag{1.11}$$

Again, by using the explicit expressions for x_d in Eq. (1.9), we can compute this payoff explicitly, yielding Eq. (1.2). $\square$

Using Proposition 1, we can compute exact payoffs and cooperation rates, as displayed in the following two tables.

Supplementary Table 1. Payoffs of CURE vs. CURE when both players apply the same threshold Δ. $R=3, S=0, T=5$ and $P=1$ unless otherwise specified.

Payoffs	No noise	1% noise	5% noise	10% noise
CURE (Δ= 1)	3	2.98656	2.93068	2.85651
CURE (Δ= 2)	3	2.98789	2.93727	2.86933
CURE (Δ= 3)	3	2.98846	2.94015	2.87505

CURE ($R=4, \Delta=1$)	4	3.96017	3.80411	3.61628

Supplementary Table 2. Cooperation rates of CURE vs. CURE when Δ**=1.** Payoff parameters: $R=3, S=0, T=5$ and $P=1$.

Cooperation rates	1% noise	5% noise	10% noise
CURE	0.986722	0.934705	0.872093

Using a similar methodology, we can also compute the players' payoffs if their thresholds Δ_A and Δ_B are different.

Proposition 2. *Under the assumptions of Proposition 1, suppose the two players now use different thresholds* Δ_A and Δ_B *when implementing CURE. Then both players have the same cooperation rate*

$$\rho = 1 - \frac{2 - 3\varepsilon + (\Delta_A + \Delta_B)(1-2\varepsilon)}{1 - 2\varepsilon^2 + (\Delta_A + \Delta_B)(1-2\varepsilon)}\varepsilon. \tag{1.12}$$

Moreover, both players obtain the same payoff, $\pi(\mathrm{CURE}_{\Delta_A}, \mathrm{CURE}_{\Delta_B}) = \pi(\mathrm{CURE}_{\Delta_B}, \mathrm{CURE}_{\Delta_A})$*, given by the expression*

$$\begin{aligned}\pi = {} & \frac{(1-\varepsilon)^2 \cdot \big(1 - 2(1-\varepsilon)\varepsilon + (\Delta_A + \Delta_B)(1-2\varepsilon)\big)}{1 - 2\varepsilon^2 + (\Delta_A + \Delta_B)(1-2\varepsilon)} \cdot R \\ & + \frac{\varepsilon(1-\varepsilon) \cdot \big((\Delta_A + \Delta_B)(1-2\varepsilon) + 2(1-\varepsilon)^2\big)}{1 - 2\varepsilon^2 + (\Delta_A + \Delta_B)(1-2\varepsilon)} \cdot (T+S) \\ & + \frac{\varepsilon^2 \cdot \big(3 - 2(3-\varepsilon)\varepsilon + (\Delta_A + \Delta_B)(1-2\varepsilon)\big)}{1 - 2\varepsilon^2 + (\Delta_A + \Delta_B)(1-2\varepsilon)} \cdot P.\end{aligned} \tag{1.13}$$

Proof of Proposition 2. The proof is similar to the proof of Proposition 1, with the only exception being that the resulting system of recursions is no longer symmetric with respect to $d=0$, but with respect to $d = (\Delta_A - \Delta_B)/2$. As a result, the average probability that the defection difference statistic is d is now given by

$$x_d = \begin{cases} \dfrac{1-2\varepsilon}{1 + (\Delta_A + \Delta_B)(1-2\varepsilon) - 2\varepsilon^2}\left(\dfrac{\varepsilon}{1-\varepsilon}\right)^{2(-d-\Delta_B)-1} & if\ d < -\Delta_B \\ \dfrac{1-2\varepsilon}{1 + (\Delta_A + \Delta_B)(1-2\varepsilon) - 2\varepsilon^2} & if\ -\Delta_B \le d \le \Delta_A \\ \dfrac{1-2\varepsilon}{1 + (\Delta_A + \Delta_B)(1-2\varepsilon) - 2\varepsilon^2}\left(\dfrac{\varepsilon}{1-\varepsilon}\right)^{2(d-\Delta_A)-1} & if\ d > \Delta_A \end{cases} \tag{1.14}$$

With these probabilities, we can compute the respective average cooperation rate of the two players with respective thresholds Δ_A and Δ_B as

$$\rho_A = \sum_{d=-\infty}^{\Delta_A} (1-\varepsilon)x_d + \sum_{d=\Delta_A+1}^{\infty} \varepsilon \cdot x_d, \tag{1.15}$$

$$\rho_B = \sum_{d=-\infty}^{-\Delta_B - 1} \varepsilon \cdot x_d + \sum_{d=-\Delta_B}^{\infty} (1-\varepsilon) x_d.$$

By plugging the values of x_d according to Eq. (1.14) into Eq. (1.15), one can verify that $\rho_A = \rho_B = \rho$. The respective result for the players' payoffs follows analogously, by considering

$$\begin{aligned}\pi_A = &\sum_{d=-\infty}^{-\Delta_B - 1} ((1-\varepsilon)^2 S + \varepsilon(1-\varepsilon)(R+P) + \varepsilon^2 T) x_d \\ &+ \sum_{d=-\Delta_B}^{\Delta_A} ((1-\varepsilon)^2 R + \varepsilon(1-\varepsilon)(T+S) + \varepsilon^2 P) x_d \\ &+ \sum_{d=\Delta_A+1}^{\infty} ((1-\varepsilon)^2 T + \varepsilon(1-\varepsilon)(R+P) + \varepsilon^2 S) x_d,\end{aligned} \tag{1.16}$$

and similarly,

$$\begin{aligned}\pi_B = &\sum_{d=-\infty}^{-\Delta_B - 1} ((1-\varepsilon)^2 T + \varepsilon(1-\varepsilon)(R+P) + \varepsilon^2 S) x_d \\ &+ \sum_{d=-\Delta_B}^{\Delta_A} ((1-\varepsilon)^2 R + \varepsilon(1-\varepsilon)(T+S) + \varepsilon^2 P) x_d \\ &+ \sum_{d=\Delta_A+1}^{\infty} ((1-\varepsilon)^2 S + \varepsilon(1-\varepsilon)(R+P) + \varepsilon^2 T) x_d.\end{aligned} \tag{1.17}$$

Again, by plugging in the values of x_d, it follows that $\pi_A = \pi_B = \pi$. □

We can use Proposition 2 to compare the payoffs of different versions of CURE_Δ against each other, as displayed in the following payoff tables:

Supplementary Table 3. Payoff matrix for different variants of CURE for 1% noise. Each value is the payoff of the row strategy in the row-column strategy game. Parameters: $R = 3, S = 0, T = 5$ and $P = 1$.

Payoffs	CURE_1	CURE_2	CURE_3	CURE_4
CURE_1	2.98656	2.98739	2.98789	2.98822
CURE_2	2.98739	2.98789	2.98822	2.98846
CURE_3	2.98789	2.98822	2.98846	2.98864
CURE_4	2.98822	2.98846	2.98864	2.98878

Supplementary Table 4. Payoff matrix for different variants of CURE for 10% noise. Each value is the payoff of the row strategy in the row-column strategy game. Parameters: $R = 3, S = 0, T = 5$ and $P = 1$.

Payoffs	$CURE_1$	$CURE_2$	$CURE_3$	$CURE_4$
$CURE_1$	2.85651	2.86444	2.86933	2.87265
$CURE_2$	2.86444	2.86933	2.87265	2.87505
$CURE_3$	2.86933	2.87265	2.87505	2.87687
$CURE_4$	2.87265	2.87505	2.87687	2.87829

According to these payoff matrices, all of the considered variants of CURE obtain approximately the same payoff against each other, with more tolerant variants obtaining slightly higher payoffs than less tolerant ones.

b. CURE vs unconditional strategies

Next, let us explore how CURE fares against the two unconditional strategies (ALLC and ALLD).

Proposition 3. *Consider an infinitely repeated prisoner's dilemma with error rate* $\varepsilon > 0$.

1. *If one player adopts the CURE strategy and the other player adopts ALLD, both players have the same average cooperation rate* $\rho_D = \epsilon$, *and they obtain the same payoff given by*
$$\pi(\text{CURE}, \text{ALLD}) = \pi(ALLD, \text{CURE}) = \varepsilon^2 R + \varepsilon(1-\varepsilon)(T+S) + (1-\varepsilon)^2 P \tag{1.18}$$
2. *If one player adopts the CURE strategy and the other player adopts ALLC, both players have the same average cooperation rate* $\rho_C = 1-\epsilon$, *and they obtain the same payoff given by*
$$\pi(\text{CURE}, \text{ALLC}) = \pi(ALLC, \text{CURE}) = (1-\varepsilon)^2 R + \varepsilon(1-\varepsilon)(T+S) + \varepsilon^2 P \tag{1.19}$$

Proof of Proposition 3. We only show the first case; the second case then follows from symmetry considerations. Suppose Alice adopts CURE, whereas Bob uses ALLD. Then, similar to the proof of Proposition 1, the average probability to observe a round in which the player's difference in cooperation is d satisfies an infinite linear system given by

$$\begin{array}{rcll}
 & \vdots & & \\
x_d & = & (1-\varepsilon)^2 x_{d-1} + 2\varepsilon(1-\varepsilon)x_d + \varepsilon^2 x_{d+1} & \text{for } d < \Delta \\
 & \vdots & & \\
x_\Delta & = & (1-\varepsilon)^2 x_{\Delta-1} + 2\varepsilon(1-\varepsilon)x_\Delta + \varepsilon(1-\varepsilon)x_{\Delta+1} & \\
x_{\Delta+1} & = & (1-\varepsilon)^2 x_\Delta + ((1-\varepsilon)^2+\varepsilon^2)x_{\Delta+1} + \varepsilon(1-\varepsilon)x_{\Delta+2} & \\
 & \vdots & & \\
x_d & = & \varepsilon(1-\varepsilon)x_{d-1} + ((1-\varepsilon)^2+\varepsilon^2)x_d + \varepsilon(1-\varepsilon)x_{d+1} & \text{for } d > \Delta+1 \\
 & \vdots & &
\end{array} \tag{1.20}$$

This homogeneous system has the following solution

$$x_d = \begin{cases} \alpha\left(\dfrac{\varepsilon}{1-\varepsilon}\right)^{2(d-\Delta)+1} & \text{for } d \le \Delta \\ \alpha & \text{for } d > \Delta, \end{cases} \tag{1.21}$$

where $\alpha \in \mathbb{R}$ is some arbitrary parameter. Denote $A = \sum_{d \le \Delta} x_d$ as the probability to observe the two players in a state in which the CURE player would cooperate. If A was positive, it would follow from Eq. (1.21) that $\alpha = A\,(1-2\varepsilon)/(\varepsilon(1-\varepsilon)) > 0$. Again, by Eq. (1.21), this would, in turn, contradict the normalization condition $\sum_{d\in\mathbb{Z}} x_d = 1$. Therefore, we conclude that $A = 0$, and thus, the CURE player behaves like an ALLD player almost surely. However, the cooperation rates and payoffs of ALLD versus ALLD are straightforward to calculate, and are given by Eq. (1.18). □

Using Proposition 3, we can calculate the payoffs of CURE against ALLC and ALLD for various error rates, as depicted in the following table.

Supplementary Table 5. Payoffs of CURE vs. unconditional strategies. $R = 3, S = 0, T = 5, P = 1$, and the tolerance level $\Delta = 1$.

Payoffs	1% noise	5% noise	10% noise
CURE	2.9899	2.9475	2.8900
ALLC	2.9899	2.9475	2.8900
CURE	1.0299	1.1475	1.2900
ALLD	1.0299	1.1475	1.2900

The above results imply that CURE is, in general, not a Nash equilibrium.

Corollary 1. *Consider an infinitely repeated prisoner's dilemma with* $2R > T + S$. *Then, for any error rate such that* $0 < \varepsilon < 1/2$, *CURE can be invaded by ALLC.*

Proof. By combining Propositions 1 and 3, we obtain

$$\pi(\text{ALLC}, \text{CURE}) - \pi(CURE, \text{CURE}) = \frac{\varepsilon(1-\varepsilon)(1-2\varepsilon)}{1-2\varepsilon^2+\Delta(2-4\varepsilon)}[(1-2\varepsilon)(2R-T-S)+2\varepsilon(R-P)] \quad (1.22)$$

Because of the conditions $2R > T + S$ and $R > P$ of the prisoner's dilemma, this difference is positive for any $0 < \varepsilon < 1/2$. □

c. An approximation for CURE vs an arbitrary memory-one strategy.

Next, suppose Alice adopts CURE and Bob adopts some memory-one strategy $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$. For a given error rate $\varepsilon > 0$, let $\tilde{\mathbf{p}} = (1-\varepsilon)\mathbf{p} + \varepsilon(\mathbf{1}-\mathbf{p})$ denote Bob's effective strategy in the presence of errors. For an interaction between the two players, let x_{d,a_1,a_2} denote the probability that in a randomly picked round, d is the cumulative difference in how often the players have cooperated and that players have chosen actions $a_1, a_2 \in \{C, D\}$, respectively, in the previous round. Here, a_1 refers to Alice's previous action, and a_2 refers to Bob's previous action. Similar to the previous analysis, we can derive a system of equations for $x_{d,a1,a2}$, given by the following four cases:

Case $d < \Delta$

$$\begin{aligned}
x_{d,CC} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d,a_1,a_2} (1-\varepsilon)\tilde{p}_{a_2 a_1} \\
x_{d,CD} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d-1,a_1,a_2} (1-\varepsilon)(1-\tilde{p}_{a_2 a_1}) \\
x_{d,DC} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d+1,a_1,a_2}\, \varepsilon\, \tilde{p}_{a_2 a_1} \\
x_{d,DD} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d,a_1,a_2} \varepsilon (1-\tilde{p}_{a_2 a_1})
\end{aligned} \tag{1.23}$$

Case $d = \Delta$

$$\begin{aligned}
x_{d,CC} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d,a_1,a_2} (1-\varepsilon)\tilde{p}_{a_2 a_1} \\
x_{d,CD} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d-1,a_1,a_2} (1-\varepsilon)(1-\tilde{p}_{a_2 a_1}) \\
x_{d,DC} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d+1,a_1,a_2} (1-\varepsilon)\tilde{p}_{a_2 a_1} \\
x_{d,DD} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d,a_1,a_2}\, \varepsilon\, (1-\tilde{p}_{a_2 a_1})
\end{aligned} \tag{1.24}$$

Case $d = \Delta + 1$

$$\begin{aligned}
x_{d,CC} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d,a_1,a_2}\, \varepsilon\, \tilde{p}_{a_2 a_1} \\
x_{d,CD} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d-1,a_1,a_2} (1-\varepsilon)(1-\tilde{p}_{a_2 a_1}) \\
x_{d,DC} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d+1,a_1,a_2} (1-\varepsilon)\tilde{p}_{a_2 a_1} \\
x_{d,DD} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d,a_1,a_2}\, (1-\varepsilon)(1-\tilde{p}_{a_2 a_1})
\end{aligned} \tag{1.25}$$

Case $d > \Delta + 1$

$$\begin{aligned}
x_{d,CC} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d,a_1,a_2}\, \varepsilon\, \tilde{p}_{a_2 a_1} \\
x_{d,CD} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d-1,a_1,a_2}\, \varepsilon\, (1-\tilde{p}_{a_2 a_1}) \\
x_{d,DC} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d+1,a_1,a_2} (1-\varepsilon)\tilde{p}_{a_2 a_1} \\
x_{d,DD} &= \sum\nolimits_{a_1,a_2 \in \{C,D\}} x_{d,a_1,a_2}\, (1-\varepsilon)(1-\tilde{p}_{a_2 a_1})
\end{aligned} \tag{1.26}$$

As a very crude approximation for this infinitely dimensional linear system we solve a truncated system. To this end, we fix some $K \gg \Delta$ and set $x_{-d,a_1,a_2} = x_{d,a_1,a_2} := 0$ for all actions $a_1, a_2 \in \{C, D\}$ and $d > K$. This yields $4(2K+1)$ equations. From these equations, we drop the equation for $x_{K,DC}$, and replace it by the following normalization condition,

$$\sum_{-K \le d \le K} \sum_{a_1,a_2} x_{d,a_1,a_2} = 1. \tag{1.27}$$

If $\tilde{x}_{d,a_1,a_2}$ is the (typically unique) solution to this truncated linear system, we can calculate the players' approximate cooperation rates as

$$\rho_A = \sum_{-K \le d \le K} \tilde{x}_{d,CC} + \tilde{x}_{d,CD} \quad \text{and} \quad \rho_B = \sum_{-K \le d \le K} \tilde{x}_{d,CC} + \tilde{x}_{d,DC}. \tag{1.28}$$

Similarly, we can compute approximate payoffs as

$$\begin{aligned} \pi_A &= \sum_{-K \le d \le K} \tilde{x}_{d,CC}R + \tilde{x}_{d,CD}S + \tilde{x}_{d,DC}T + \tilde{x}_{d,DD}P \\ \pi_B &= \sum_{-K \le d \le K} \tilde{x}_{d,CC}R + \tilde{x}_{d,CD}T + \tilde{x}_{d,DC}S + \tilde{x}_{d,DD}P \end{aligned} \tag{1.29}$$

The following table shows the approximated payoffs for $\Delta = 1$ and $K = 5 \cdot 10^3$. The results are in remarkably good agreement with the corresponding simulations shown in Supplementary Table 8.

Supplementary Table 6. Payoffs of CURE vs an arbitrary memory-one strategy. Parameters: $R = 3, S = 0, T = 5$ and $P = 1$, and the tolerance level $\Delta = 1$.

Payoffs	1% noise	5% noise	10% noise
CURE	2.2500	2.2500	2.2500
TFT	2.2500	2.2500	2.2500
CURE	2.9081	2.6676	2.5153
GTFT(0.1)	2.9081	2.6676	2.5153
CURE	2.9671	2.8466	2.7211
GTFT(0.3)	2.9671	2.8466	2.7211
CURE	2.9800	2.9008	2.8056
GTFT(0.5)	2.9800	2.9008	2.8056
CURE	2.2500	2.2500	2.2500
WSLS	2.2500	2.2500	2.2500

d. Fairness and stability of CURE

An important property of CURE is that it always enforces a fair outcome between the two players in the following sense.

Proposition 4. *Consider a repeated 2-player 2-action game with symmetric payoffs such that $T \ge S$, and suppose players are subject to errors at rate $\varepsilon \ge 0$. Suppose Alice adopts the CURE strategy, whereas Bob adopts an arbitrary strategy such that the resulting realized payoffs π_A and π_B exist. Then with probability 1, the payoffs satisfy $\pi_A = \pi_B$.*

Proof. For a given instantiation of the game between the two players, denote $e_A(t)$ and $e_B(t)$ as the indicator function that is equal to 1 if the respective player cooperates in round t, and equal to 0 otherwise. Let $n_A(\tau) = \sum_{t=1}^{\tau}(1 - e_i(t))$ denote how often Alice has defected up to round τ, and define $n_B(\tau)$ analogously. Then $d(\tau) = n_B(\tau) - n_A(\tau)$. Using this notation, we can write the players' realized payoffs as

$$\begin{aligned}\pi_A &= \lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau} e_A(t)e_B(t)R + e_A(t)(1-e_B(t))S + (1-e_A(t))e_B(t)T + (1-e_A(t))(1-e_B(t))P \\ \pi_B &= \lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau} e_A(t)e_B(t)R + e_A(t)(1-e_B(t))T + (1-e_A(t))e_B(t)S + (1-e_A(t))(1-e_B(t))P\end{aligned} \tag{1.30}$$

It follows that

$$\pi_B - \pi_A = \lim_{\tau\to\infty}\frac{T-S}{\tau}\sum_{t=1}^{\tau} e_A(t) - e_B(t) = (T-S)\lim_{\tau\to\infty}\left(\frac{n_B(\tau)}{\tau} - \frac{n_A(\tau)}{\tau}\right) = (T-S)\lim_{\tau\to\infty}\frac{d(\tau)}{\tau}. \tag{1.31}$$

Now assume that $\pi_B > \pi_A$. Then, it follows from Eq. (1.31) that there is some t_0 such that $d(t) > \Delta$ for all $t > t_0$. It follows from the definition of CURE, that Alice acts for all but finitely many rounds like a defector, and therefore, $\lim_{\tau\to\infty} n_A(\tau)/\tau = 1-\varepsilon$. However, because $\lim_{\tau\to\infty} n_B(\tau)/\tau \leq 1-\varepsilon$, Eq. (1.31) yields $\pi_B - \pi_A \leq 0$, a contradiction. The assumption $\pi_B < \pi_A$ yields a similar contradiction. In that case, the CURE player acts like an unconditional cooperator in all but finitely many rounds. □

Proposition 4 allows us to derive an interesting result about the stability of CURE. As we have shown, CURE is not a Nash equilibrium in the presence of errors. Instead, players can gain a payoff advantage by deviating to ALLC. However, the following two results show that nevertheless, CURE has reasonable robustness properties when the error rate vanishes, or when errors are sufficiently rare.

Proposition 5. *Consider a repeated 2-player game with 2 actions, such that payoffs satisfy the inequalities* $2R > T+S > 2P$ *and* $T \geqslant S$*. Then, for any threshold* Δ*, the strategy profile* $(CURE_\Delta, CURE_\Delta)$ *is a subgame perfect equilibrium in the limit of rare errors,* $\varepsilon \to 0$.

Proof. To show the result, we need to show that no deviation from $CURE_\Delta$ is profitable, after no previous history of the game. To this end, consider an arbitrary history such that the current round's defection difference statistic is d.

Then the continuation payoff of a focal player who continues to use $CURE_\Delta$ is *R*. This follows from the observation that independent of the current value of *d*, two $CURE_\Delta$ players will reach $|d| \leq \Delta$ in finite time with certainty. From then on, the players' continuation payoff will be given by *R*; the finitely many rounds before that do not affect the players' long-run payoffs.

Now suppose there is a deviation such that the focal player obtains a continuation payoff larger than *R*. Then it follows from Proposition 4 that also the co-player gets a payoff larger than *R* (since the co-player continues to adopt $CURE_\Delta$). This implies that the joint payoff of both players is larger than 2*R*, which contradicts the assumption that 2*R* is the maximum payoff that is achievable in any single round. □

In addition to the previous result for vanishing errors, we can also show a robustness property for CURE when errors are sufficiently rare. To this end, we say a strategy $\hat{\sigma}$ is an approximate Nash equilibrium with respect to some given threshold $\hat{\varepsilon}$ if any other strategy σ can gain at most a payoff advantage of $\hat{\varepsilon}$ against $\hat{\sigma}$, i.e. if $\pi(\sigma,\hat{\sigma})-\pi(\hat{\sigma},\hat{\sigma})\leq\hat{\varepsilon}$ for all σ (such strategies are sometimes referred to as $\hat{\varepsilon}$-equilibria). The following shows that CURE is an $\hat{\varepsilon}$-equilibrium if the error rate ε is sufficiently small.

Proposition 6. *Consider a repeated prisoner's dilemma with payoffs R,S,T,P, and fix an arbitrary threshold $\hat{\varepsilon}$. Then CURE is an $\hat{\varepsilon}$-equilibrium for all $\varepsilon\leq\frac{\hat{\varepsilon}}{max(2R-T-S,R-P)}$.*

Proof. The maximum payoff that two players can jointly achieve in a repeated game with errors is twice the payoff that they get if they both intend to cooperate in every round,

$$\pi(\mathrm{ALLC},\mathrm{ALLC})=(1-\varepsilon)^2R+\varepsilon(1-\varepsilon)(T+S)+\varepsilon^2P. \tag{1.32}$$

Because for an arbitrary opponent strategy σ, CURE always obtains the same payoff, it follows that

$$\pi(\sigma,\mathrm{CURE})\leq(1-\varepsilon)^2R+\varepsilon(1-\varepsilon)(T+S)+\varepsilon^2P. \tag{1.33}$$

Therefore, by Eq. (1.2), we obtain the estimate

$$\begin{aligned}\pi(\sigma,\mathrm{CURE})-\pi(CURE,\mathrm{CURE})&\leq\frac{\varepsilon(1-\varepsilon)(1-2\varepsilon)}{1-2\varepsilon^2+\Delta(2-4\varepsilon)}[(1-2\varepsilon)(2R-T-S)+2\varepsilon(R-P)]\\&\leq\varepsilon\max(2R-T-S,R-P)\leq\hat{\varepsilon}.\end{aligned} \tag{1.34}$$

□

Supplementary Section 2. Payoffs and cooperation rates obtained through computer simulations

To validate the above analytical results, we have implemented independent computer simulations to estimate the players' payoffs and cooperation rates. To this end, we consider three different error rates (1%, 5%, 10%), and we take the payoffs of Axelrod's tournament[2] as a baseline ($R = 3, S = 0, T = 5$ and $P = 1$). Using these parameters, we explore two scenarios: games between two players who adopt the same strategies, and games between a cumulative reciprocator and another player who adopts one of nine well-known strategies for comparison. In addition to ALLC and ALLD, these strategies include tit-for-tat, contrite tit-for-tat[3], generous tit-for-tat (with different generosity levels), win-stay lose-shift, and an extortionate strategy[4]. Moreover, as two examples of strategies that cannot be implemented with finite memory, we consider 'SoftMajority' and 'HardMajority'[5]. A SoftMajority player cooperates in the first round, and cooperates if the number of times the opponent has cooperated is no less than the number of times the opponent has defected. Otherwise, the SoftMajority player defects. A HardMajority player defects on the first move, and keeps defecting unless the number of times the opponent has cooperated is greater than the number of times the opponent has defected.

a. Games between two players who adopt the same strategy

Supplementary Table 7. Average payoffs when both players adopt the same strategy. From left to right, payoffs for the repeated prisoner's dilemma are obtained for no noise, 1%, 5%, and 10% noise levels. From top to bottom, there are CURE, TFT-type, cooperative and defective strategies.

Gaming Strategy	**Noise rate**			
	No noise (**0%**)	Low (**1%**)	Medium (**5%**)	High (**10%**)
CURE(Δ=0)	3	2.979999	2.900239	2.801839
CURE(Δ=1)	3	2.986557	2.930680	2.856508
CURE(Δ=2)	3	2.987889	2.937267	2.869329
CURE(Δ=3)	3	2.988461	2.940148	2.875050
CURE (*R*=4, Δ=1)	4	3.960168	3.804112	3.616276
TFT	3	2.251015	2.248285	2.250255
CTFT	3	2.669795	2.619060	2.570005
GTFT(G=0.1)	3	2.907585	2.667374	2.515275
GTFT(G=0.3)	3	2.967007	2.846495	2.721121
GTFT(G=0.5)	3	2.980035	2.900715	2.805665
ALLC	3	2.989885	2.947495	2.889995
WSLS	3	2.951255	2.776353	2.601810
WSLS(*R*=4)	4	3.911758	3.593417	3.267800
SoftMajority	3	2.989906	2.929455	2.794111

ALLD	1	1.029905	1.147615	1.290415
HardMajority	1	1.029963	1.147405	1.369848
Extort2	1.000115	1.136155	1.493630	1.740825

Supplementary Table 7 shows the average payoff of a player when playing the repeated prisoner's dilemma (RPD) with another player who uses the same strategy. The results indicate that CURE copes well with noise. Cooperation is well-maintained as the noise rate ranges from 1% to 10%. The average payoffs of both players increase as the tolerance level Δ increases. In comparison with the performances of other classic strategies, the payoff of a pair of CURE players is slightly lower than that of a pair of ALLC players, while it is comparable to the payoff of two SoftMajority players. GTFT performs well at the low noise rate; but its payoff decreases to a relatively lower level when the noise increases. The performance of WSLS is very similar to GTFT. CURE has a relatively higher payoff compared to GTFT and WSLS when the noise rate is medium and high (i.e., at 5% and 10%). The other tested strategies are less successful in retaining cooperation against themselves in noisy environments.

b. Games between a CURE player and a player who adopts one of the selected strategies

We also test the payoffs and cooperation rates in the RPD between a CURE player and another player who uses one of several well-known strategies (Supplementary Table 8).

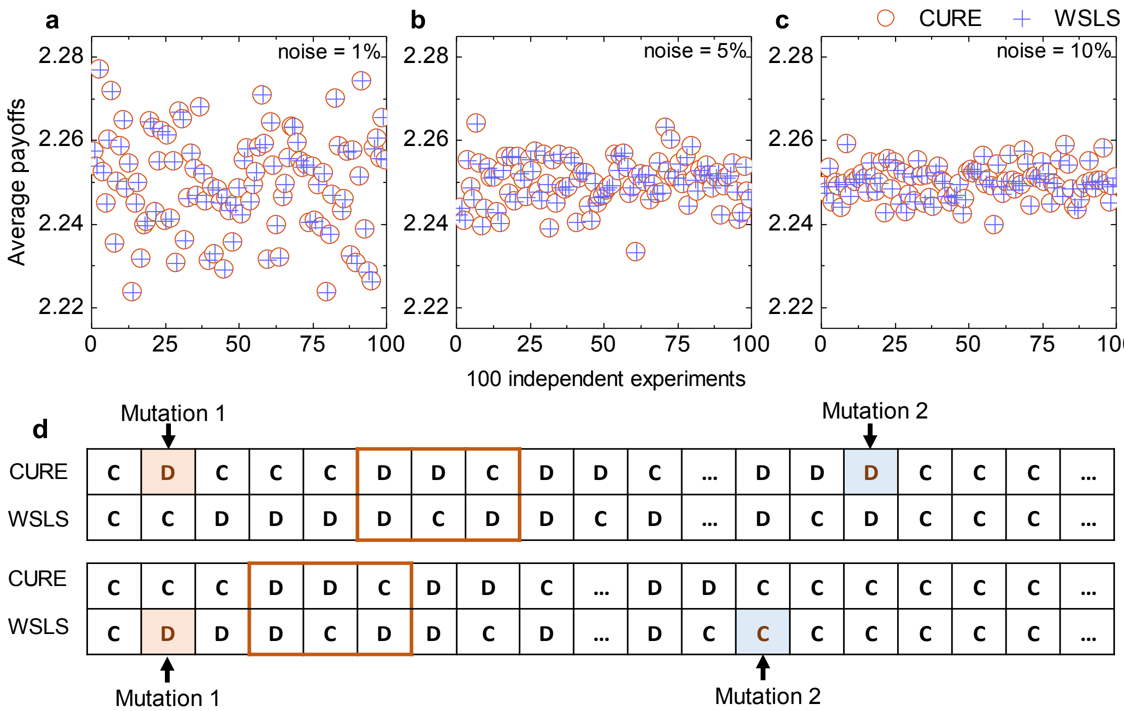
Supplementary Table 8. Average payoffs and cooperation rates of CURE (Δ= 1) against selected strategies. From left to right, payoffs are obtained for no noise, 1%, 5%, and 10% noise levels. From top to bottom, there are the payoffs when CURE interacts with several selected strategies, and the difference in payoffs between the two sides.

Gaming Strategy	**Noise rate**							
	No noise (**0%**)		Low (**1%**)		Medium (**5%**)		High (**10%**)	
	Payoff	**Cooperation rates**	**Payoff**	**Cooperation rates**	**Payoff**	**Cooperation rates**	**Payoff**	**Cooperation rates**
CURE	3	1	2.986557	0.986775	2.930680	0.934694	2.856508	0.872152
CURE	3	1	2.986557	0.986775	2.930680	0.934694	2.856508	0.872152
d(CR-CR)	**0**	**0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
CURE	3	1	2.251152	0.500904	2.246734	0.498665	2.250205	0.500318
TFT	3	1	2.251225	0.500889	2.246815	0.498649	2.250283	0.500302
d(CR-T)	**0**	**0**	**-0.000073**	**0.000015**	**-0.000081**	**0.000016**	**-0.000078**	**0.000016**
CURE	3	1	2.749441	0.754327	2.691164	0.717555	2.634868	0.688831
CTFT	3	1	2.749501	0.754315	2.691230	0.717542	2.634920	0.688829
d(CR-CT)	**0**	**0**	**-0.000060**	**0.000012**	**-0.000066**	**0.000013**	**-0.000052**	**0.000002**
CURE	3	1	2.907732	0.915000	2.667168	0.736632	2.516348	0.643413
GTFT(0.1)	3	1	2.907778	0.914990	2.667224	0.736620	2.516399	0.643402

d(CR-GT)	**0**	**0**	**-0.000056**	**0.000010**	**-0.000056**	**0.000012**	**-0.000051**	**0.000011**
CURE	3	1	2.966883	0.967980	2.846386	0.864728	2.721007	0.772590
GTFT(0.3)	3	1	2.966915	0.967974	2.846416	0.864722	2.721049	0.772582
d(CR-GT)	**0**	**0**	**-0.000032**	**0.000006**	**-0.000030**	**0.000006**	**-0.000042**	**0.000008**
CURE	3	1	2.980097	0.980484	2.900952	0.909250	2.805319	0.833204
GTFT(0.5)	3	1	2.980099	0.980484	2.900941	0.909252	2.805311	0.833205
d(CR-GT)	**0**	**0**	**-0.000002**	**0.000000**	**0.000011**	**-0.000002**	**0.000008**	**-0.000001**
CURE	3	1	2.990643	0.989620	2.948772	0.949198	2.892204	0.898982
AllC	3	1	2.988761	0.989996	2.945226	0.949907	2.886759	0.900071
d(CR-C)	**0**	**0**	**0.001882**	**-0.000376**	**0.003548**	**-0.000709**	**0.005445**	**-0.001089**
CURE	3	1	2.990475	0.989663	2.948973	0.949215	2.891768	0.898927
SoftMajority	3	1	2.988948	0.989969	2.945169	0.949975	2.886774	0.899926
d(CR-SM)	**0**	**0**	**0.001528**	**-0.000306**	**0.003804**	**-0.000760**	**0.004994**	**-0.000999**
CURE	3	1	2.249640	0.499799	2.249858	0.499929	2.249803	0.499883
WSLS	3	1	2.249717	0.499784	2.249932	0.499914	2.249878	0.499868
d(CR-W)	**0**	**0**	**-0.000077**	**0.000015**	**-0.000074**	**0.000015**	**-0.000075**	**0.000015**
R=4 CURE	4	1	2.498530	0.499530	2.499094	0.499718	2.499895	0.499986
WSLS	4	1	2.498606	0.499515	2.499169	0.499703	2.499973	0.499970
d(CR-W)	**0**	**0**	**-0.000076**	**0.000015**	**-0.000075**	**0.000015**	**-0.000078**	**0.000016**
CURE	0.999980	0.000002	1.029555	0.010383	1.146963	0.050810	1.288744	0.100965
ALLD	1.000080	0.0	1.031416	0.010011	1.150627	0.050077	1.293803	0.099954
d(CR- D)	**-0.000100**	**0.000002**	**-0.001861**	**0.000372**	**-0.003664**	**0.000733**	**-0.005059**	**0.001011**
CURE	1.000202	0.000130	1.115546	0.050590	1.439270	0.187360	1.684724	0.285129
Extort2	1.000302	0.000110	1.115744	0.050550	1.439440	0.187320	1.684891	0.285096
d(CR-E2)	**-0.000100**	**0.000020**	**-0.000198**	**0.000004**	**-0.000170**	**0.000040**	**-0.000167**	**0.000033**
CURE	2.999970	1	2.970880	0.979892	2.894839	0.922271	2.795762	0.851149
HardMajority	3.000020	0.999990	2.969422	0.980183	2.891396	0.922959	2.791311	0.852040
d(CR-HM)	**-0.000050**	**0.000001**	**0.001458**	**-0.000191**	**0.003443**	**-0.000688**	**0.004451**	**-0.000891**

In general, CURE ensures an equal outcome against any opponent's strategy. When the CURE player encounters a player who uses an exploitative strategy, especially the unconditional exploitation strategy (ALLD), CURE still restricts the payoff difference. The simulated data shown in Supplementary Tables 7 and 8 are in good agreement with the analytical and approximated results provided in Section 1 of the Supplementary Information.

The Supplementary Tables 1-8 reveal several remarkable regularities. First, with respect to a strategy's payoff against itself, CURE outcompetes almost every strategy other than ALLC (Supplementary Table 7). Second, when we look at interactions between CURE and any other strategy, we indeed find that both players obtain the same payoff (Supplementary Table 8). This result is independent of the error rate and CURE's tolerance levels. Somewhat surprisingly, we finally note that even if CURE's co-player adopts a partner strategy, the resulting dynamics does not necessarily lead to full cooperation. Supplementary Fig. 1 illustrates an example. Here, CURE is matched with a win-stay lose-shift opponent. For positive error rates, the two players yield an approximate payoff of $(R + S + T + P)/4$ (which is below the mutual cooperation payoff R that either strategy achieves against itself). The same payoff is achieved when CURE interacts with TFT (Supplementary Tables 6,8). These results suggest that although CURE is generally cooperative, it does not cooperate with any other cooperative strategy in the presence of errors.



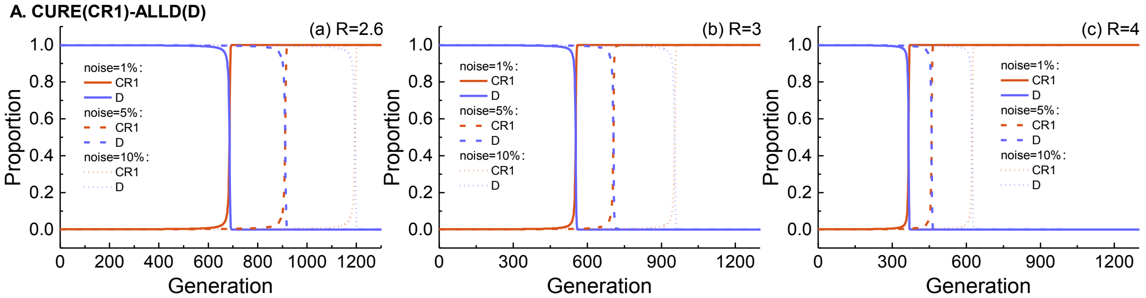
Supplementary Fig. 1. The competition between CURE and WSLS in the noisy repeated prisoner's dilemma. a-c, We have run 100 independent simulations to explore the dynamics among a cumulative reciprocator and a player adopting win-stay lose-shift (WSLS). The panels show the players' payoffs for each simulation. While there are some fluctuations between simulations, the difference between the two players' payoffs is negligible in each single simulation. On average, payoffs are close to the theoretically predicted $(R + S + T + P)/4 = 2.25$ for the infinitely repeated game. In particular, payoffs are below the mutual cooperation payoff of R. **d,** To explain why the two players fail to consistently coordinate on mutual cooperation, suppose both players initially start with mutual cooperation. Then a single error by either player can lead to a continuing cycle of alternating defection (i.e., $\mathrm{DD} \to \mathrm{DC} \to \mathrm{CD} \to \mathrm{DD} \to \cdots$). This cycle only stops if there occurs another error in a round in which CURE is expected to cooperate and WSLS is expected to defect. Simulations are based on the payoffs used in Axelrod's tournament, $R = 3, S = 0, T = 5, P = 1$. For CURE, we use a tolerance level of $\Delta = 1$.

Supplementary Section 3. Evolutionary dynamics in two-strategy populations

So far, we have assumed the players' strategies to be fixed. In the following, we explore what happens if players are part of an evolving population in which successful strategies are expected to spread. We begin with the simplest possible scenario in which there are only two possible strategies in the population. Specifically, we compare CURE with a tolerance level of $\Delta = 1$ to each of the nine other strategies that are used in previous strategy comparison (Supplementary Section 2). Fig.2 illustrates our results. In these simulations, CURE is initially adopted by 0.1% of the population (accordingly, we speak of the other strategy as the 'resident'). In this section, the illustrated dynamical processes are generated through computational simulation under three different values of R (i.e., $R = 2.6, 3.0, \text{and } 4.0$), so as to complement the results of Fig.2 (main text section "CURE and population dynamics"). The results for different payoff values and other initial frequencies of CURE are illustrated in Supplementary Figs.2-7.

a. CURE invades exploitive strategies (ALLD, HardMajority, and Extort2)

A. CURE(CR1)-ALLD(D)

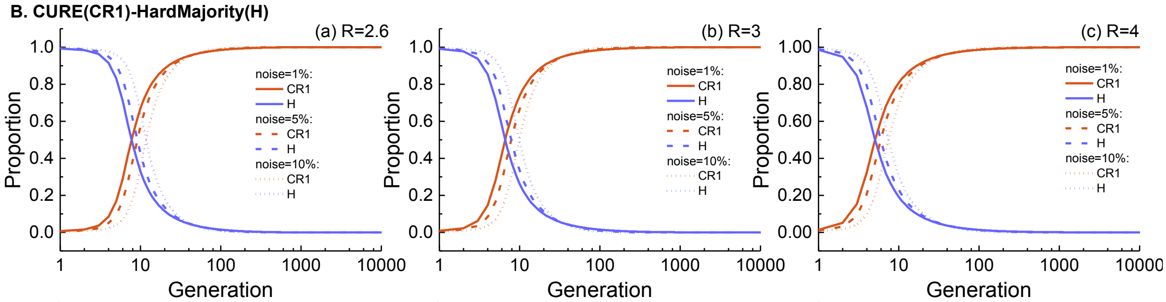


R=2.6	1% CR1	1% D	5% CR1	5% D	10% CR1	10% D
CR1	2.5971	1.02976	2.5813	1.1462	2.5526	1.2857
D	1.0303	1.02985	1.1475	1.1465	1.2873	1.286

R=3.0	1% CR1	1% D	5% CR1	5% D	10% CR1	10% D
CR1	2.9866	1.02978	2.9307	1.14723	2.8565	1.28963
D	1.03035	1.0299	1.14846	1.1475	1.29136	1.29

R=4.0	1% CR1	1% D	5% CR1	5% D	10% CR1	10% D
CR1	3.96	1.02988	3.8042	1.1498	3.6163	1.2996
D	1.03045	1.03	1.151	1.145	1.30134	1.3

B. CURE(CR1)-HardMajority(H)

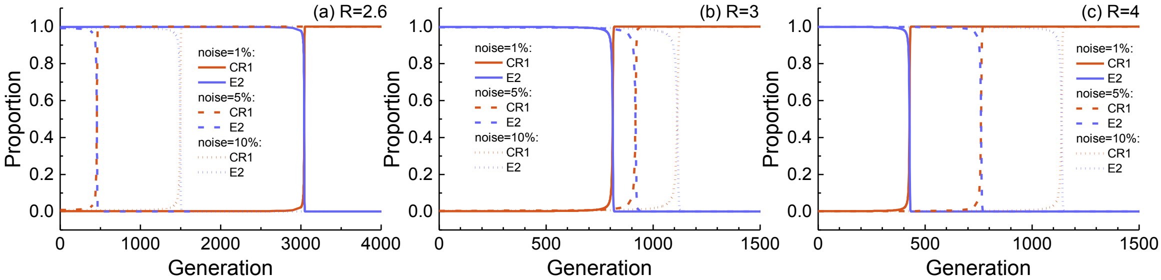


R=2.6	1% CR1	1% H	5% CR1	5% H	10% CR1	10% H
CR1	2.5971	2.5871	2.5813	2.5554	2.5526	2.4873
H	2.5866	1.0298	2.5542	1.1566	2.4859	1.3333

R=3.0	1% CR1	1% H	5% CR1	5% H	10% CR1	10% H
CR1	2.9866	2.9823	2.9307	2.8958	2.8565	2.7658
H	2.9817	1.0299	2.8946	1.1655	2.7643	1.3316

R=4.0	1% CR1	1% H	5% CR1	5% H	10% CR1	10% H
CR1	3.96	3.9583	3.8042	3.7854	3.6163	3.5490
H	3.9578	1.03	3.7843	1.1797	3.5476	1.3695

C. CURE(CR1)-Extort2(E2)



R= 2.6	1%		5%		10%	
	CR1	E2	CR1	E2	CR1	E2
CR1	2.5971	1.15493	2.5813	1.51114	2.5526	1.73699
E2	1.15494	1.1563	1.51115	1.5189	1.737	1.7418

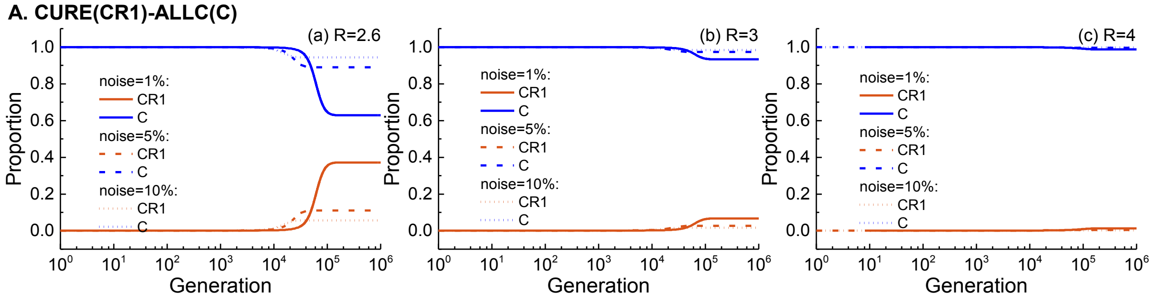
R= 3.0	1%		5%		10%	
	CR1	E2	CR1	E2	CR1	E2
CR1	2.9866	1.15847	2.9307	1.54319	2.8565	1.79130
E2	1.15849	1.1592	1.54320	1.5485	1.79131	1.7952

R= 4.0	1%		5%		10%	
	CR1	E2	CR1	E2	CR1	E2
CR1	3.96	1.16374	3.8042	1.59993	3.6163	1.89991
E2	1.16376	1.1638	1.59994	1.600	1.89992	1.89997

Supplementary Fig. 2. CURE invades the ALLD (the upper layer A), HardMajority (the middle layer B), and the Extort2 population (the lower layer C). CURE, with an initial proportion of 0.001, can invade ALLD and HardMajority populations, regardless of R and the noise rate. The lower the level of noise and the larger the mutual cooperation payoff R, the faster the invasion. It takes less than 100 generations for CURE to fully invade HardMajority. With respect to Extort2, CURE must reach a critical initial proportion to take over the population, depending on R and the noise rate. When $R = 2.6$, the critical value is 0.008 at 5% noise, and the critical value is 0.006 at 10% noise; when $R = 3.0$, the critical value is 0.004 at both 5% and 10% noises. In all other conditions, an initial 0.001 proportion of CURE players can invade the Extort2 population. Invasion is fastest for large values of R and substantial error rates. Each payoff matrix associating with each frequency curve shows the approximated payoffs. Double-precision data is used in the simulations.

b. Population dynamics of CURE and highly cooperative strategies (ALLC and SoftMajority)

A. CURE(CR1)-ALLC(C)

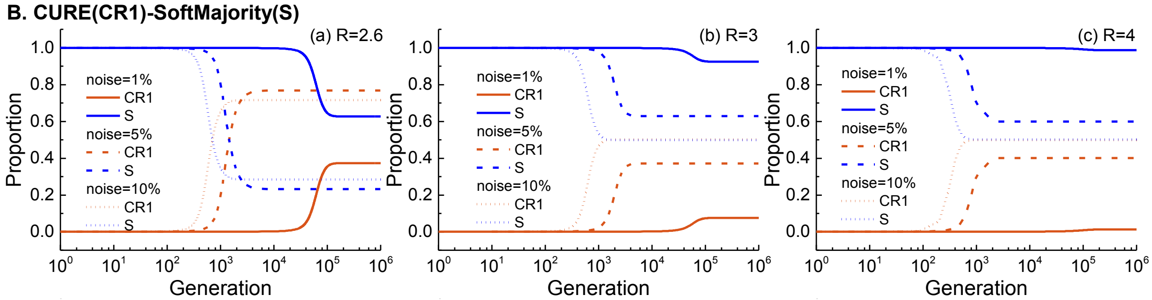


R= 2.6	1%		5%		10%	
	CR1	C	CR1	C	CR1	C
CR1	2.5971	2.5981	2.5813	2.5870	2.5526	2.5668
C	2.5976	2.5979	2.5859	2.5865	2.5651	2.5660

R= 3.0	1%		5%		10%	
	CR1	C	CR1	C	CR1	C
CR1	2.9866	2.9901	2.9307	2.9480	2.8565	2.8906
C	2.9896	2.9899	2.9468	2.9475	2.8890	2.8900

R= 4.0	1%		5%		10%	
	CR1	C	CR1	C	CR1	C
CR1	3.96	3.9701	3.8042	3.8502	3.6163	3.7003
C	3.9696	3.9700	3.8490	3.8500	3.6986	3.7000

B. CURE(CR1)-SoftMajority(S)



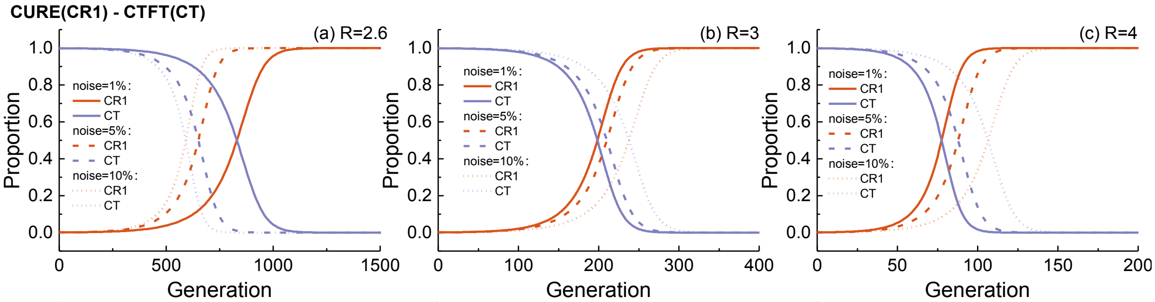
R= 2.6	1%		5%		10%	
	CR1	S	CR1	S	CR1	S
CR1	2.5971	2.5981	2.5813	2.5871	2.5526	2.5668
S	2.5976	2.5979	2.5858	2.5721	2.5651	2.5353

R= 3.0	1%		5%		10%	
	CR1	S	CR1	S	CR1	S
CR1	2.9866	2.9901	2.9307	2.9480	2.8565	2.8890
S	2.9895	2.9899	2.9468	2.9385	2.8873	2.8580

R= 4.0	1%		5%		10%	
	CR1	S	CR1	S	CR1	S
CR1	3.96	3.9701	3.8042	3.8503	3.6163	3.6956
S	3.9696	3.9700	3.8490	3.8203	3.6938	3.6184

Supplementary Fig. 3. Coexistence of CURE with ALLC (the upper layer A) and SoftMajority (the lower layer B). Initially, the frequency of CURE is 0.001, and that of the other strategy is 0.999. The frequency of CURE at the steady state decreases when R increases from 2.6 to 4.0, as well as when the noise rate increases from 1% to 10% (except for the coexistence of CURE with SoftMajority under $R = 2.6$). The higher the noise rate, the faster the steady state is reached.

c. CURE invades TFT-family strategies

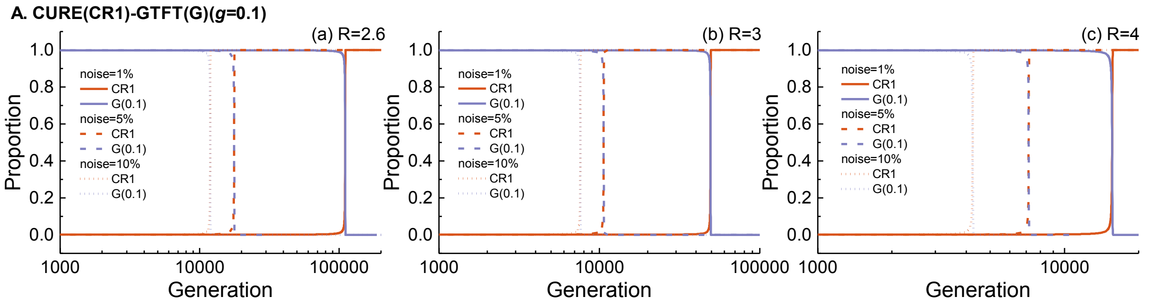


R= 2.6	1% CR1	1% CT	5% CR1	5% CT	10% CR1	10% CT
CR1	2.5971	2.54207	2.5813	2.50722	2.5526	2.46184
CT	2.54208	2.5238	2.50723	2.4846	2.46185	2.4380

R= 3.0	1% CR1	1% CT	5% CR1	5% CT	10% CR1	10% CT
CR1	2.9866	2.74011	2.9307	2.69266	2.8565	2.63405
CT	2.74012	2.6587	2.69267	2.6177	2.63406	2.57

R= 4.0	1% CR1	1% CT	5% CR1	5% CT	10% CR1	10% CT
CR1	3.96	2.23222	3.8042	3.15686	3.6163	3.06494
CT	2.23223	2.9921	3.15687	2.9499	3.06495	2.8998

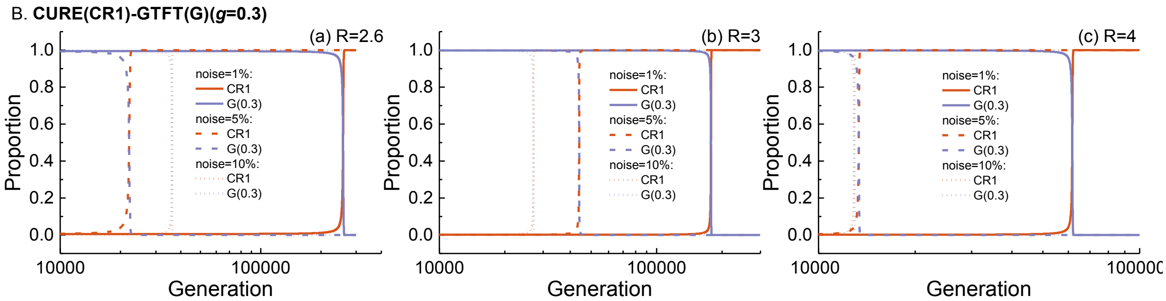
Supplementary Fig. 4. CURE invades CTFT populations. (a) - (c) show the evolutionary dynamics of the two strategies under R=2.6, 3, and 4. In general, it takes relatively few generations for CURE to invade CTFT. The invasion time decreases with R. When $R = 2.6$, the invasion time is shorter when the noise rate is higher; in contrast, when $R \geq 3$, the invasion occurs more quickly at lower noise (1%).



R= 2.6	1% CR1	1% G(0.1)	5% CR1	5% G(0.1)	10% CR1	10% G(0.1)
CR1	2.5971	2.572993	2.5813	2.45046	2.5526	2.350003
G(0.1)	2.572998	2.572995	2.45047	2.45045	2.350009	2.350011

R= 3.0	1% CR1	1% G(0.1)	5% CR1	5% G(0.1)	10% CR1	10% G(0.1)
CR1	2.9866	2.90805	2.9307	2.66763	2.8565	2.515286
G(0.1)	2.90806	2.90809	2.66764	2.66766	2.515292	2.5153

R= 4.0	1% CR1	1% G(0.1)	5% CR1	5% G(0.1)	10% CR1	10% G(0.1)
CR1	3.96	3.74585	3.8042	3.21037	3.6163	2.928579
G(0.1)	3.74586	3.7458	3.21038	3.2106	2.928584	2.92857

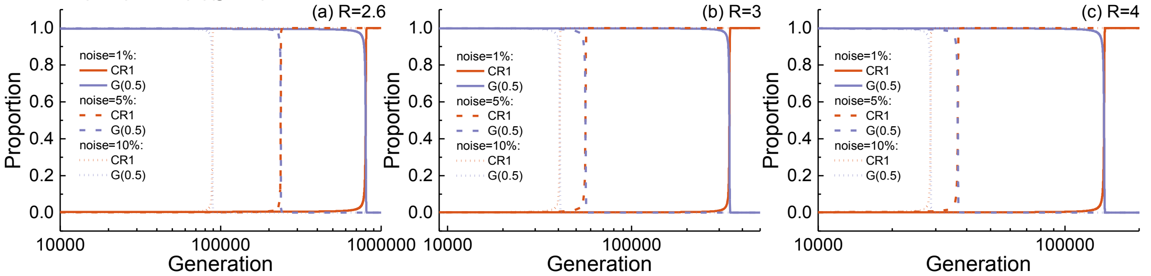


R= 2.6	1% CR1	1% G(0.3)	5% CR1	5% G(0.3)	10% CR1	10% G(0.3)
CR1	2.5971	2.592208	2.5813	2.547386	2.5526	2.482251
G(0.3)	2.592212	2.59223	2.547389	2.54742	2.482254	2.48225

R= 3.0	1% CR1	1% G(0.3)	5% CR1	5% G(0.3)	10% CR1	10% G(0.3)
CR1	2.9866	2.967132	2.9307	2.846576	2.8565	2.721023
G(0.3)	2.967135	2.967137	2.846579	2.8466	2.721026	2.72108

R= 4.0	1% CR1	1% G(0.3)	5% CR1	5% G(0.3)	10% CR1	10% G(0.3)
CR1	3.96	3.90449	3.8042	3.594729	3.6163	3.318153
G(0.3)	3.904493	3.90447	3.594732	3.5946	3.318157	3.31823

C. CURE(CR1)-GTFT(G)(*g*=0.5)



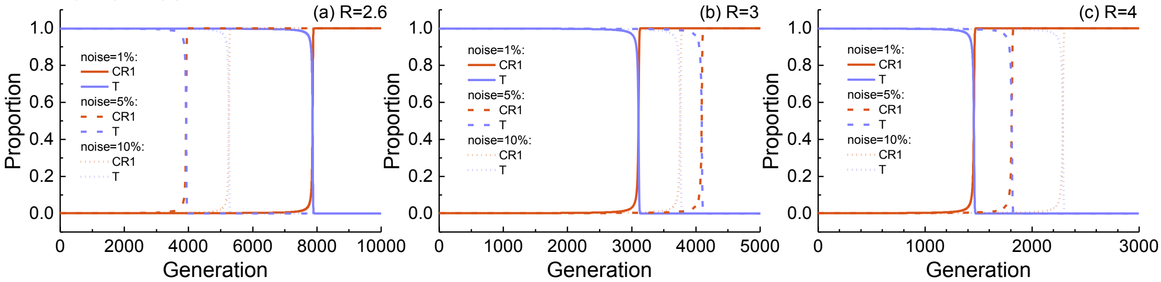
R=2.6	1%		5%		10%	
	CR1	G(0.5)	CR1	G(0.5)	CR1	G(0.5)
CR1	2.5971	2.59554	2.5813	2.57025	2.5526	2.52775
G(0.5)	2.59554	2.59555	2.57025	2.57025	2.52775	2.52774

R=3.0	1%		5%		10%	
	CR1	G(0.5)	CR1	G(0.5)	CR1	G(0.5)
CR1	2.9866	2.9800	2.9307	2.90084	2.8565	2.80558
G(0.5)	2.9800	2.9800	2.90084	2.90079	2.80558	2.80555

R=4.0	1%		5%		10%	
	CR1	G(0.5)	CR1	G(0.5)	CR1	G(0.5)
CR1	3.96	3.94118	3.8042	3.72724	3.6163	3.49993
G(0.5)	3.94118	3.94116	3.72724	3.72733	3.49993	3.5

Supplementary Fig. 5. CURE invades GTFT populations. For GTFT, the top row (A) illustrates the case of a generosity level of $g = 0.1$, the middle row (B) illustrates the case of $g = 0.3$, and the bottom row (C) illustrates the case $g = 0.5$. In (B), when $R = 2.6$, the critical initial proportion for CURE to invade is 0.005 under 1% noise, and the critical value under 5% is 0.004. In (C), the critical initial proportion is 0.004 at 1% noise when $R = 2.6$, and 0.002 at 5% and 10% noise when $R = 4.0$. The critical value under other conditions is 0.001.

CURE(CR1)-TFT(T)



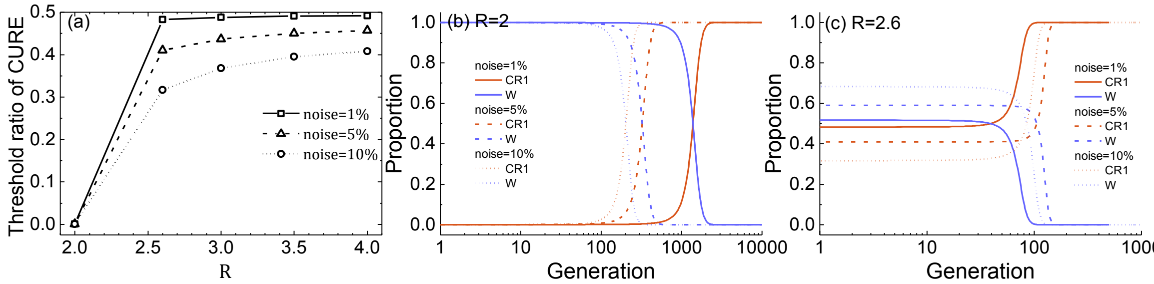
R=2.6	1%		5%		10%	
	CR1	T	CR1	T	CR1	T
CR1	2.5971	2.14981	2.5813	2.15008	2.5526	2.1500
T	2.14982	2.1501	2.15009	2.1498	2.15001	2.15

R=3.0	1%		5%		10%	
	CR1	T	CR1	T	CR1	T
CR1	2.9866	2.24982	2.9307	2.24966	2.8565	2.249992
T	2.24983	2.2498	2.24967	2.2499	2.25	2.249999

R=4.0	1%		5%		10%	
	CR1	T	CR1	T	CR1	T
CR1	3.96	2.50031	3.8042	2.50009	3.6163	2.50003
T	2.50032	2.4997	2.50010	2.49991	2.50004	2.50006

Supplementary Fig. 6. CURE invades TFT populations. (a) - (c) show the evolutionary dynamics of the two strategies under *R*=2.6, 3, and 4. In general, it takes fewer generations for CURE to invade TFT than invading GTFT, but more generations than invading CTFT. Invasion occurs more quickly when R is high. When $R = 2.6$, the required time for invasion is shorter when the noise rate is higher. In contrast, when $R \geq 3$, the time for invasion is shorter when there is little noise (1%).

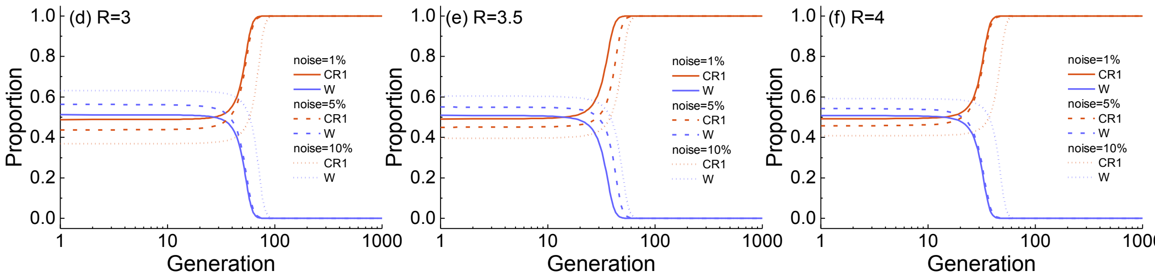
d. CURE invades WSLS



CURE(CR1)-WSLS(W)

R=2.0	1% CR1	1% W	5% CR1	5% W	10% CR1	10% W
CR1	2.0129	1.999995	2.0572	1.999997	2.0967	1.999987
W	2.000002	1.9904	2.000005	1.9595	1.999995	1.9360

R=3.5	1% CR1	1% W	5% CR1	5% W	10% CR1	10% W
CR1	3.4734	2.37503	3.3674	2.37492	3.2364	2.37502
W	2.37504	3.4315	2.37493	3.1850	2.37503	2.9351



R=2.6	1% CR1	1% W	5% CR1	5% W	10% CR1	10% W
CR1	2.5971	2.14996	2.5813	2.14998	2.5526	2.14996
W	2.14997	2.5668	2.14999	2.4497	2.14997	2.3356

R=3.0	1% CR1	1% W	5% CR1	5% W	10% CR1	10% W
CR1	2.9866	2.24998	2.9307	2.25000	2.8565	2.24998
W	2.24999	2.9511	2.25001	2.7765	2.24999	2.6020

R=4.0	1% CR1	1% W	5% CR1	5% W	10% CR1	10% W
CR1	3.96	2.49982	3.8042	2.50011	3.6163	2.49995
W	2.49983	3.9118	2.50012	3.5935	2.49996	3.2681

Supplementary Fig. 7. CURE invades WSLS populations. (a) The initial critical proportion of CURE to invade WSLS at different noise rates (accurate to 0.001). Below this value, CURE cannot invade WSLS. (b) - (f) show the evolutionary dynamics of the two strategies when the initial proportion of CURE is exactly at the critical point under $R = 2.0, 2.6, 3.0, 3.5,$ and 4.0, respectively. CURE can invade more quickly when R is large. At $R = 2.0$, this invasion occurs more quickly when there is a high level of noise. In contrast, when $R \geq 3$, CURE invades more quickly when there is little noise. When $R = 2.6$, noise has a non-monotonic effect on the speed of invasion. The invasion is fastest for a noise rate of 1%, and it is slowest for an intermediate noise rate of 5%.

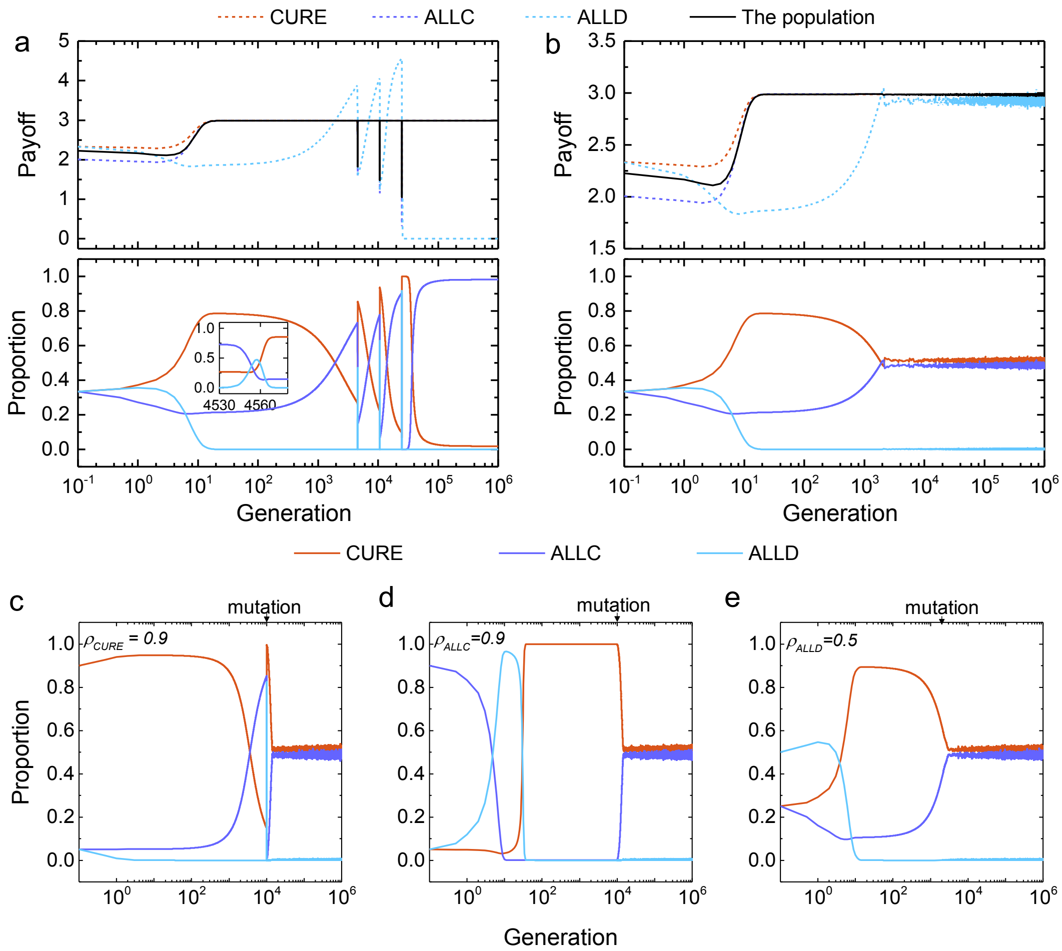
Supplementary Section 4. Evolutionary dynamics among ALLD, ALLC, and CURE

In this section, we complement the previous simulations among pairs of strategies by analyzing the evolutionary competition among three strategies, ALLD, ALLC, and CURE. Despite the overall strong performance of CURE, the previous results suggest that cumulative reciprocity is susceptible to invasion by unconditional cooperators. Once unconditional cooperators are common, they may in turn facilitate the emergence of defectors. A similar destabilizing effect has been reported for tit-for-tat; the three-strategy dynamics of ALLC, ALLD, and TFT have been studied, for example, by Imhof et al.[6] and Brandt & Sigmund[7]. Their analysis suggests that the instability of tit-for-tat against unconditional cooperators can eventually lead to a complete breakdown of cooperation. In the following, we explore the dynamics when we replace tit-for-tat by cumulative reciprocity.

a. Simulated dynamics in populations with ALLD, ALLC, and CURE

We investigate the dynamics of ALLC, ALLD, and CURE using the same simulations as in the two-strategy case (see Methods for detailed description). Initially, all three strategies are equally abundant. When we run simulations without mutations, we first observe cycles between cooperators, defectors, and reciprocators. Over time, however, defectors are driven to extinction. Eventually, the population consists of a large majority of unconditional cooperators, stabilized by a small fraction of cumulative reciprocators (Supplementary Fig.8a). In contrast, in simulations with a positive mutation rate, ALLD persists at small frequencies. As a result, the remaining population members are more likely to adopt cumulative reciprocity. Eventually, about 51.7% use CURE, 48.2% use ALLC, and 0.1% of the population use ALLD (Supplementary Fig.8b).

In addition to these simulation for populations with a uniform initial frequency distribution, we also explore the influence of different initial distributions. Specifically, we study three different initial distributions, namely CURE-dominant, ALLC-dominant, and ALLD-dominant. The results show that the initial distribution of strategies does not affect the strategies' eventual frequencies when mutations are introduced. Before the introduction of mutations, we observe the following dynamics. When CURE dominates the population initially (Supplementary Fig.8c), ALLC invades CURE. When ALLC dominates initially (Supplementary Fig.8d), ALLD first invades ALLC, but is replaced by CURE. Finally, when ALLD dominates initially (Supplementary Fig.8e), CURE first replaces ALLD, and then gets invaded by ALLC. Although ALLC can invade into CURE when the frequency of ALLD is very low, it does not completely replace CURE. Eventually, CURE and ALLC coexist.



Supplementary Fig. 8. Simulated dynamics among cooperators, defectors, and cumulative reciprocators. Here, we explore the competition between ALLC, ALLD, and CURE with simulations. Initially, all three strategies are assumed to be equally abundant. **a,** When strategies reproduce without mutations, we first observe cyclical dynamics. CURE is invaded by ALLC, which is invaded by ALLD, which in turn leads to the evolution of CURE (see also the inset that highlights this cyclic pattern at round the 4,560th generation). Eventually, however, the frequency of defectors approaches zero; the final population predominantly consists of unconditional cooperators. **b,** We have repeated the previous simulations, but this time we allow for mutations after 2,000 generations. Here, ALLC and CURE eventually coexist in almost equal proportions. We also examine different initial distributions, namely CURE-dominant (**c**), ALLC-dominant (**d**), and ALLD-dominant (**e**). Mutations are introduced at the 10,000th generation in plots **c** and **d**. The final abundance of each strategy is largely independent of the initial strategy distribution. Parameters: $T = 5, R = 3, P = 1, S = 0$, $\Delta = 1$, and 1% noise.

b. Replicator dynamics among ALLD, ALLC, and CURE

To further explore the evolutionary competition among these three strategies, we conduct a dynamic analysis between ALLD, ALLC, and CURE in an infinite population. According to the previously discussed Propositions 1 and 3, payoffs are given by the following payoff matrix:

$$\pi = \begin{pmatrix} \pi_{DD} & \pi_{DC} & \pi_{DR} \\ \pi_{CD} & \pi_{CC} & \pi_{CR} \\ \pi_{RD} & \pi_{RC} & \pi_{RR} \end{pmatrix} \tag{4.1}$$

with entries

$$\begin{aligned}
&\pi_{DD} = \pi_{DR} = \pi_{RD} = \varepsilon^2 R + \varepsilon(1-\varepsilon)(T+S) + (1-\varepsilon)^2 P,\\
&\pi_{CC} = \pi_{CR} = \pi_{RC} = (1-\varepsilon)^2 R + \varepsilon(1-\varepsilon)(T+S) + \varepsilon^2 P,\\
&\pi_{DC} = (1-\varepsilon)^2 T + \varepsilon(1-\varepsilon)(R+P) + \varepsilon^2 S,\\
&\pi_{CD} = (1-\varepsilon)^2 S + \varepsilon(1-\varepsilon)(R+P) + \varepsilon^2 T,\\
&\pi_{RR} = \frac{(1-\varepsilon)^2\cdot(1-2(1-\varepsilon)\varepsilon+2\Delta(1-2\varepsilon))}{1-2\varepsilon^2+2\Delta(1-2\varepsilon)}\cdot R + \frac{2\varepsilon(1-\varepsilon)\cdot(\Delta(1-2\varepsilon)+(1-\varepsilon)^2)}{1-2\varepsilon^2+2\Delta(1-2\varepsilon)}\cdot(T+S) +\\
&\frac{\varepsilon^2\cdot(3-2(3-\varepsilon)\varepsilon+2\Delta(1-2\varepsilon))}{1-2\varepsilon^2+2\Delta(1-2\varepsilon)}\cdot P.
\end{aligned} \tag{4.2}$$

We analyze the dynamics in a population of infinite size. Let x_D, x_C, and x_R denote the fraction of defectors, cooperators, and cumulative reciprocators, respectively. Then, the expected payoff of each strategy is given by

$$\begin{aligned}
\bar{\pi}_D &= \pi_{DD}\cdot x_D + \pi_{DC}\cdot x_C + \pi_{DR}\cdot x_R\\
\bar{\pi}_C &= \pi_{CD}\cdot x_D + \pi_{CC}\cdot x_C + \pi_{CR}\cdot x_R\\
\bar{\pi}_R &= \pi_{RD}\cdot x_D + \pi_{RC}\cdot x_C + \pi_{RR}\cdot x_R.
\end{aligned} \tag{4.3}$$

We define the average payoff of the population as $\bar{\pi} = \bar{\pi}_D x_D + \bar{\pi}_C x_C + \bar{\pi}_R x_R$. We model the evolutionary dynamics using the replicator equation:

$$\dot{x}_i = x_i(\bar{\pi}_i - \bar{\pi}). \tag{4.4}$$

That is, we assume the frequency of a strategy increases in time if and only if it performs better than average in the current population. A rock-scissors-paper-like cycle is inherent in this three-population game with payoffs obtained from Eq. (4.2). On the edge between ALLC and ALLD (i.e., $x_R = 0$), defectors succeed. On the edge between ALLD and CURE ($x_C = 0$), CURE takes over. Finally, on the edge between CURE and ALLC ($x_D = 0$), the cooperators succeed, thereby closing the cycle. It follows that there is no stable fixed point on the boundary of the state space.

We can find the position of the interior fixed point by setting $\bar{\pi}_D = \bar{\pi}_C = \bar{\pi}_R$ and solving for those x_D, x_C, x_R that satisfy the constraint $x_D + x_C + x_R = 1$. For a positive error rate and no mutations, there is a unique fixed point in the interior of the state space. In the limit of rare errors, $\varepsilon \to 0$, we obtain the following expression for the abundance of defectors, cooperators, and cumulative reciprocators in this fixed point as

$$x_D^* = 0,\quad x_C^* = \frac{R-P}{T-P},\quad x_R^* = \frac{T-R}{T-P}. \tag{4.5}$$

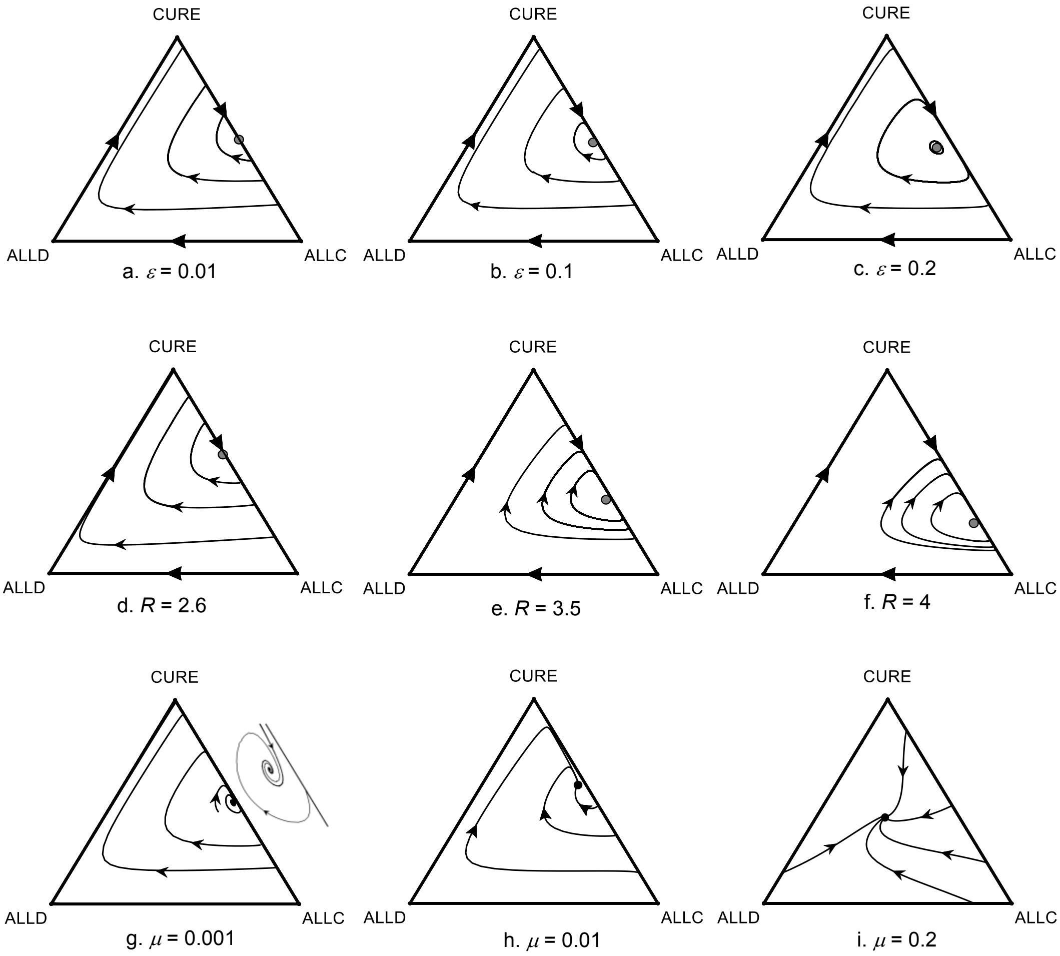
First, the abundance of strategies is independent of CURE's threshold Δ. Second, defectors are entirely missing in the unique equilibrium. For positive error rates, the equilibrium moves into the interior of the state space, but remains close to the fixed point shown in the Equation (4.5) when errors are rare. Using the techniques of Hofbauer and Sigmund (Section 7.7)[8], one can show that any such interior fixed point is surrounded by closed periodic orbits, suggesting that almost all initial populations result in oscillations between ALLC, ALLD, and CURE, as illustrated in Supplementary Fig. 9(a-f). As ε increases from 0.01 to 0.2, the fixed point moves away from the edge of ALLC and CURE. As R increases from 2.6 to 4.0, the fixed point moves along the edge of CURE and ALLC, from the vicinity of CURE to the vicinity of ALLC.

When mutations are introduced, by considering the following ordinary differential equation,

$$\dot{x}_i = x_i(\bar{\pi}_i - \bar{\pi} - 3\mu) + \mu, \tag{4.6}$$

where μ is the uniform mutation rate for all strategies[1], those oscillations around the fixed point break down. Instead, eventually all initial populations converge to the unique fixed point, and the interior fixed point becomes globally stable, as shown in Supplementary Fig. 9(g-i). As μ increases, the fixed point moves from the edge between ALLC and CURE to the center of the triangle. These results based on replicator dynamics are in line with our earlier simulation results.

Overall, the evolutionary dynamics among ALLC, ALLD, and CURE are remarkably different from the ALLC-ALLD-TFT system. Only when discriminators use cumulative reciprocity, cooperation can persist. Of course, this analysis for three particular strategies does not rule out that other exploitative strategies (different from ALLD) may eventually take over. However, it illustrates that cumulative reciprocity can act as an effective mechanism to restrain unconditional defectors.

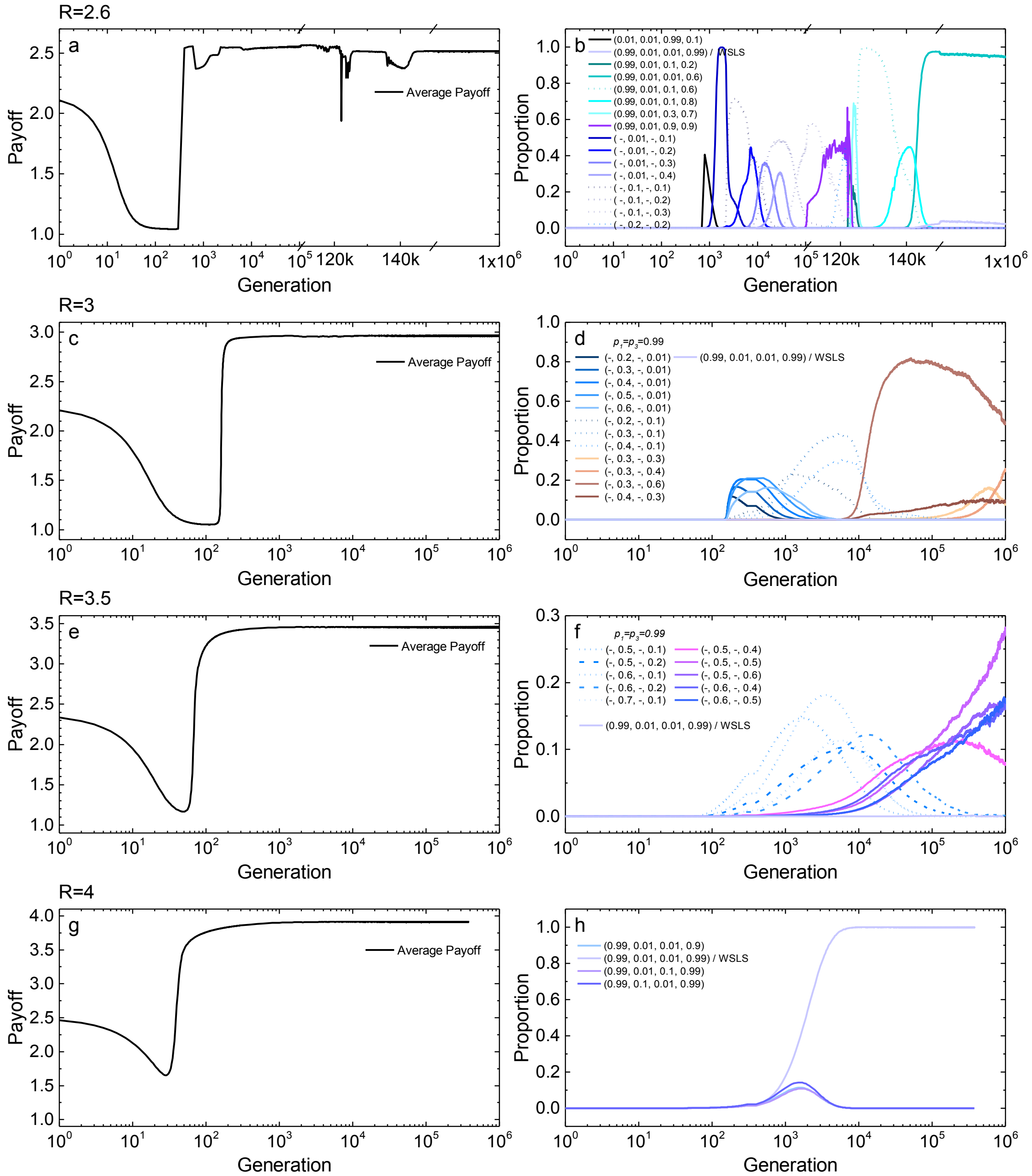


Supplementary Fig. 9. Replicator dynamics of cooperation, defection, and cumulative reciprocity. Panels **a~f** display solutions of the classical replicator equation (4.4) using the parameters $R = 3, S = 0, T = 5, P = 1$, and $\Delta = 1, \varepsilon = 0.1$ unless otherwise specified. Each corner represents a homogeneous population; each edge represents a population in which two of the three strategies coexist. The point indicates the unique fixed point of the dynamics. Curves in the interior indicate representative trajectories. Panels **a-c** show that the fixed point moves away from the edge of ALLC and CURE as ε increases. As R increases from 2.6 to 4.0, Panels **d-f** show that the fixed point moves along the edge between CURE and ALLC, from close to CURE to close to ALLC, reflecting the result of Eq. (4.5). Panels **g-i** represent solutions for the replicator equation with mutations, as defined by Eq. (4.6). Here, all orbits converge to the interior fixed point. As μ increases, the fixed point moves away from the edge between CURE and ALLC, towards the center of the triangle.

Supplementary Section 5. Evolution of CURE in memory-one populations

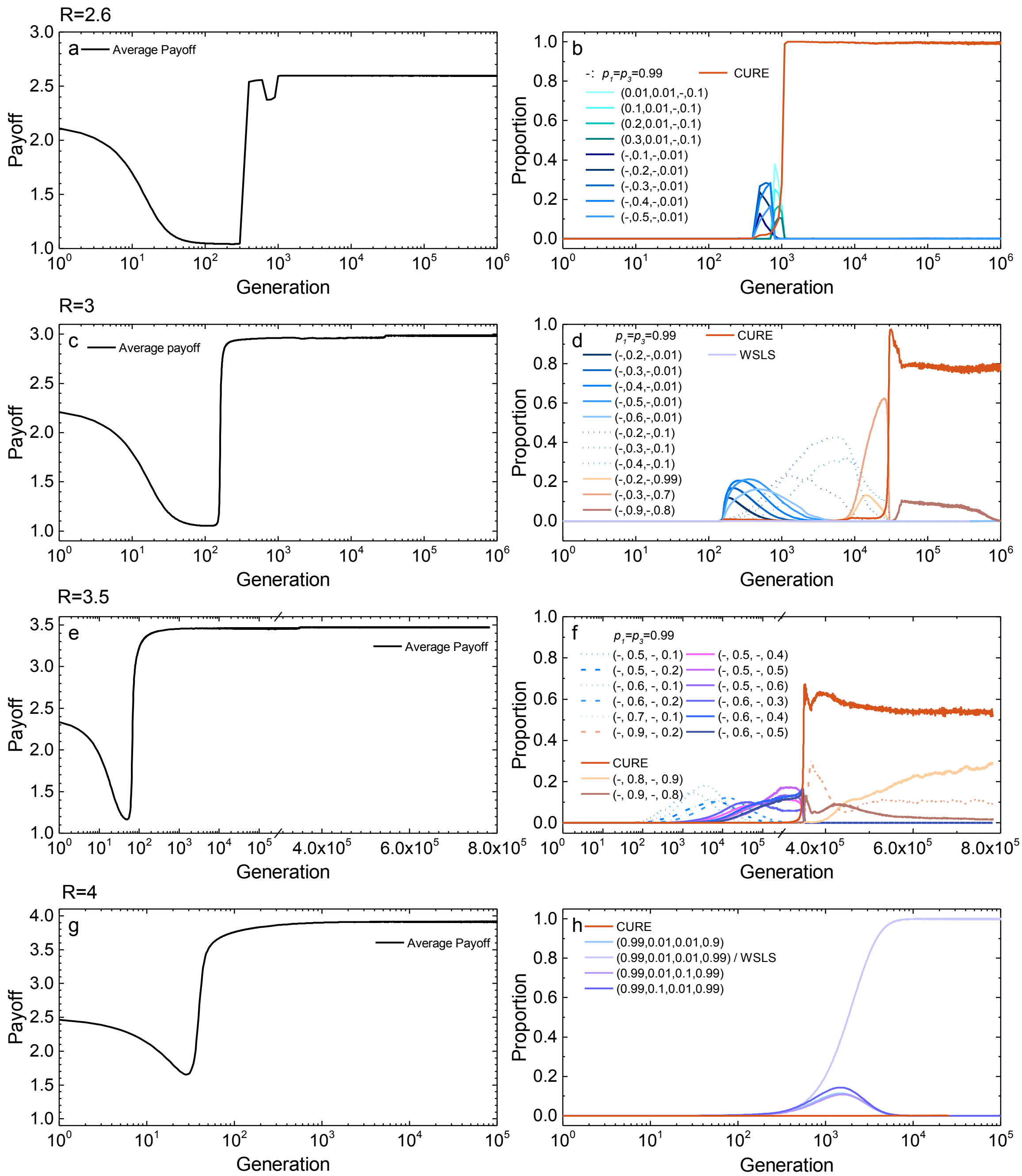
In this section, we report simulation results on the performance of CURE in heterogeneous populations of memory-1 players, as a supplement to the main text section "CURE and population dynamics". For these simulations, we let R range from 2.6 to 4.0, for a tolerance level $\Delta = 0, 1$ and 2, and we consider the dynamics with and without mutations, respectively.

a. The evolution of memory-one strategies without CURE (as a baseline)

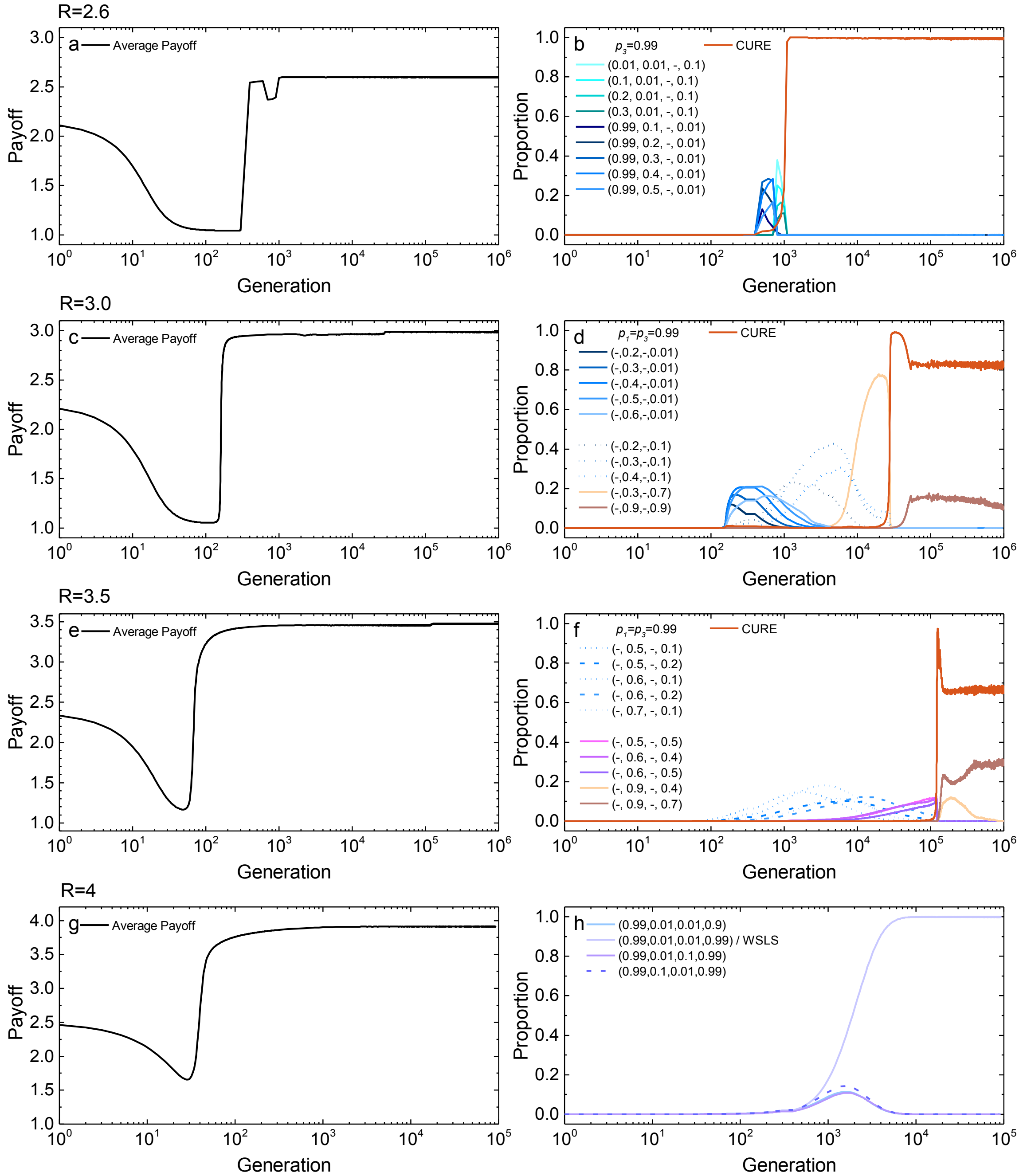


Supplementary Fig. 10. Evolution among memory-one strategies in the presence of mutations. Panels **a**, **c**, **e**, and **g** show the dynamics of the population's average payoff when $R = 2.6, 3.0, 3.5,$ and 4.0, respectively; whereas Panels **b**, **d**, **f**, and **h** illustrate the corresponding dynamics of the strategy frequencies. The winning strategy shifts from the TFT-family to WSLS as R increases from 3.0 to 4.0. When $R = 2.6$, GTFT-type strategies are predominant in around the first 100,000 generations. These GTFT-type strategies are then gradually replaced by strategies with a stronger tendency to defect. In general, we observe a declining average payoff during the first stage of the simulations, followed by a rapid increase, indicating the trend from defection to cooperation. However, when $R = 2.6$, cooperation is quite unstable. In Panels **b**, **d**, **f**, and **h**, the strategies that achieve a proportion of no-less-than 0.1 are displayed, with WSLS being specifically added. The same is true for the following figures of this section.

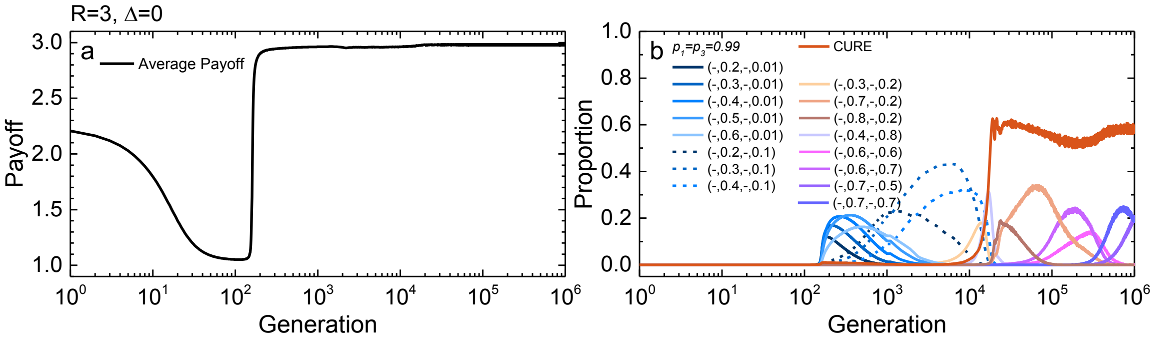
b. Incorporating CURE into the pool of memory-one strategies



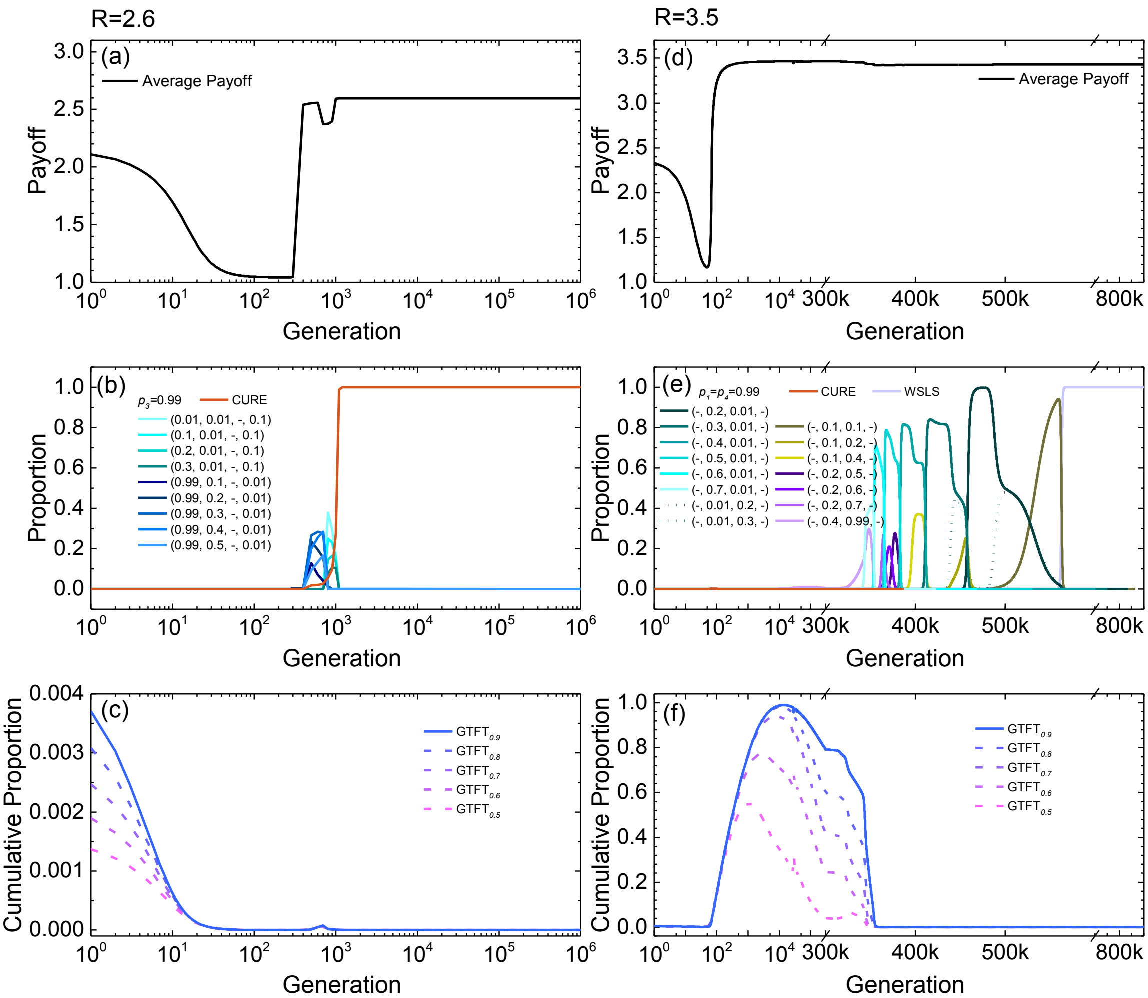
Supplementary Fig. 11. Co-evolution of CURE (Δ= 1) and memory-one strategies in the presence of mutations. The wine-colored line in **b, d, f**, and **h** indicates the CURE strategy. Panels **a** and **b** reflect the case of $R = 2.6$. Here, TFT-family strategies (0.99, 0.1~0.5, 0.99, 0.01) prevail and the population's average payoff rapidly increases in the early stage of simulation. They are then overtaken by more exploitative strategies of (0.01~0.3, 0.01 0.99, 0.1) and the average payoff decreases around the 1000th generation. Eventually, CURE invades those strategies and takes over. When $R = 3.0$ and $R = 3.5$, CURE tends to coexist with various highly cooperative strategies. As R increases, the frequency of CURE decreases. When $R = 4.0$, WSLS rapidly dominates.



Supplementary Fig. 12. Co-evolution of CURE (Δ= 2) and memory-one strategies in the presence of mutations. The results are generally consistent with those in the previous Supplementary Fig. 11 for the case of Δ= 1. Panels **a**, **c**, **e**, **g** show the evolution of population average payoff under *R*=2.6, 3.0, 3.5, and 4. Panels **b**, **d**, **f**, **h** reflect the evolution of strategies correspondingly.



Supplementary Fig. 13. Co-evolution of CURE (Δ= 0) and memory-one strategies in the presence of mutations. Here, we depict simulations for $R = 3$, panel **a** shows the evolution of payoff and **b** shows the evolution of strategies. The results show that CURE still occupies around 60% of the population, although the proportion of CURE is lower than that in the case of higher Δ.

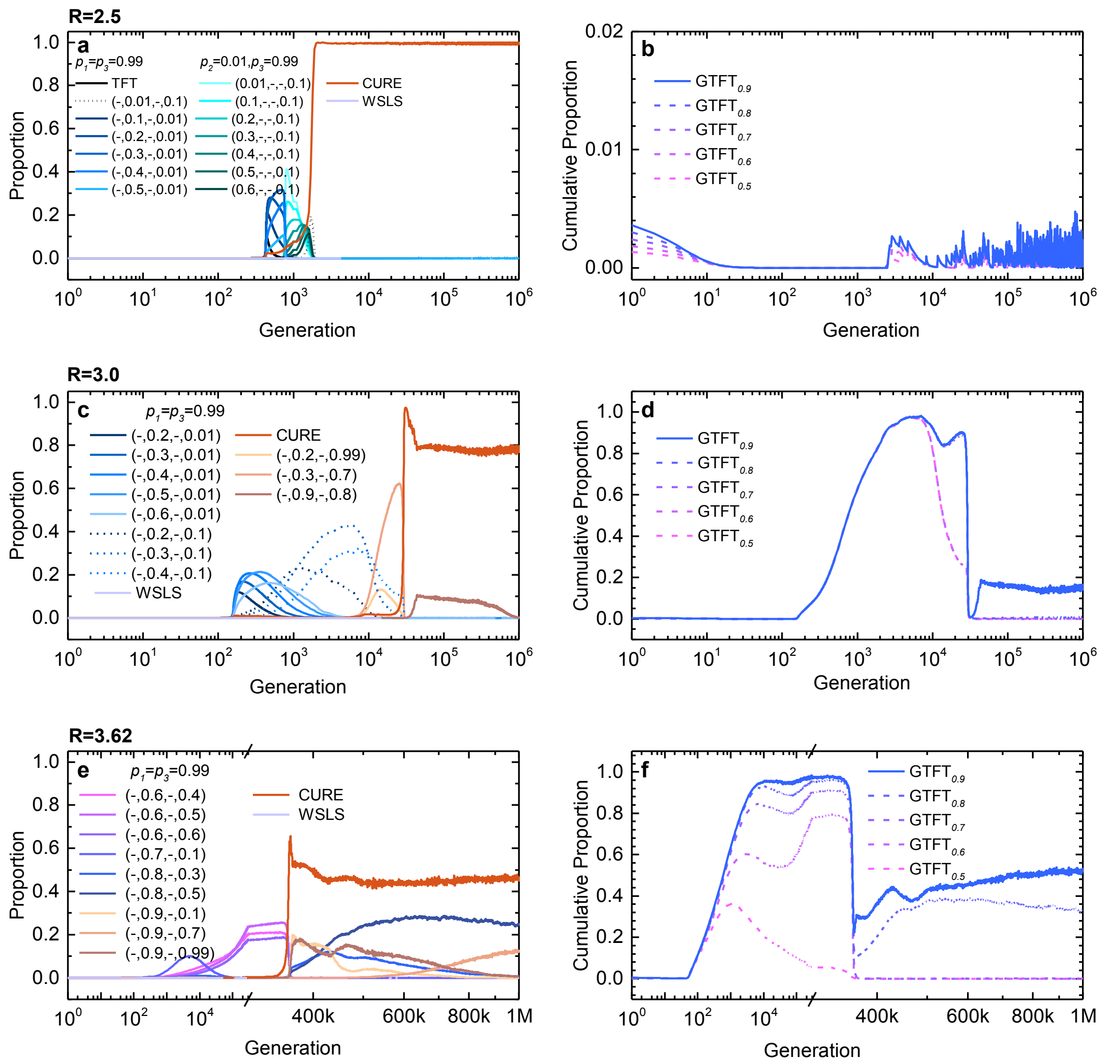


Supplementary Fig. 14. Evolutionary dynamics without mutations when $\Delta = 1$**.** This figure supplements Panel (f) of Fig. 3 in the main text. The dynamics of the population's average payoff are exhibited in Panels (a) and (d), respectively, for $R = 2.6$ and 3.5. Correspondingly, Panels (b) and (e) represent the evolution of strategies that reach at least a proportion of 10% of the population. Panels (c) and (f) depict the evolutions of GTFT-type strategies. For $R = 2.6$, the results show that CURE takes over the population after the 1,300th generation. By comparison, when $R = 3.5$, the population eventually adopts WSLS. Here, the entire process is characterized by the continuous competition of a series of strategies with the form of $(0.99, p_2, p_3, 0.99)$, and WSLS eventually dominates the population. By comparing the three cases $R = 2.6, 3.0$ [Panel(f) of Fig. 3 in the main text], and 3.5, we see a transition from the absolute dominance of CURE (for $R = 2.6$) to a high proportion of CURE (for $R = 3.0$) and then to an absolute dominance of WSLS (for $R = 3.5$). This trend is the same as in the simulations with mutations.

Supplementary Section 6. The critical value of R for the transition from CURE to GTFT and WSLS

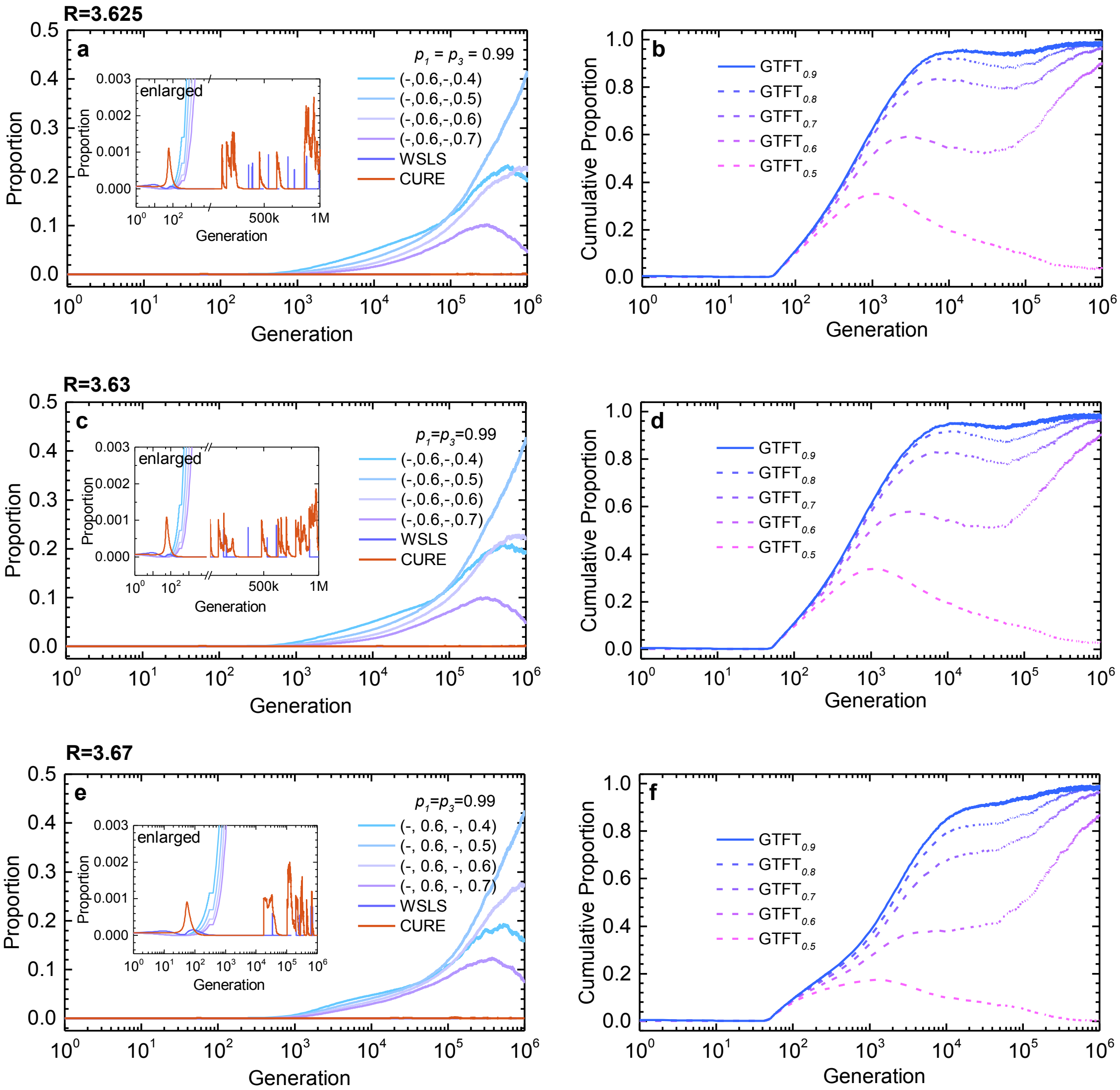
The previous results show how the most dominant strategy changes as we change the reward R for mutual cooperation, from CURE to GTFT and then to WSLS. In the following, we explore this transition in more detail, to support the respective discussion (Fig. 4) in the main text.

a. Strategy CURE dominates the population when $R < R_1^* = 3.625$



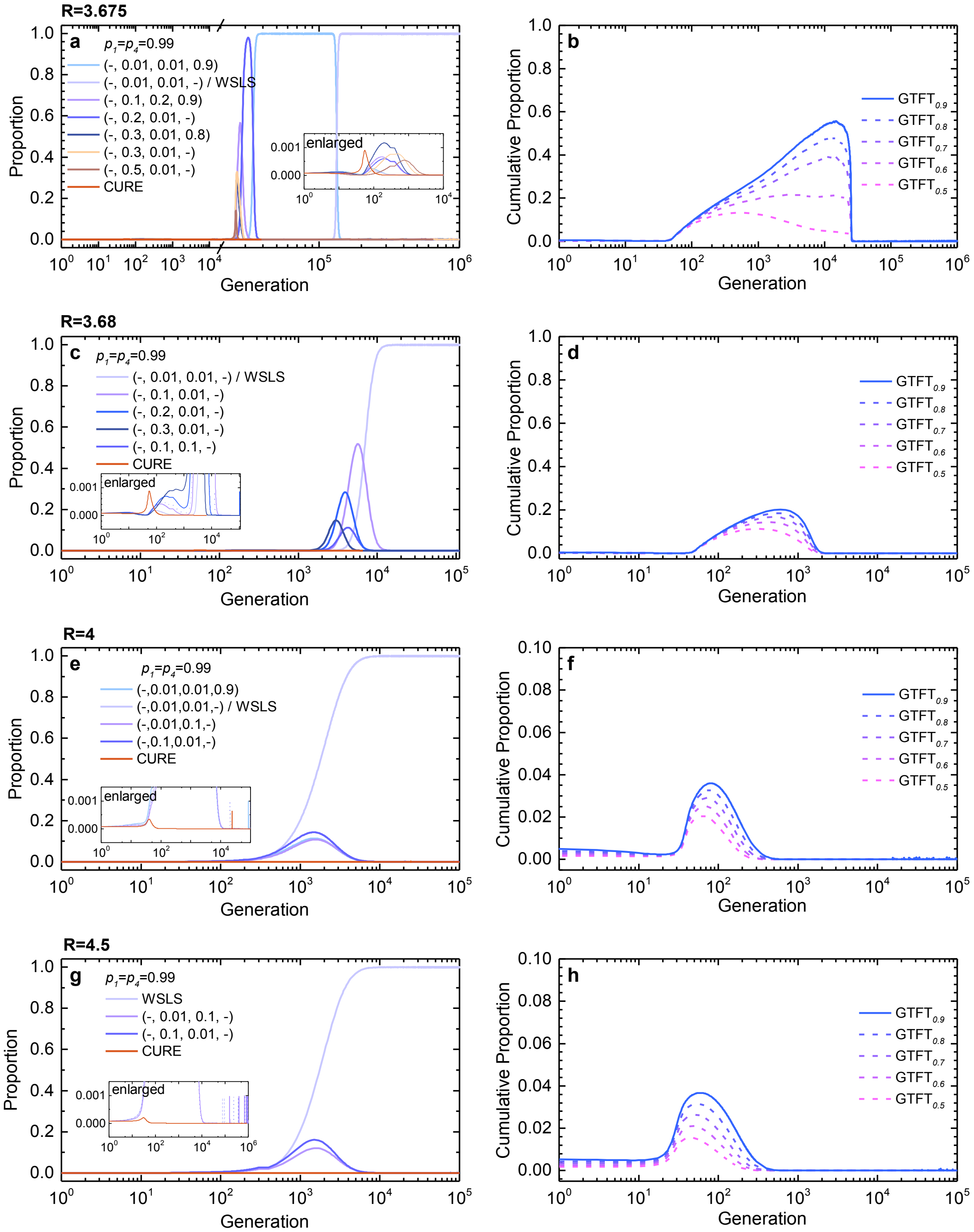
Supplementary Fig. 15. Frequency dynamics of major strategies in the cases of $R =$ $2.5, 3.0,$ and 3.62. Panels **a, c,** and **e** show the frequency dynamics of the major strategies. Panels **b, d**, and **f** show the cumulative proportions for different sets of GTFT strategies (i.e., $GTFT_{0.5}$, $GTFT_{0.6}$, …, $GTFT_{0.9}$). Here, the major strategies refer to those strategies that are played by at least 10% of the population during evolution (the same applies to the following figures). When $R = 2.5$, after approximately 1,000 generations, CURE is adopted by 99% of the population. When $R = 3.0$, CURE is still adopted by more than 80% of the population. Finally, when $R = 3.62$, the pattern is similar to the case of $R = 3.0$. However, the overall proportion of GTFT-type strategies is now around 50%, while cumulative reciprocators make up the remaining 50% of the population. The frequency of WSLS is negligible in this range of R.

b. GTFT-type strategies dominate the population when $R \in [3.625, 3.675)$



Supplementary Fig. 16. Frequency dynamics of major strategies in the cases of $R = 3.625, 3.63$, and 3.67. Panels **a, c**, and **e** show the frequency dynamics of the major strategies. Panels **b, d**, and **f** show the cumulative proportions for different sets of GTFT strategies (i.e., $GTFT_{0.5}$, $GTFT_{0.6}$, …, $GTFT_{0.9}$). When $R = 3.625$, the advantage of GTFT becomes apparent. GTFT-type strategies, more specifically the strategies (0.99, 0.6, 0.99, 0.4~0.7), keep prevailing after around 100 generations. CURE and WSLS may temporarily occur through mutation, but they rapidly diminish afterwards, as shown in the inset of Panel **a**. When $R = 3.63$ and 3.67, the same pattern can be observed.

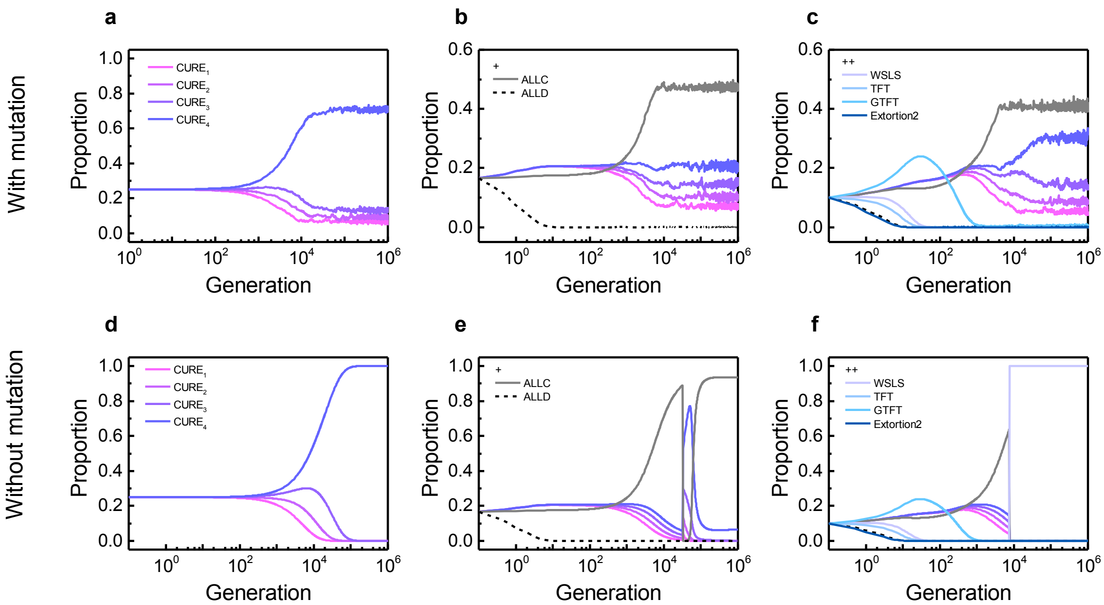
c. Strategy WSLS dominates the population when $R \geq 3.675$



Supplementary Fig. 17. Frequency dynamics of major strategies in the cases of $R = 3.675, 3.68, 4$, and 4.5. In Panels **a** and **b**, GTFT-type strategies prevail in the first 26,000 generations, at which point WSLS takes over. The strategy (0.99, 0.2, 0.01, 0.99) first briefly dominates and is then replaced by a more retaliatory strategy (0.99, 0.01, 0.01, 0.9). Finally, WSLS dominates with a proportion of 0.999 at the 150,000th generation. As R increases from 3.68 to 4.5 (Panels **c-f**), WSLS dominates the population more rapidly, and CURE's short-term advantage in the first 100 steps disappears, as shown in the insets in Panels **a, c, e**, and **g**.

Supplementary Section 7. Co-evolution of CURE strategies with different thresholds

In our previous evolutionary analyses, we considered one instantiation of CURE at a time. That is, we assumed that all cumulative reciprocators in a population use the same threshold Δ. In the following, we explore the case when different variants of CURE compete, similar to the static comparison in Supplementary Tables 3 and 4. To this end, we consider three kinds of competitions. First, we only consider the four variants $CURE_1$, $CURE_2$, $CURE_3$, and $CURE_4$. Second, we additionally allow for the unconditional strategies ALLC and ALLD. And third, we also allow for the strategies WSLS, TFT, GTFT and Extortion2. The results are displayed in Supplementary Fig. 18.



Supplementary Fig. 18. Evolutionary competition among different variants of cumulative reciprocity. To explore the performance of different variants of CURE, we implement simulations similar to the ones before. Simulations either allow for mutations (Panels **a-c**), or they do not (Panels **d-f**). We consider three scenarios. Either only CURE strategies compete (left column), CURE strategies compete with ALLC and ALLD (middle column), or CURE strategies additionally compete with WSLS, TFT, GTFT, and Extortion2 (right column).

We observe the following regularities. First, when only different variants of CURE compete, then the most tolerant variant is most abundant, as expected from the static payoffs displayed in Supplementary Tables 3 and 4 (Supplementary Fig. 18a,d). Second, if we additionally allow for unconditional strategies, we still observe the evolution of high cooperation rates. Here, ALLC tends to co-exist with different variants of cumulative reciprocity (Supplementary Fig. 18b,e). Finally, we observe the same qualitative result when we add WSLS, TFT, GTFT, and Extortion2, and when mutations continually introduce rare strategies (Supplementary Fig. 18c). However, without mutations, WSLS eventually takes over (Supplementary Fig. 18f). The inclusion of several variants of CURE thus helps to sustain cooperation, and it can lead to stable coexistences between cumulative reciprocators and unconditional cooperators.

Supplementary Section 8. Comparison of CURE with selected memory-*k* strategies

When analyzing the evolutionary performance of CURE, so far we have focused on comparisons between CURE and memory-1 strategies. The space of memory-1 strategies is arguably among the best explored strategy spaces for direct reciprocity[1]. This space is rich enough to include many well-known strategies (such as ALLD, TFT[2], and WSLS[9,10]). At the same time, the space is small enough to be explored systematically. While a comparison of CURE to more general memory-*k* strategies might be desirable, the mere size of these strategy spaces makes such a comparison difficult. Even if we restrict attention to deterministic strategies, there are 2^{16} = 65,536 memory-2 strategies and $2^{64} \approx 1.8 \cdot 10^{19}$ memory-3 strategies. In the following, we thus confine ourselves to compare CURE to a small number of select memory-2 and memory-3 strategies highlighted by the literature. We describe these strategies in the following.

One strategy class highlighted previously is the class of so-called All-or-None strategies[11,12]. An AON_k strategy cooperates initially, and it continues to cooperate as long as both players picked the same action in each of the past *k* rounds. However, if the two players chose a different action in one of the past *k* rounds, an AON_k player defects. These strategies have a number of desirable properties: when two AON_k players meet, they mutually cooperate; when they are exploited, they punish the co-player with k rounds of defection; and when one player defected by mistake, a group of AON_k players restores cooperation after *k* rounds. Moreover, it can be shown that All-or-None strategies are a Nash equilibrium of the (infinitely) repeated prisoner's dilemma if the parameter *k* is sufficiently large[12]. In the special case of k=1, AON_1 reproduces the well-known WSLS strategy.

Another interesting strategy is the memory-2 strategy *TFT-ATFT* (Tit-for-Tat Anti-Tit-for-Tat), introduced by Yi, Ki Baek, and Choi[13]. According to this strategy, a player normally uses *TFT*; however, after committing an error, the player turns to *ATFT* and returns to *TFT* either when mutual cooperation is recovered, or when the opponent unilaterally defects twice in a row. Similar to AON_k, this strategy has a number of remarkable properties: it is "efficient" (meaning that two *TFT-ATFT* players obtain the mutual cooperation payoff); it is "distinguishable" (it is able to identify ALLC players and to exploit them); and finally, it is "defensible" (in the absence of errors, it cannot be exploited by any other strategy).

Finally, we also consider the memory-3 strategy *CAPRI*, first introduced by Murase and Ki Baek[14]. *CAPRI* is an instance of a "friendly rival": With such a strategy, a player can prevent exploitation by any other strategy, while making sure that both players get the mutual cooperation payoff when both players use the *CAPRI* strategy. *CAPRI* is defined by the following table[14], that describes the focal player's next action depending on the outcome of the past three rounds:

Focal player's actions in previous rounds (*t*-3,*t*-2,*t*-1)	Co-player's actions in previous rounds (*t*-3, *t*-2, *t*-1)							
	CCC	CCD	CDC	CDD	DCC	DCD	DDC	DDD
CCC	C	D	D	D	C	D	D	D
CCD	C	D	C	D	D	D	D	D
CDC	D	C	D	D	C	D	D	D
CDD	D	D	D	D	D	D	D	D
DCC	C	D	C	D	C	D	C	D
DCD	D	D	D	D	D	D	D	D
DDC	D	D	D	D	C	D	C	C
DDD	D	D	D	D	D	D	C	D

During the first rounds of the game (i.e., when players do not have the outcome of three previous rounds in memory), one may choose the action as if players always cooperated prior to the first round. To compare these strategies to *CURE*, we first use simulations to compute the respective payoff matrix for three different error rates, see Supplementary Table 9.

Supplementary Table 9. Payoff matrixes for CURE (Δ=1) and selected memory-*k* strategies. $R = 3$, $S = 0$, $T = 5$, $P = 1$.

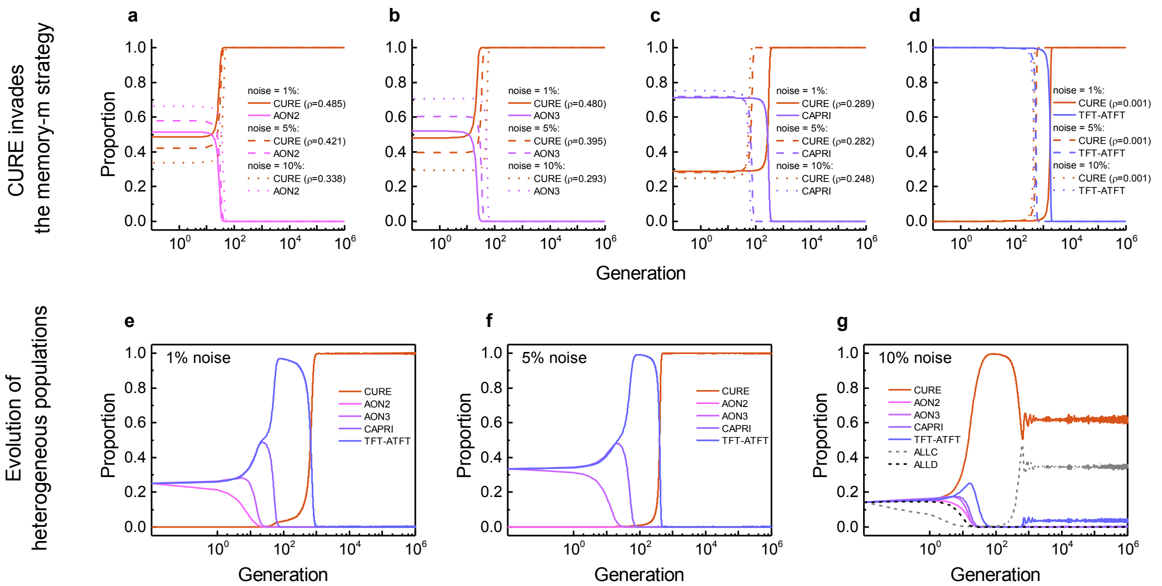
1% noise	CURE	AON_2	AON_3	CAPRI	TFT-ATFT
CURE	2.986557	1.762267	1.607052	2.778874	2.925951
AON2	1.762268	2.913053	1.268792	1.891112	1.926866
AON3	1.607053	2.475399	2.875767	1.740724	1.768704
CAPRI	2.778874	1.891595	1.723237	2.863047	2.308540
TFT-ATFT	2.925951	1.873180	1.706650	2.275944	2.917400

5% noise	CURE	AON_2	AON_3	CAPRI	TFT-ATFT
CURE	2.930674	1.807408	1.639563	2.212150	2.714125
AON_2	1.807409	2.621748	1.345135	1.896018	2.059097
AON_3	1.639564	2.385384	2.481695	1.789358	1.921958
CAPRI	2.212151	1.906516	1.715254	2.493231	2.255157
TFT-ATFT	2.714126	1.825519	1.641596	2.104119	2.689697

10% noise	CURE	AON_2	AON_3	CAPRI	TFT-ATFT
CURE	2.856513	1.858170	1.688148	1.942936	2.555169
AON_2	1.858171	2.365837	1.442986	1.898546	2.185789
AON_3	1.688149	2.292817	2.172193	1.841235	2.087034
CAPRI	1.942937	1.932507	1.723010	2.242920	2.251159
TFT-ATFT	2.555170	1.795748	1.596147	1.973461	2.528471

Several remarks are in order. First, we note that the four strategies *CURE*, AON_2, AON_3, and *CAPRI* cannot be invaded by any other of these five strategies. In contrast, *TFT-ATFT* can be invaded by *CURE* for all considered error probabilities. This observation suggests that while *TFT-ATFT* is defensible as the error rate approaches zero, it ceases to be robust for positive error probabilities. Second, among the first four strategies, it is CURE that achieves the highest payoff against itself. This advantage is particularly pronounced in the 10% noise condition. Here, CURE still achieves a payoff close to the social optimum ($\pi = 2.856$ compared to $R = 3$), whereas all other strategies obtain a payoff of 2.528 or less. This observation, together with our result that CURE enforces fairness against any other player (Proposition 4) implies that CURE is risk-dominant[15] in any direct competition with either *AON₂*, *AON₃*, or *CAPRI*.

To further explore this latter observation, we have conducted further simulations, following the basic scheme of the simulations shown in SI Section 3 ("Evolutionary dynamics in two-strategy populations"). That is, we match CURE in a pairwise competition with each of the other four strategies. In each case we observe that CURE can evolve even if it is initially adopted only by a relative minority of the population, as shown in Supplementary Fig. 19a-d. With further simulations, we show that CURE also evolves when it is simultaneously competing with all four other strategies (Supplementary Fig.19e,f), or when additionally ALLC and ALLD are added to the population (Supplementary Fig.19g).



Supplementary Fig. 19. Evolutionary dynamics among CURE, AON$_2$, AON$_3$, CAPRI, and TFT-ATFT. a-d, Using the same simulation scheme as in SI Section 3, we first explore the evolutionary dynamics when CURE competes with each of the other strategies in isolation. In each case we find that CURE can evolve even if its initial share is less than 50% of the population. **e,f,** We also explore the evolutionary dynamics when all five strategies are present in the population simultaneously. We find that CURE can evolve even if it is initially played only by 0.1% of the population. **g,** When the error rate is large, and when we additionally include unconditional cooperators and defectors, we find that populations eventually converge to a mixture of CURE, ALLC, and TFT-ATFT.

Supplementary Section 9. A variant of cumulative reciprocity with discounting

The strategy of cumulative reciprocity that we considered in the main text weighs all past decisions equally. If Alice adopts CURE and Bob unilaterally defected many rounds ago, this has the same impact on Alice's next decision as a unilateral defection in the very last round. Instead, one may consider variants of cumulative reciprocity in which the most recent events receive relatively more weight. One way to incorporate this idea is by assuming that Alice updates her defection difference statistic as follows. Initially, she sets $d(0) = 0$. After each round, this statistic is then updated according to the rule,

$$d(t+1) = \lambda d(t) + \rho_{xy}(t). \tag{9.1}$$

In this expression, λ is a parameter that determines the weight of past decisions, with $0 \leq \lambda \leq 1$. The second parameter $\rho_{xy}(t)$ encodes the outcome of the very last round, and it is defined by

$$\rho_{CC}(t) = \rho_{DD}(t) = 0, \quad \rho_{CD}(t) = 1, \quad \rho_{DC}(t) = -1. \tag{9.2}$$

In the special case of no discounting of the past, $\lambda = 1$, this notion of cumulative reciprocity recovers the version of CURE we have explored throughout. In the other limit, for $\lambda = 0$, and setting the tolerance level to $\Delta = 0$, this notion of cumulative reciprocity instantiates the classical strategy Firm-but-Fair. For intermediate values of λ, this strategy implements a discounted version of cumulative reciprocity, where past cooperation imbalances are gradually forgotten. We refer to such a strategy as D-CURE.

While a comprehensive analysis of this model extension is beyond the scope of this paper, in the following we briefly mention results for a particular instance of D-CURE, for which we set $\lambda = 0.99$ and $\Delta = 1$. In Supplementary Table 10, we report numerically estimated payoffs of this strategy against itself, against CURE, and against the nine strategies considered in Supplementary Section 2, "Payoffs and cooperation rates obtained by computer simulations". The data is obtained by computing an average of 10^4 interactions, each running for 10^7 rounds. For the game parameters, we use the usual values, $R = 3, T = 5, S = 0, P = 1$.

A few aspects of these results are worth highlighting. Perhaps most importantly, D-CURE does no longer seem to enforce fairness. For example, for an error rate of 10%, the payoff of D-CURE against ALLD is approximately 1.257, whereas ALLD's payoff is 1.407. This suggests that D-CURE is less effective in invading populations of defectors. On the other hand, again for an error rate of 10% the payoff of D-CURE against ALLC is 2.924, which now exceeds ALLC's payoff of 2.835. In particular, while CURE is vulnerable to invasion by ALLC, the discounted version D-CURE is stable against ALLC. These results suggest that there might be an intermediate value of λ that optimally balances D-CURE's robustness against ALLC and D-CURE's competitiveness with ALLD. Finding this optimal value of λ could be an exciting direction for future explorations.

Supplementary Table 10. Average payoffs and cooperation rates of D-CURE against selected strategies. From left to right, payoffs and cooperation rates are obtained for no noise, 1%, 5%, and 10% noise levels. We consider games between D-CURE and selected alternative strategies.

Gaming Strategy	**Noise rates**							
	No noise (**0%**)		Low (**1%**)		Medium **(5%)**		High (**10%**)	
	Payoff	**Cooperation rates**	**Payoff**	**Cooperation rates**	**Payoff**	**Cooperation rates**	**Payoff**	**Cooperation rates**
D-CURE	3	1	2.984906	0.985104	2.922686	0.927441	2.841460	0.859556
D-CURE	3	1	2.984906	0.985104	2.922686	0.927441	2.841460	0.859556
d(DCR-DCR)	**0**	**0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
D-CURE	3	1	2.984827	0.985026	2.922678	0.927435	2.841455	0.859547
CURE	3	1	2.984827	0.985026	2.922678	0.927435	2.841455	0.859547
d(DCR-CR)	**0**	**0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
D-CURE	3	1	2.730337	0.787254	2.383898	0.580915	2.307247	0.541968
TFT	3	1	2.759060	0.781509	2.424340	0.572827	2.349237	0.533569
d(DCR-T)	**0**	**0**	**-0.028723**	**0.005745**	**-0.040442**	**0.008088**	**-0.04199**	**0.008399**
D-CURE	3	1	2.976902	0.980789	2.868607	0.885099	2.716358	0.765742
CTFT	3	1	2.982879	0.979594	2.880817	0.882656	2.735473	0.761919
d(DCR-CT)	**0**	**0**	**-0.005977**	**0.001195**	**-0.01221**	**0.002443**	**-0.019115**	**0.003823**
D-CURE	3	1	2.967402	0.970130	2.847221	0.866855	2.721822	0.775022
GTFT(0.3)	3	1	2.970500	0.969511	2.850922	0.866114	2.726854	0.774016
d(DCR-GT)	**0**	**0**	**-0.003098**	**0.000619**	**-0.003701**	**0.000741**	**-0.005032**	**0.001006**
D-CURE	3	1	2.997715	0.986073	2.970814	0.938044	2.923798	0.882209
ALLC	3	1	2.978079	0.989999	2.911036	0.949997	2.834860	0.900001
d(DCR-C)	**0**	**0**	**0.019636**	**-0.003926**	**0.059778**	**-0.011953**	**0.088938**	**-0.017792**
D-CURE	3	1	2.997714	0.986070	2.970257	0.937180	2.917295	0.879199
SoftMajority	3	1	2.978079	0.989999	2.910534	0.949098	2.829291	0.896798
d(DCR-SM)	**0**	**0**	**0.019635**	**-0.003929**	**0.059723**	**-0.011918**	**0.088004**	**-0.017599**
D-CURE	3	1	2.384956	0.606663	2.261643	0.526015	2.245444	0.515243
WSLS	3	1	2.424920	0.598671	2.305651	0.517214	2.289832	0.506365
d(DCR-W)	**0**	**0**	**-0.039964**	**0.007992**	**-0.044008**	**0.008801**	**-0.044388**	**0.008878**
D-CURE	0.985673	0.014327	1.013055	0.026676	1.122019	0.074274	1.256928	0.130069
ALLD	1.057307	0	1.096441	0.009999	1.243369	0.050002	1.407251	0.100002
d(DCR- D)	**-0.071634**	**0.014327**	**-0.083386**	**0.016677**	**-0.12135**	**0.024272**	**-0.150323**	**0.030067**
D-CURE	1.070196	0.052638	1.218306	0.119151	1.572208	0.266847	1.800528	0.352855
Extort2	1.140360	0.038605	1.291348	0.104542	1.647835	0.251722	1.873336	0.338294
d(DCR-E2)	**-0.070164**	**0.014033**	**-0.073042**	**0.014609**	**-0.075627**	**0.015125**	**-0.072808**	**0.014561**
D-CURE	3.0	1.0	2.990373	0.985114	2.925332	0.913859	2.820449	0.834824
HardMajority	3.0	1.0	2.971115	0.989021	2.870013	0.924801	2.746344	0.849603
d(DCR-HM)	**0.0**	**0.0**	**0.019258**	**-0.003907**	**0.055319**	**-0.010942**	**0.074105**	**-0.014779**

Supplementary Section 10. Cumulative reciprocity beyond the prisoner's dilemma.

Although the prisoner's dilemma has become the major paradigm to study cooperation[1,16], not all social dilemmas are readily described by this model. Instead, in many applications, the players' payoffs in any given round are not determined by a fixed payoff matrix. Rather, the player's action today modifies the feasible payoffs in future[17-19]. Similarly, many existing social dilemmas involve more than two players. In the following, we describe how cumulative reciprocity can be extended to these more general scenarios.

Stochastic prison's dilemma games setup.

We first consider the case that payoffs fluctuate in time, depending on the player's previous actions. To this end, we use stochastic games[20], in which players can be in different states over time. Fig.5a shows an example. In this game, there are two players who experience two possible states. In each state, players interact in a different prisoner's dilemma. The first state corresponds to a more profitable environment in which mutual cooperation yields high benefits. The second state corresponds to a depleted environment with lower payoffs. Players only find themselves in the more profitable state if they both cooperated in the previous round; if one or both players defected, they experience the depleted environment.

To extend the definition of cumulative reciprocity to stochastic games, we assume that a cumulative reciprocator now has separate defection difference statistics $d_s(k)$ for each possible state s. When the co-player Bob defects in state s while Alice cooperates, $d_s(k)$ increases by one. In the converse scenario, in which Alice defects and Bob cooperates, $d_s(k)$ decreases by one. CURE cooperates in a given state s if and only if $d_s(k)$ does not exceed the predetermined threshold Δ. In addition to CURE, we also consider pure memory-one strategies, $(p_{CC}, p_{CD}, p_{DC}, p_{DD}) \in \{0,1\}^4$. Therefore, there are $2^4 = 16$ memory-one strategies and a single CURE strategy in the population.

Repeated public goods games setup.

We can similarly extend the notion of cumulative reciprocity to repeated multiplayer social dilemmas (Fig.5d). To do so, we consider an example of a public good game among $n = 4$ players. In this game, cooperation means to make a contribution of $c > 0$ to a public pool. Total contributions are multiplied by a factor of r with $1 < r < n$. The resulting amount is then equally divided among all group members (even those who did not contribute).

Again, it is straightforward to generalize CURE to this scenario. In this case, Alice's defection difference statistic $d(k)$ records how often Alice has defected up to round k, compared to how often all other group members have defected so far *on average*. In the public goods game studied in Figure 5, pure memory-one strategies are defined as $p_{ij} = (p_{C3}, p_{C2}, p_{C1}, p_{C0}, p_{D3}, p_{D2}, p_{D1}, p_{D0}) \in \{0,1\}^8$. The entries of this vector give the cooperation probability in the next round depending on the number of cooperating co-players *j* and the focal player's own action *i* in the last round. For example, the All-or-None strategy[11] is represented by the vector (1,0,0,0,0,0,0,1). For the simulations, we consider $2^8 = 256$ memory-one strategies and a single implementation of CURE.

Simulation Methods. We explored the evolutionary dynamics of CURE in these scenarios through computer simulations. Similar to the simulation process in the main text, each simulation consists of two steps. First, we obtain the payoffs $\pi(\sigma_i, \sigma_j)$ between two strategies σ_i and σ_j (see Methods). Second, we calculate the strategies' frequencies during the process through the "survival of the fittest" in a noisy environment based on the obtained payoffs between pairs of strategies, following Nowak & Sigmund's approach[21]. However, in the stochastic game and the multiplayer game scenarios, since the cumulative payoff of strategy σ_i may be negative, we project the fitness of strategy σ_i to the interval [0,1] through the normalized equation $f(\sigma_i) = (f(\sigma_i) - f_{min})/(f_{max} - f_{min})$, where f_{max}, f_{min} are the highest and lowest cumulative payoffs of each generation, respectively. We denote the overall fitness of all strategies by $\bar{f} = \sum_{i=1}^{n} x_i f(\sigma_i)$. The frequency of σ_i in the next generation is determined to be $x_i' = x_i * f(\sigma_i)/\bar{f}$. This elementary updating process is repeated for many generations. Mutations are introduced at the beginning in stochastic games and multiplayer games. The mutation rate is set to 10%. When a mutation happens, all other strategies decrease their proportions to 99.9%, while a strategy is randomly selected to increase its proportion by 0.1%.

Results. The results of our simulations are summarized in Fig. 5. For stochastic games, we again observe that CURE enforces fairness (Fig.5a); it readily evolves among memory-one strategies (Fig.5b); and it is particularly successful in those parameter regions in which memory-one strategies fail to sustain cooperation (Fig.5c). We observe similar outcomes for multiplayer games. In particular, CURE is again able to unilaterally enforce a fair outcome, independent of the co-players' strategies (Fig.5d). Moreover, CURE can readily evolve in populations of memory-one players (Fig.5e), and it is particularly strong in those parameter regions in which memory-one strategies alone cannot sustain cooperation (Fig.5f).

These results illustrate how CURE can sustain cooperation in social dilemmas with two actions, cooperation and defection. Future work could explore how this principle can be extended to social dilemmas with continuous action spaces[22,23]. Such games are more complex. In particular, there does not seem to be a consensus on what it means for such games to be a social dilemma. As a first approximation, one could consider games in which there is a unique Nash equilibrium that is not socially optimal. In that case 'cooperation' might be identified with playing the action associated to the social optimum, whereas any other action might be considered a 'defection'. The corresponding version of CURE would again cooperate as long as the defection difference statistic is below the threshold Δ. Once the statistic is above Δ, CURE would switch and play the action associated with the unique Nash equilibrium. Exploring the properties of such an extended version of CURE represents an exciting direction for future work.

Supplementary Section 11. Cumulative reciprocity in an economic experiment

This section supplements the description of "Cumulative reciprocity and human play" in the main text.

General setup. To explore the relevance of cumulative reciprocity on human decision making, we designed a simple behavioral experiment. Participants played the repeated prisoner's dilemma for at least 20 rounds. After that, the game would stop with probability 50% each round, to avoid end-game effects. For the payoffs per round, we use the payoff matrix by Axelrod scaled by a factor of two (values are in British pence)

	C	D
C	15p	0p
D	25p	5p

(10.1)

We consider two different treatments. The first treatment is the *treatment without errors*. Here, the players' decisions are implemented perfectly. The second treatment is the *treatment with errors.* Here, whenever a focal participant chooses to cooperate (defect), there is a 10% chance that the decision is mis-implemented as a defection (cooperation). In that case, the focal participant learns that there was an implementation error; however, the participant's co-player does not. Hence, the co-player cannot tell whether an observed defection was intended or the result of an error.

Experimental methods. To implement the experiment, we created a game software using oTree[24]. To recruit participants, we used the online platform Prolific, and we only allowed UK residents to participate. In total, we recruited 189 participants; 15 out of these participants dropped out during the game's instructions. In addition, one group of 2 participants dropped out during the actual experiment (in the treatment without errors). Overall, we thus report data on 172 participants (43 groups of two players each for each of the two treatments). Each participant received a fixed show-up fee and a variable bonus. The show-up fee was £1. The bonus was the sum of all the payoffs the participant received during the repeated prisoner's dilemma. On average, it took participants 10-20 minutes to complete the experiment, and the average bonus paid was £2.65.

Before entering the actual experiment, participants had to read two pages with instructions. The first page informed participants that their decisions are anonymous, and that they can withdraw from the experiment at any point. The second page contained the rules of the game and three comprehension questions. Participants who failed to answer the comprehension questions were asked again (until they passed all questions). After these introduction pages, participants were randomly assigned to a treatment, and randomly matched with another participant. Each round, before making their decisions, participants were reminded about each other's cumulative past decisions. After interacting in the repeated prisoner's dilemma, participants could provide some feedback on the game software and on how they made their decisions. In addition, we asked some basic demographic questions. Screenshots of the introduction pages and the general game software, are provided in our online repository.

For these behavioral experiments, we have obtained IRB approval by the Ethics Committee of the Medical Faculty of the University of Kiel (D 613/21, October 29, 2021).

Statistical methods. Our statistical analysis reports data on the 172 participants who completed the experiment. To make comparisons across different groups with different game lengths, we only analyze behavior during the first 20 rounds of the game (for which we have data for all groups). All statistical tests reported in the following are two-tailed and non-parametric. For the overall cooperation statistics, we use pairs of players as our statistical unit (i.e., we have $n_1 = 43$ observations in the treatment without errors, and $n_2 = 43$ observations in the treatment with errors). For classifying the players' strategies, we treat each player as a statistical unit (i.e., in that case we have 86 observations in each treatment). We use these tests to explore (*i*) whether the two treatments differ in how cooperative participants are, (*ii*) whether there are any significant changes in the participants' behaviors over time, and (iii) whether any strategy is better able to predict observed behaviors than the other considered strategies.

To explore the participants' strategies, we compare each player's decision in any given round to the decision the player would have made when using some well-known strategies. The set of well-known strategies we consider are CURE (with threshold $\Delta = 3$) and the four memory-one strategies GRIM[1], Tit-for-Tat (TFT[2], Firm-but-fair (FBF)[1] and Win-Stay Lose-Shift (WSLS)[9,10]. The latter four strategies are exactly the pure memory-one strategies that can sustain cooperation in a Nash equilibrium in the game without errors[25]. All these memory-one strategies cooperate in the first round. Thereafter, their behavior is described by the conditional probabilities $\mathbf{p} = (\mathrm{p}_{CC}, \mathrm{p}_{CD}, \mathrm{p}_{DC}, \mathrm{p}_{DD})$ given by

$$\begin{aligned} \mathbf{p}_{\mathrm{GRIM}} &= (1,0,0,0) \\ \mathbf{p}_{\mathrm{TFT}} &= (1,0,1,0) \\ \mathbf{p}_{\mathrm{FBF}} &= (1,0,1,1) \\ \mathbf{p}_{\mathrm{WSLS}} &= (1,0,0,1) \end{aligned} \qquad (10.2)$$

In addition, we also include the strategies AON_2, TFT-ATFT, and CAPRI, which are described in SI Section 8, "Comparison of CURE with selected memory-2 and memory-3 strategies". For implementing these strategies, we used the respective definitions given by respective articles, see Eq. [1] in Hilbe et al[12], Table 1 in Yi et al[13], and Table 3 in Murase and Ki Baek[14]. We note that the respective papers introduce these strategies as memory-2 and memory-3 strategies for infinitely repeated games, without specifying the players' actions in the very first rounds of the game. To have a well-defined strategy for a finitely repeated game, we assume players act in the first few rounds as if they had cooperative decisions in memory for the missing rounds.

For each of these strategies, we compare (*i*) the decision this strategy would predict for the next round, given the players' previous (realized) behavior, and (*ii*) the player's actual decision in the next round. As a result of this comparison, we record how many decisions of a player are correctly predicted by the respective strategy. In particular, we are interested in how often a given strategy is able to correctly predict all twenty decisions of a participant.

Results. The results of the experiment are summarized in Fig. 6. We observe the following patterns:

1. As one may expect, we observe more cooperation in the treatment without errors. When we compare the cooperation rate in the treatment without errors to the *realized actions* in the treatment with errors, the difference is significant (without errors: 73.3% versus with errors: 62.7%; $Z = 2{:}371$, $p = 0{:}018$, Fig. 6a). There is no significant difference if we compare the treatment without errors to the *intended actions* in the treatment with errors (cooperation rate according to intended actions: 66.6%, $Z = 1{:}748$, $p = 0{:}081$).

2. The cooperation rates are slightly decreasing over the course of the experiment (Fig. 6b). In the treatment without errors, this change is significant (cooperation rates during first 10 rounds: 76.7%, last 10 rounds: 69.8%, Wilcoxon matched-pairs test, $Z = 2{:}819$, $p = 0{:}005$). In the

treatment with errors, the change is only significant for intended actions (first 10 rounds: 70.1%, last 10 rounds: 63.1%, $Z = 2{:}124$, $p = 0{:}034$). The change is not significant for realized actions (first 10 rounds: 64.8%, last 10 rounds: 60.5%, $Z = 1{:}383$, $p = 16{:}6\%$).

3. All considered strategies seem to describe the participant's behavior equally well in the treatment without errors (Fig. 6c). Here, each of the eight considered strategies can equally predict the behavior of approximately 40% of the participants (most of the perfectly predicted participants happened to cooperate in all 20 rounds). As a result, we do not find significant differences with respect to the number of correctly predicted participants across the different strategies (Fisher's exact test: $p > 0.85$ for each comparison between pairs of strategies).

   The game with errors makes it easier to distinguish between the eight strategies. Here, we observe that CURE performs best. While CURE perfectly predicts the behavior of 13 subjects, all other strategies only predict the behavior of 0 participants (WSLS, GRIM, AON_2, TFT-ATFT, CAPRI), 1 participant (TFT), or 2 participants (FBF). Based on these numbers, the number of participants correctly predicted by CURE is significantly larger than the number of participants correctly predicted by any of the other strategies (Fisher's exact test, $p < 0.006$ for each pairwise comparison between CURE and any other strategy).

Supplementary References

1. Sigmund, K. *The Calculus of Selfishness*, (Princeton University Press, 2010).
2. Axelrod, R. & Hamilton, W.D. The evolution of cooperation. *Science* **211**, 1390-1396 (1981).
3. Boerlijst, M.C., Nowak, M.A. & Sigmund, K. The logic of contrition. *Journal of Theoretical Biology* **185**, 281-293 (1997).
4. Press, W.H. & Dyson, F.J. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences* **109**, 10409-10413 (2012).
5. Mittal, S. & Deb, K. Optimal strategies of the iterated prisoner's dilemma problem for multiple conflicting objectives. *IEEE Transactions on Evolutionary Computation* **13**, 554-565 (2009).
6. Imhof, L.A., Fudenberg, D. & Nowak, M.A. Tit-for-tat or win-stay, lose-shift? *Journal of theoretical biology* **247**, 574-580 (2007).
7. Brandt, H. & Sigmund, K. The good, the bad and the discriminator—errors in direct and indirect reciprocity. *Journal of Theoretical Biology* **239**, 183-194 (2006).
8. Hofbauer, J. & Sigmund, K. *Evolutionary games and population dynamics*, (Cambridge university press, 1998).
9. Kraines, D. & Kraines, V. Pavlov and the prisoner's dilemma. *Theory and Decision* **26**, 47-79 (1989).
10. Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* **364**, 56-58 (1993).
11. Pinheiro, F.L., Vasconcelos, V.V., Santos, F.C., Pacheco, J.M. & Daunizeau, J. Evolution of All-or-None Strategies in Repeated Public Goods Dilemmas. *Plos Computational Biology* **10**, e1003945 (2014).
12. Hilbe, C., Martinez-Vaquero, L.A., Chatterjee, K. & Nowak, M.A. Memory-n strategies of direct reciprocity. *Proceedings of the National Academy of Sciences* **114**, 4715-4720 (2017).
13. Do Yi, S., Baek, S.K. & Choi, J.-K. Combination with anti-tit-for-tat remedies problems of tit-for-tat.

Journal of Theoretical Biology **412**, 1-7 (2017).
14. Murase, Y. & Baek, S.K. Five rules for friendly rivalry in direct reciprocity. *Scientific Reports* **10**, 1-9 (2020).
15. Nowak, M.A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646-650 (2004).
16. Hilbe, C., Chatterjee, K. & Nowak, M.A. Partners and rivals in direct reciprocity. *Nature Human Behaviour* **2**, 469-477 (2018).
17. Hilbe, C., Šimsa, Š., Chatterjee, K. & Nowak, M.A. Evolution of cooperation in stochastic games. *Nature* **559**, 246-249 (2018).
18. Su, Q., Mcavoy, A., Wang, L. & Nowak, M.A. Evolutionary dynamics with stochastic game transitions. *Proceedings of the National Academy of Sciences* **116**, 201908936 (2019).
19. Tilman, A.R., Plotkin, J.B. & Akay, E. Evolutionary games with environmental feedbacks. *Nature Communications* **11**(2020).
20. Shapley & L., S. Stochastic Games. *Proceedings of the National Academy of Sciences* **39**, 1095-1100 (1953).
21. Nowak, M.A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355**, 250-253 (1992).
22. Killingback, T. & Doebeli, M. The continuous prisoner's dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *The American Naturalist* **160**, 421-438 (2002).
23. McAvoy, A. & Hauert, C. Autocratic strategies for iterated games with arbitrary action spaces. *Proceedings of the National Academy of Sciences* **113**, 3573-3578 (2016).
24. Chen, D.L., Schonger, M. & Wickens, C. oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments. *Journal of Behavioral and Experimental Finance*, 88-97 (2016).
25. Akin, E. The iterated prisoner's dilemma: good strategies and their dynamics. *Ergodic Theory, Advances in Dynamical Systems*, 77-107 (2016).

Acknowledgments

It takes 150 meaningful relationships to make a doctor

Inspired by Dunbar and an old wisdom.

I would first like to express my immense gratitude to my supervisors, Christian Hilbe and Oliver Hauser. For teaching me your respective crafts, helping perfect old skills and learn new techniques. Christian, thank you so much for taking the time to show me the ropes of the trade, for your patience when nothing was working, and for your insightful comments on my results and work. Oliver, thank you for being my experimental specialist from afar. Your mentorship and advice have carried me all the way here. I couldn't have asked for a better advisory team.

I am also grateful for the collaborators and mentors who have followed my journey from up close and afar. I would like to thank Arne for creating such a great working atmosphere in his host department and for reassuring me that my work should definitely be a chapter! I wouldn't be thinking about science and human behaviour the way I am now if it wasn't for Moshe. Thank you for making me look at social behaviour from a different angle and for encouraging me to seek external collaborations. And to Bethany for her valuable wisdom and experience in running experiments, and being a fellow psychologist in this sea of mathematicians.

I must thank my fellow dynos for creating a group with such a positive and supportive atmosphere. First my doctoral fellows, Marta and Saptarshi who have been with me every step of the way: starting all together made this experience that much more special. Nik for the mood-lifting office banter and all the python support – I shall always remember the difference between a method and a function now! Thanks also to new and old members: Ben, Julien, Logan, Vivien, Xiaomin, and Yosuke. The conversations on such a range of topics of expertise has been invaluable.

Now, a special mention to my PR family. Amanda, Alejandra, Alvaro, Christin, Dana, Dharanish, Emil, Ernesto, Félix, Florence, Gaurav, Gosia, Guilhem, Ian, Jule[‡], Michael R, Nikhil, Roman, Saumil, Sedigheh, and Xiaowen. The theory lunches, lockdown Fridays, whiskey nights (sorry again about that), BBQs on the beach, swimming sessions, and dance parties made Plön a special place. Thank you also for asking me the weirdest questions and forcing me to think deeper and better about my projects. A family would not be complete without brothers who tease you relentlessly but are always there to make you laugh or give you a shoulder to cry on: The true Indian boys, Nikhil, Saumil and Saptarshi.

Over the bungalows and far away lays the main building where “real” biologists reside. Alec, Alejandro, Aleksa, Alexandros, Amor, Beatriz, Charly, Demetris, Elio, Enea, Felipe, Gisela, Jun, Karem, Lavisha, Louk, Pauline, Quentin‡, and Stella. The institute is alive thanks to you.

Beyond my colleagues, collaborators and comrades-in-arms, I would like to thank all those who have helped me, directly or indirectly, by giving their time and effort or by spreading enthusiasm for social behaviour. I am thinking in particular of Angela, who goes above and beyond for her students, Ursula for helping us navigate the German bureaucracy, my TAC member Daniel for offering advice when needed, and Hinrich, who kindly agreed to review this thesis. And of course, to all those I have met along the way of my Doctoral career and who shared the passion for human behaviour, including Alejandro, Amine, Andrea, Augusto, Diogo, Hirotaka, Julio, Luuk, Maija, Valentin, Viviana, Yuan, Maria, Jorge, Jens, Hanna, Guiliana and Angelo.

As one of the few who chose to live in the big city, I have to thank the Kielers who gave me a life outside of the institute. Aleksa, Antonia, Goda, Jule, and Meghna.

Je ne peux pas omettre ma vraie famille qui m’a soutenue et encouragée de loin. Merci Didier d’avoir implanté l’idée d’un doctorat si profondément que je n’ai jamais envisagé de faire autre chose, mais surtout merci à mes parents, Enrico et Tina, pour leur amour et fierté qui m’ont donné le courage et l’assurance d’embarquer dans ce voyage scientifique à l’étranger.

I would like to finish by (re)mentioning individuals who have particularly marked my time as a Doctoral student for their friendship and love, or at the very least for having my toothbrush at their house for when work, DB or social life had me trapped in Plön. Florence, thank you for being my french speaking person and co-instigator of fun times. Saptarshi, thank you for going on adventures with me and for the lively discussions. Saumil, thank you for never judging my mistakes and rejoicing with me of life’s small victories. Louk, thank you for letting me invade your house, for feeding me when procrastination had the better of me and for hugging me always, even when I don’t need it. These reciprocal relationships will stay with me forever.

‡Honorary members and theory wannabes

Biography

Charlotte Scarlet Lucy Rossetti was born on the 30th of November 1994 in Geneva, Switzerland. She left her home country in 2014 to study psychology at the University of Nottingham in the United Kingdom. After receiver her Bsc in Psychology, her interests for strategic decision-making led her to pursue a Msc in Behavioural and Economics Science at the University of Warwick. There, her Master thesis on communication in the centipede game received the prize of best project in her cohort and she graduated with distinction in 2019. To expand her skills and horizons, she took up a field assistant position at the Inkawu Vervet Project in South Africa where she worked on vervet monkeys' social and strategic behaviour. Armed with a background in psychology, economics and animal behaviour, she started as a PhD candidate under Dr. Christian Hilbe and Prof. Oliver Hauser in 2020 researching human social behaviour both experimentally and theoretically. The results of this work is described in this thesis.