Evolution of cooperation via (in)direct reciprocity under imperfect information

by

Laura Schmid

October, 2021

A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:

Chris Wojtan, Chair
Krishnendu Chatterjee
Dan Alistarh
Ulrich Berger

Institute of Science and Technology

The thesis of Laura Schmid, titled *Evolution of cooperation via (in)direct reciprocity under imperfect information*, is approved by:

Supervisor: Krishnendu Chatterjee, IST Austria, Klosterneuburg, Austria

Signature: ____________________

Committee Member: Dan Alistarh, IST Austria, Klosterneuburg, Austria

Signature: ____________________

Committee Member: Ulrich Berger, Vienna University of Economics and Business, Vienna, Austria

Signature: ____________________

Defense Chair: Chris Wojtan, IST Austria, Klosterneuburg, Austria

Signature: ____________________

Signed page is on file

IST Austria Thesis, ISSN: 2663-337X

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: ______________________________

Laura Schmid
October, 2021

Signed page is on file

Abstract

Indirect reciprocity in evolutionary game theory is a prominent mechanism for explaining the evolution of cooperation among unrelated individuals. In contrast to direct reciprocity, which is based on individuals meeting repeatedly, and conditionally cooperating by using their own experiences, indirect reciprocity is based on individuals' reputations. If a player helps another, this increases the helper's public standing, benefitting them in the future. This lets cooperation in the population emerge without individuals having to meet more than once.

While the two modes of reciprocity are intertwined, they are difficult to compare. Thus, they are usually studied in isolation. Direct reciprocity can maintain cooperation with simple strategies, and is robust against noise even when players do not remember more than their partner's last action. Meanwhile, indirect reciprocity requires its successful strategies, or social norms, to be more complex. Exhaustive search previously identified eight such norms, called the "leading eight", which excel at maintaining cooperation.

However, as the first result of this thesis, we show that the leading eight break down once we remove the fundamental assumption that information is synchronized and public, such that everyone agrees on reputations. Once we consider a more realistic scenario of imperfect information, where reputations are private, and individuals occasionally misinterpret or miss observations, the leading eight do not promote cooperation anymore. Instead, minor initial disagreements can proliferate, fragmenting populations into subgroups.

In a next step, we consider ways to mitigate this issue. We first explore whether introducing "generosity" can stabilize cooperation when players use the leading eight strategies in noisy environments. This approach of modifying strategies to include probabilistic elements for coping with errors is known to work well in direct reciprocity. However, as we show here, it fails for the more complex norms of indirect reciprocity. Imperfect information still prevents cooperation from evolving.

On the other hand, we succeeded to show in this thesis that modifying the leading eight to use "quantitative assessment", i.e. tracking reputation scores on a scale beyond good and bad, and making overall judgments of others based on a threshold, is highly successful, even when noise increases in the environment. Cooperation can flourish when reputations are more nuanced, and players have a broader understanding what it means to be "good."

Finally, we present a single theoretical framework that unites the two modes of reciprocity despite their differences. Within this framework, we identify a novel simple and successful strategy for indirect reciprocity, which can cope with noisy environments and has an analogue in direct reciprocity. We can also analyze decision making when different sources of information are available. Our results help highlight that for sustaining cooperation, already the most simple rules of reciprocity can be sufficient.

Acknowledgements

First of all, my thanks go out to my advisor Krish for his continuous support and guidance throughout my PhD, and for the many opportunities he has given me to grow as a researcher. Krish always did his best to help me, mentor me, and watch out for me in all matters during my doctorate. I am very happy to have been able to explore many different projects and ideas in this friendly environment, and to challenge myself. I am also grateful to the other members of my thesis committee, Dan and Ulrich, for all their help and advice.

I furthermore want to thank my contributors, starting with Christian, who was an integral part of my PhD right from the start. His mentorship and our many collaborations and discussions have been a great source of support, and I have learned a lot from him. My appreciation also goes out to Martin for all the efforts he invested into our joint work. Additionally, I would like to thank to everyone else that I had the pleasure to work on projects with over the past years, such as Stefan, Juho, Pepa, and Maria, for their valuable inputs and help.

Finally, my thanks go out to my family and friends for their constant support.

The research for this thesis has partially been supported via the European Research Council Start Grant 279307 Graph Games, the European Research Council CoG 863818 (ForM-SMArt), as well as the Austrian Science Fund (FWF) Grant P23499-N23, Austrian Science Fund (FWF) under grant Z211-N23 (Wittgenstein Award), and FWF Nationale Forschungsnetzwerke Grant S11407-N23 Rigorous Systems Engineering/Systematic Methods in Systems Engineering.

About the Author

Laura Schmid completed a BSc in Engineering Physics at TU Vienna and forewent her MSc thesis and graduation in the same subject in favor of joining IST Austria in September 2016. Her main interests relate to evolutionary game theory with a focus on mechanisms for the evolution of cooperation. In her central research project, Laura studied (in)direct reciprocity under noisy, private, and incomplete information, and published works in high-impact journals such as Nature Human Behaviour and PNAS. In the course of her PhD, she additionally wrote a paper on the surprising role memory can play in direct reciprocity, pursued projects on the interface of game theory and distributed systems, and contributed to a book chapter on modeling sociological phenomena using evolutionary game theory. She gave various talks on her results at venues such as Harvard University, MIT Media Lab, the Conference on Principles of Distributed Systems (OPODIS), University of Vienna, and the University of Pennsylvania.

Aside from doing science, Laura also studied piano at the Music and Arts University of the City of Vienna (formerly Konservatorium Wien Privatuniversität) in her teens, and worked at the Vienna State Opera as an assistant stage manager for subtitling for eight years. Her hobbies include music, learning languages, skeet shooting, and boxing.

List of Collaborators and Publications

This thesis is based on the following publications:

- C. Hilbe, L. Schmid, J. Tkadlec, K. Chatterjee, and M. A. Nowak. Indirect reciprocity with private, noisy, and incomplete information. *Proceedings of the National Academy of Sciences USA*, 115:12241–12246, 2018
- Laura Schmid, Pouya Shati, Christian Hilbe, and Krishnendu Chatterjee. Evolution of indirect reciprocity under assessment and action generosity. *Scientific Reports*, 11, 2021
- L. Schmid, K. Chatterjee, C. Hilbe, and M.A. Nowak. A unified framework of direct and indirect reciprocity. *Nature Human Behaviour*, 2021

Chapter 4 is based on joint work (as of yet unpublished) with Christian Hilbe (Max Planck Research Group Dynamics of Social Behavior, Max Planck Institute for Evolutionary Biology, Plön, Germany) Farbod Ekbatani (London School of Economics, United Kingdom) and Krishnendu Chatterjee (IST Austria).

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Game theory, originally introduced by von Neumann and Morgenstern [vNM53], aims to mathematically analyze strategic decision making in competitive situations and interactions with multiple participants (players) who can influence each other's outcomes [FT98, NRTV07]. The classic framework usually makes some key assumptions about the players' rationality, their beliefs and cognitive abilities when reasoning about strategies that maximize players' benefits (payoffs) in a given situation, building on fundamental concepts such as Nash equilibria. In this general form, game theory has become a very broad field with multiple applications, starting with its foundations in economics and extending to a wide array of problems in computer science as well as political science and biology. Game-theoretic approaches are used to describe and solve a diverse range of problems, such as resource allocation, public goods provision, virus inoculation, voting, network routing, bargaining, auction bidding, and market competition.

Applications of game theory in (evolutionary) biology gave rise to the field of *evolutionary game theory*, starting with the work of John Maynard-Smith [MSP73]. As a generic approach to evolutionary dynamics [MS82, HS98], evolutionary game theory now is not only a central area of mathematical biology, but has also developed to have many applications even beyond life sciences, such as describing the evolution of language, or sociological phenomena. In contrast to classic game theory, it considers populations of players with bounded rationality who interact with other population members in a game. Individuals' behavior is governed by their strategies, which tell them how to act in a given situation. The players' payoffs from their interactions – which depend on the actions of the co-players and therefore on the abundance of different strategies – are assumed to be a proxy for their *evolutionary fitness*. Success in the underlying game is thus translated into increased evolutionary fitness, such that good strategies spread faster, whereas disadvantageous strategies rather go extinct.

In the following, we first embed the work presented in this thesis into the state of the art in evolutionary game theory, and then give an overview of our contributions.

1.1 State of the field

One prominent question raised in evolutionary game theory is how cooperation among unrelated individuals can be sustained. The emergence of cooperative behavior might seem puzzling from an evolutionary standpoint: after all, helping others at own cost can easily lead to a personal disadvantage over time, even though altruism is beneficial for the population as a whole. Selfishness seems like a better option in an evolutionary competition, since natural selection is competitive, and altruistic behavior often is exploited once free-riders, i.e. defectors, start appearing in a group. This conflict between personal advantage and group advantage is usually modeled in the form of social dilemmas like the Prisoner's Dilemma [AH81] or a public goods game [SP16]. Players in these games prefer to defect against their partners, while mutual cooperation would lead to higher overall benefits. The literature on evolutionary game theory has highlighted various mechanisms to describe how cooperation can however still evolve and be maintained. One of the most prominent such mechanisms is reciprocity, which means that players conditionally cooperate with each other, and preferentially help those who have been helpful in the past. Most humans are conditional cooperators, and many daily life interactions are based on reciprocity, be it repaying a favor or building up trust with someone over time [Tri71, Sug86, Now06, Sig10].

Evolutionary game theory distinguishes between two modes of reciprocity: direct and indirect reciprocity. Direct reciprocity [AH81, NS92, HS97, PD12, HNS13, SP14a, SP14c, Aki16, PHRZ15, HRZ15, MH16b, IM18, HCN18, GvV18, RHR+18] is based on repeated interactions between players. Individuals use their own experience to decide whether to cooperate or defect with an opponent. This requires that the same individuals interact repeatedly, which enables them to respond to their interaction partner in future transactions. This is modeled in the framework of repeated games, for example using the iterated Prisoner's Dilemma. While each player still prefers to defect if they only interact once, cooperation becomes feasible when they interact over multiple rounds [MS06, HvCN18, vVGRN12]. In such repeated games, players can take into account the game's previous history when deciding what to do next. Strategies are then rules that tell the player whether to cooperate or not given the history of the game. In this context, the concept of memory comes into play. Individuals are constrained in how many previous interactions they remember, governed by their limited memory capacity. A very commonly considered class of strategies are so-called memory-1 strategies, which have players take into account only their own and their co-player's last move. A special case of memory-1 strategies is the even simpler class of reactive strategies, where players only remember the co-player's last move [NS90]. Reactive strategies such as Tit-for-Tat [AH81] only require minimal information, yet they are very successful in maintaining cooperation. Even when players occasionally make mistakes or environments are noisy, cooperation can evolve with simple strategies of direct reciprocity, e.g. with the help of Generous Tit-for-Tat [NS92, Mol85]. Similarly successful strategies using a larger memory capacity have also been identified [HS97, Lin97, HMVCN17].

Indirect reciprocity [NS98b, LH01, OI04, SSP18, Sig12, Kan92], on the other hand, lets players also consider others' experiences and cooperate with those opponents that have a good reputation. By helping somebody, people increase their own public standing, which benefits them in future interactions if reputation is a sufficiently valuable commodity. In contrast to direct reciprocity, indirect reciprocity does not require any two individuals to interact with one another more than once. Population members instead continually assess others' actions and then act based on the resulting reputations. Indirect reciprocity thus transfers the principles of direct reciprocity to the group level, and builds upon a group's collective memories. Since individuals need to develop a sense of which behaviors should be regarded as good or bad, indirect reciprocity has become the subfield of evolutionary

game theory that explores the evolution of moral behavior [Ale87]. The dynamics of indirect reciprocity are again modeled with simple social dilemma situations, where a donor is asked whether or not to pay a personal cost to provide help to a recipient, i.e. whether to cooperate or defect. The donor's action is observed by other population members, who then decide how the donor's reputation should be updated in light of the observed behavior [BS05]. How exactly actions are assessed, and how these assessments are translated into future actions, is governed by *social norms* in the community. These social norms thus consist of two components: an assessment rule that specifies how a player should judge an observed action, and an action rule that tells the donor whether or not to help the recipient, potentially depending on some context like the recipient's current reputation. Many previous studies have dealt with the question of which norms excel most at facilitating and sustaining cooperation, and how complex they need to be. While there is important work that shows that first-order information, e.g. remembering a co-player's last two actions, can already suffice to sustain cooperation via indirect reciprocity [Ber11, BG16], most of the well-known social norms studied in the literature are very complex. Influential previous work [OI04, OI06] identified eight highly successful such norms of higher complexity: the "leading eight". These norms' complexity lies in the fact that players take third-order information into account when making assessments; this means that they contextualize an observed action with the reputation of the donor and the reputation of the receiver. The eight highlighted norms can then differ, for example, in how they judge the cooperation of a good player with a bad player.
Yet, previous research on the leading eight norms is based on rather stringent assumptions on the players' environment. The success of these norms so far had been demonstrated in settings where information is public, all players share the same synchronized opinion about everyone else, and where all players observe every single interaction without misunderstandings. This is a very idealized scenario, which is rarely mirrored in real life. The role of *imperfect information* in indirect reciprocity was less clear. How do the leading eight in contrast fare in more realistic settings, where all these assumptions are removed and players may have differing opinions about others spurred by occasional misunderstandings and missed observations? How easily can indirect reciprocity thus sustain cooperation in noisy environments with private and incomplete information? This still remained a natural, but non-trivial question. The first part of this thesis addresses this issue, and demonstrates the shortcomings of successful social norms and subsequent breakdown of cooperation in noisy environments.

While the two modes of reciprocity are conceptually intertwined, the models previously used to study them are strikingly different. The gap between those models seems wide, in particular when it comes to strategic complexity. After all, in most studies, successful strategies of direct reciprocity are very simple, while still being able to cope with errors and noise. When players use the well-known reactive strategy Generous Tit-for-Tat, they just need to remember one bit of information to be able to sustain cooperation, even when players make occasional mistakes. Describing the resulting dynamics is then well feasible. In contrast, successful strategies of indirect reciprocity usually take the form of elaborate social norms like the leading eight. Deciding how to evaluate the reputation of a given co-player or how to act towards a given co-player can depend on a significant amount of context that players need to keep track of. In most models, these decisions depend on the co-player's previous actions, but at the same time on the reputation of the co-player's previous co-players (which in turn depends on those players' last co-players). This significantly complicates the analysis of indirect reciprocity, and makes simpler

strategies for indirect reciprocity desirable - in particular simpler strategies that can additionally cope with the effects of imperfect information.
As a result of the differences between models of direct and indirect reciprocity, the two modes are usually studied in isolation. This now begs the question if there is a possibility to unite the two kinds of reciprocity despite their fundamental contrasts. How should we, for example, answer the question of whether to help someone who treated us well in the past, but has a bad public reputation? Can one find a unified model of reciprocity that assumes an imperfect information setting right from the start? And is there a possibility then that successful strategies in indirect reciprocity can take forms more similar to the simple strategies of direct reciprocity? This is what the second part of this thesis answers.

1.2 Our contributions

Much of the previous work on the leading eight strategies of indirect reciprocity assumes that a central authority tracks players' behavior and in turn assigns updated reputations after every interaction in the population. In Chapter 2, we first show the fundamental result that once we remove this assumption of an idealized setting featuring synchronized and fully public information, the leading eight strategies of indirect reciprocity do not promote cooperation anymore.
For this, we consider a scenario where individuals privately keep track of others' reputations instead of perfectly agreeing on everyone's standing. Every player is equipped with a reputation repository representing a collection of the reputations he assigns to each member of the population including himself. We also assume a noisy environment: players' observations are subject to perception errors, i.e. misunderstandings. Furthermore, information is incomplete: not every interaction is observed by everyone, such that not every player updates their reputation repository after every single interaction.

In this setting, we show that minor initial disagreements about others' reputations can proliferate, eventually fully fragmenting communities into different subgroups that assign a good reputation to every player inside the group, and a bad reputation to every outsider. This creates a deadlock. Thus, in the presence of noise caused by perception errors, most leading eight strategies are not stable anymore. Even if a leading eight norm evolves, which only happens for three of the eight norms, the emergence of cooperative behavior is effectively prevented. These modestly successful three norms then lead to positive cooperation rates in the population, but in comparison to the public information scenario, these rates are drastically reduced. We additionally prove the analytical result that while the more robust leading eight strategies can recover from single intial disagreements in a noisy environment, it may take them a long time to do so, which implies that they cannot maintain full cooperation. This study sheds light on the interplay of moral norms, reputation formation, and the evolution of cooperation aside from idealized settings.

This finding naturally calls for a way to mitigate the strong negative impact of noisy environments on indirect reciprocity. In this context, we considered two approaches to this problem.

In Chapter 3, we first demonstrate a negative result. Here, we introduce elements of "generosity" into the leading eight. This implies that contrary to the baseline model in Chapter 2, we assume that players' strategies are not completely deterministic. Instead individuals now are equipped with probabilities to cooperate when their norm would usually prescribe to defect ("action generosity"), and/or to assign a good reputation to a

player they would usually assign a bad reputation to ("assessment generosity"). We hence distinguish between two kinds of generosity, and explore the effect that each of them has on the capacity of the leading eight norms to sustain cooperation. This approach is inspired by the related literature on direct reciprocity, where the role of generosity has traditionally been highlighted. There, the strategy Generous Tit-for-Tat has been shown to be successful when the classical Tit-for-Tat strategy is not. By occasionally cooperating against defectors, Generous Tit-for-Tat is able to reestablish cooperation even in the presence of noise. This makes generosity a powerful mechanism in the realm of repeated games to cope with errors that would usually prove detrimental.
However, this does not hold for indirect reciprocity. In fact, we show that assessment generosity acts as yet another seed of disagreement, as it leads to private reassessment of others' reputations. This further weakens the performance of the leading eight, letting cooperation rates decrease further without exception. On the other hand, action generosity only leads to a public show of cooperation, which does not have the same immediate negative consequences, as it does not lead to the same mutual disagreements as assessment generosity. However, while this mode of generosity can sometimes slightly enhance cooperation, the leading eight norms still fail to evolve. These results further highlight the importance of coordination and communication in societies with complex social norms. In such societies, individual acts of generosity can be detrimental as they naturally introduce disagreements between different population members.

On the other hand, in Chapter 4, we present a separate, successful method of making the leading eight norms more robust against the effects of noise and private information. This method requires us to drop a different standard assumption made about the leading eight, namely, taking players' reputations to be binary. In the baseline model of the leading eight that is commonly studied in the literature, individuals only distinguish between the labels "good" and "bad". In contrast, in this chapter, we study the case where reputations are more nuanced, i.e. on a scale between a minimal and a maximal value. Players privately keep track of each others' integer reputation scores instead of binary labels, adding to and subtracting from these scores depending on their norms and the observations they make. For an individual to then make an overall judgment of a peer in order to get the context for assessment or decision whether to cooperate or not, they compare this co-player's score with a threshold. A larger range of values for the reputation scores therefore results in a greater insensitivity of overall judgment to single assessments, i.e. greater tolerance for single bad events. This more refined assessment system hence can act as a buffer for misjudgments. The approach in this chapter draws inspiration from previous work on the simple Image Scoring norm of indirect reciprocity, which has been shown to be less vulnerable to errors when scores are not binary. Our results show that such a model of "quantitative assessment" can indeed work for the leading eight norms, as it corrects and to a certain degree prevents disagreements between players using the same norm. This makes it more likely for four of the leading eight norms to evolve, giving high cooperation rates for these four norms even in environments with significant noise levels. Our findings let us eventually draw the conclusion that for a leading eight norm to be stable both under public and private information, it cannot be "gullible" by making it too easy for bad players to gain a good reputation.
We thus show the power of more nuanced opinions when a community uses complex social norms. Too rash judgments can break down cooperation in noisy environments, whereas a certain amount of tolerance can help sustain it.

Finally, in Chapter 5, we then considered the question of how to unify direct and indirect reciprocity despite their stark differences and the quite contrasting models for the two modes. Here, we introduced a single theoretical framework that works for both, assuming a setting where information can be private, noisy and imperfect. We assume that individuals can choose between direct and indirect reciprocity: they can consider only actions directed at them or also interactions between third parties. Our framework then contains many previous models of direct and indirect reciprocity as special cases. We describe the interaction between the two modes with the help of a central equation that lets us explicitly calculate payoffs, and derive a number of results starting from this point. For one, we analyze how to make decisions when different sources of information are available. Here, we find that winning strategies take third-party interactions into account only when the information is sufficiently reliable. Otherwise they ignore indirect information and trust only their personal experience. Generally, the two modes need different environments to emerge, both in terms of the length of the game as well as in terms of how noisy and prone to mutations the evolutionary process itself is.

Second, we find a novel simple and powerful strategy of indirect reciprocity, which we call "Generous Scoring", and which has an immediate analogue (Generous Tit-for-Tat) in direct reciprocity. Generous Scoring only requires a minimum amount of information, namely only what the co-player did last. A player with this strategy then always assigns a good reputation to those who have previously cooperated, but even someone who has been observed to have defected can be forgiven with some probability. By extending the mathematical theory of direct reciprocity to indirect reciprocity, we prove important theoretical properties of Generous Scoring: it is a Nash equilibrium with respect to arbitrary mutants, just like its direct reciprocity analogue Generous Tit-for-Tat. Hence, if everyone adopts one of these strategies, no other norm can take over even if it is much more sophisticated. More generally, we hereby show that direct and indirect reciprocity can be explored with the same equations, give rise to similar strategies, and naturally complement each other in sustaining cooperation. Our results suggest that already the most simple rules of reciprocity can be sufficient for the evolution of cooperation. Our framework however not only links direct and indirect reciprocity by what they have in common, but also allows us to understand their differences in more depth.

CHAPTER 2

Indirect reciprocity with private, noisy, and incomplete information

Indirect reciprocity is a mechanism for cooperation based on shared moral systems and individual reputations. It assumes that members of a community routinely observe and assess each other, and that they use this information to decide who is good or bad, and who deserves cooperation. When information is transmitted publicly, such that all community members agree on each other's reputation, previous research has highlighted eight crucial moral systems. These 'leading-eight' strategies can maintain cooperation and resist invasion by defectors. However, in real populations individuals often hold their own private views of others. Once two individuals disagree about their opinion of some third party, they may also see its subsequent actions in a different light. Their opinions may further diverge over time. Herein, we explore indirect reciprocity when information transmission is private and noisy. We find that in the presence of perception errors, most leading-eight strategies cease to be stable. Even if a leading-eight strategy evolves, cooperation rates may drop considerably when errors are common. Our research highlights the role of reliable information and synchronized reputations to maintain stable moral systems.

2.1 Introduction

Humans treat their reputations as a form of social capital [Ale87, RN12, MS10]. They strategically invest into their good reputation when their benevolent actions are widely observed [EF09, JHTM11, OIN15], which in turn makes them more likely to receive benefits in subsequent interactions [WM00, MSBK01, SS06, BKO05, vAS16, SMUE16, CGVP16]. Reputations undergo constant changes in time. They are affected by rumors and gossip [SKSM07], which themselves can spread in a population and develop a life of their own. Evolutionary game theory explores how good reputations are acquired, and how they affect subsequent behaviors, using the framework of indirect reciprocity [NS05, Now06, Sig10, Sig12]. This framework assumes that members of a population routinely observe and assess each other's social interactions. Whether a given action is perceived as good depends on the action itself, the context, and on the social norm employed by the population. Behaviors that yield a good reputation in one society may be condemned in others. A crucial question thus becomes which social norms are most conducive to maintain cooperation in a population.

Different social norms can be ordered according to their complexity [SSP18], and according to the information that is required to assess a given action [BS04, OI04]. According to 'first-order norms', the interpretation of an action only depends on the action itself. When a donor interacts with a recipient in a social dilemma, the donor's reputation improves if she cooperates, whereas her reputation drops if she defects [NS98b, NS98a, Oht04, BS05, Ber11, BG16]. According to 'second-order norms', the interpretation of an action additionally depends on the reputation of the recipient. The recipient's reputation provides the context of the interaction. It allows observers to distinguish between justified and unjustified defections [LH01, PB03, SA07]. For example, the standing strategy only considers it as wrongful to defect against well-reputed recipients; donors who defect against bad recipients do not suffer from an impaired reputation [Sug86]. According to 'third-order norms', observers need to additionally take the donor's reputation into account. In this way, assessment rules of higher order are increasingly able to give a more nuanced interpretation of a donor's action, but they also require observers to store and process more information.

When subjects are restricted to binary norms, such that reputations are either 'good' or 'bad', an exhaustive search shows there are eight third-order norms that maintain cooperation [OI04, OI06]. These 'leading-eight strategies' are summarized in **Tab. 1**, and we refer to them as L1–L8. None of them is exclusively based on first-order information, whereas two of them (called 'Simple Standing' and 'Stern Judging', [CSP06, SSP16]) require second-order information only. There are several universal characteristics that all leading-eight strategies share. For example, against a recipient with a good reputation, a donor who cooperates should always obtain a good reputation, whereas a donor who defects should gain a bad reputation. The norms differ, however, in how they assess actions towards bad recipients. Whereas some norms allow good donors to preserve their good standing when they cooperate with a bad recipient, other norms disincentivise such behaviors.

Ohtsuki and Iwasa have shown that if all members of a population adopt a leading-eight strategy, stable cooperation can emerge [OI04, OI06]. Their model, however, assumes that the players' images are synchronized; two population members would always agree on the current reputation of some third population member. The assumption of publicly available and synchronized information greatly facilitates a rigorous analysis of the reputation dynamics. Yet in most real populations, different individuals may have access to different kinds of information, and thus they might disagree on how they assess others. Their opinions may well be correlated, but they will not be correlated *perfectly.* Once individuals disagree in their initial evaluation of some person, their views may further diverge over time. How such initial disagreements spread may itself depend on the social norm employed by the population. While some norms can maintain cooperation even in the presence of rare disagreements, other norms are more susceptible to deviations from the public information assumption [Uch10, US13, OSN17, OSN18]. Here, we explore systematically how the leading-eight strategies fare when information is private, noisy, and incomplete. We show that under these conditions, most leading-eight strategies cease to be stable. Even if a leading-eight strategy evolves, the resulting cooperation rate may be drastically reduced.

Assessment rule	L1 *	L2 Consistent Standing	L3 Simple Standing	L4 *	L5 *	L6 Stern Judging	L7 Staying	L8 Judging
Good cooperates with Good	g	g	g	g	g	g	g	g
Good cooperates with Bad	g	b	g	g	b	b	g	b
Bad cooperates with Good	g	g	g	g	g	g	g	g
Bad cooperates with Bad	g	g	g	b	g	b	b	b
Good defects against Good	b	b	b	b	b	b	b	b
Good defects against Bad	g	g	g	g	g	g	g	g
Bad defects against Good	b	b	b	b	b	b	b	b
Bad defects against Bad	b	b	g	g	g	g	b	b

Action rule	L1	L2	L3	L4	L5	L6	L7	L8
Good meets Good	C	C	C	C	C	C	C	C
Good meets Bad	D	D	D	D	D	D	D	D
Bad meets Good	C	C	C	C	C	C	C	C
Bad meets Bad	C	C	D	D	D	D	D	D

Table 2.1: **The leading-eight strategies of indirect reciprocity.** There are eight strategies, called the 'leading-eight', that have been shown to maintain cooperation under public assessment [OI04, OI06]. Each such strategy consists of an assessment rule and of an action rule. The assessment rule determines whether a donor is deemed good (g) or bad (b). This assessment depends on the context of the interaction (on the reputations of the donor and the recipient), and on the donor's action (C or D). The action rule determines whether to cooperate with a given recipient when in the role of the donor. A donor's action may depend on her own reputation, as well as on the reputation of the recipient. All of the leading-eight strategies agree that cooperation against a good player should be deemed as good, whereas defection against a good player should be deemed bad. They disagree in how they evaluate actions towards bad recipients.

2.2 Results

2.2.1 Model setup

We consider a well-mixed population of size N. The members of this population are engaged in a series of cooperative interactions. In each round, two individuals are randomly drawn, a donor and a recipient. The donor can then decide whether to transfer a benefit b to the recipient at own cost c, with $0<c<b$. We refer to the donor's two possible actions as cooperation (transferring the benefit) and defection (not doing anything). Whereas the donor and the recipient always learn the donor's decision, each other population member independently learns the donor's decision with probability $q>0$. Observations may be subject to noise: we assume that all players who learn the donor's action may misperceive it with probability $\epsilon > 0$, independently of the other players. In that case, a player misinterprets the donor's cooperation as defection, or conversely, the donor's defection as cooperation. After observing an interaction, population members independently update their image of the donor according to the information they have (**Fig. 2.1**).

To do so, we assume that each individual is equipped with a strategy that consists of

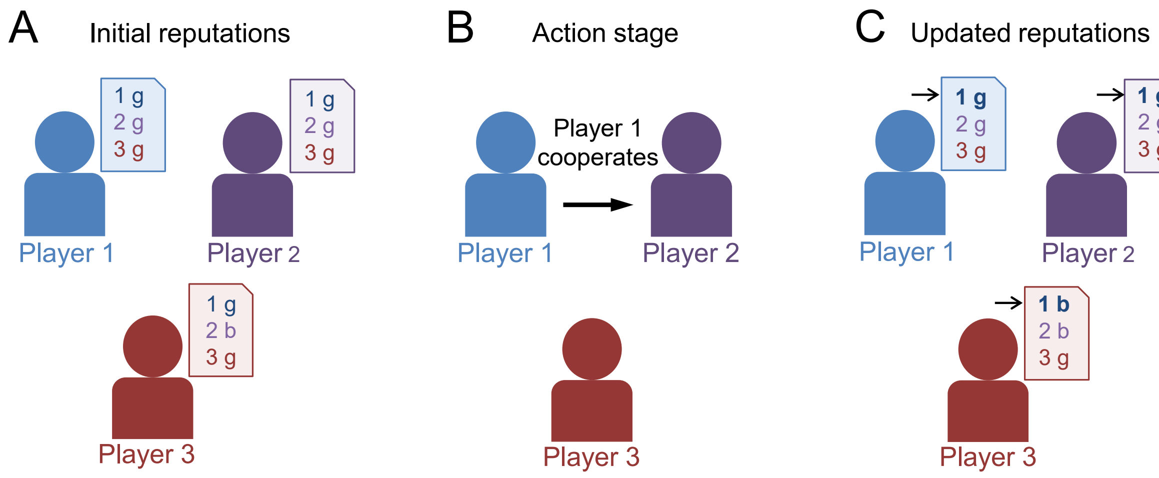

Figure 2.1: **Under indirect reciprocity, individual actions are continually assessed by all population members.** (A) We consider a population of different players. All players hold a private repository where they store which of their co-players they deem as either good (g) or bad (b). Different players may hold different views on the same co-player. In this example, player 2 is considered to be good from the perspective of the first two players, but he is considered to be bad by player 3. (B) In the action stage, two players are randomly chosen, a donor (here, player 1) and a recipient (here, player 2). The donor can then decide whether or not to cooperate with the recipient. The donor's decision may depend on the stored reputations in her own private repository. (C) After the action stage, all players who observe the interaction update the donor's reputation. The newly assigned reputation may differ across the population even if all players apply the same social norm. This can occur (*i*) when individuals already disagreed on their initial assessments of the involved players, (*ii*) when some subjects do not observe the interaction and hence do not update the donor's reputation accordingly, or (*iii*) when there are perception errors.

an assessment rule and of an action rule. The player's assessment rule governs how players update the reputation they assign to the the donor. Here we consider third-order assessment rules. That is, when updating the donor's reputation, a player takes the donor's action into account, as well as the donor's and the recipient's previous reputation. Importantly, when two observers differ in their initial assessment of a given donor, they may also disagree on the donor's updated reputation, even if both apply the same assessment rule and observe the same interaction (**Fig. 2.1c**). The second component of a player's strategy, the action rule, determines which action to take when chosen to be the donor. This action may depend on the player's own reputation, as well as on the reputation of the recipient. A player's payoff for this indirect reciprocity game is defined as the expected benefit obtained as a recipient, reduced by the expected costs paid when acting as a donor, averaged over many rounds (see **Materials and Methods**, Section 2.4 for details).

2.2.2 Analysis of the reputation dynamics

We first explore how different social norms affect the dynamics of reputations, keeping the strategies of all players fixed. To this end, we use the concept of image matrices [Uch10, US13, OSN17]. These matrices record, at any point in time, which reputations players assign to each other. In **Fig. 2.2**, we show a snapshot of these image matrices for eight different scenarios. In all scenarios, the population consists in equal proportions of a leading-eight strategy, of ALLC (unconditional cooperators who regard everyone as good), and of ALLD (unconditional defectors who regard everyone as bad). Depending on the leading-eight strategy considered, the reputation dynamics in these scenarios can differ considerably.

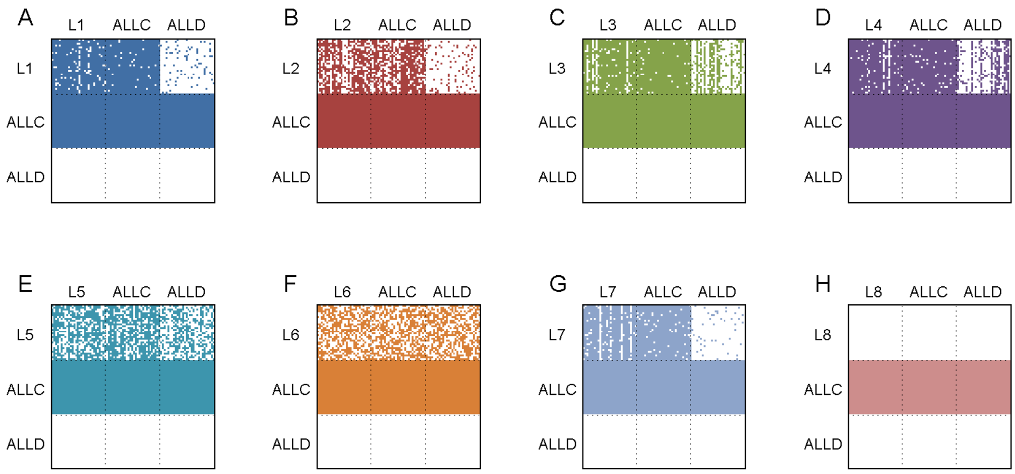

Figure 2.2: **When individuals base their decisions on noisy private information, their assessments may diverge.** Models of private information need to keep track of which player assigns which reputation to which co-player at any given time. These pairwise assessments are represented by image matrices. Here, we represent these image matrices graphically, assuming that the population consists of equal parts of a leading-eight strategist, of unconditional cooperators (ALLC) and unconditional defectors (ALLD). A colored dot means that the corresponding row-player assigns a good reputation to the column-player. Without loss of generality, we assume that ALLC players assign a good reputation to everyone, whereas ALLD players deem everyone as bad. The assessments of the leading-eight players depend on the co-player's strategy and on the frequency of perception errors. We observe that two of the leading-eight strategies are particularly prone to errors: L6 ('Stern Judging') eventually assigns a random reputation to any co-player, while L8 ('Judging') eventually considers everyone as bad. Only the other six strategies separate between conditionally cooperative strategies and unconditional defectors. Each box shows the image matrix after $2 \cdot 10^6$ simulated interactions in a population of size $N = 3 \cdot 30 = 90$. Perception errors occur at rate $\varepsilon = 0.05$, and interactions are observed with high probability, $q = 0.9$.

First, for four out of the eight scenarios, a substantial proportion of leading-eight players assigns a good reputation to ALLD players. The average proportion of ALLD players considered as good by L3, L4, L5, L6 is given by 31%, 31%, 42%, and 50%, respectively (**Fig. 2.3**).

In the eyes of these four leading-eight strategies, a bad player who defects against another bad player deserves a good reputation (**Tab. 1**). In particular, ALLD players can easily gain a good reputation whenever they encounter another ALLD player. Moreover, the higher the proportion of ALLD players in a population, the more readily they obtain a

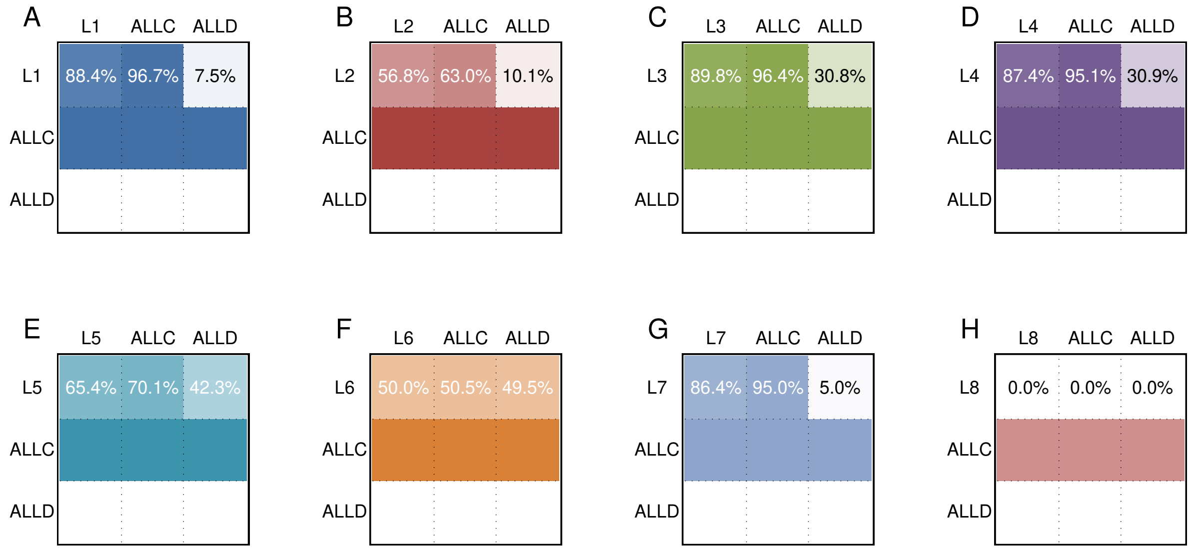

Figure 2.3: **The average images players have of each other converge over time.** While **Fig. 2.2** provides a snapshot of the reputation dynamics at a particular point in time, here we show how often players consider any other player as good on average. While ALLC players deem everyone as good and ALLD players deem everyone as bad, independent of the error rate, the leading-eight strategies differ in their assessments. Typically, a leading-eight player is most likely to consider ALLC players as good, followed by other leading-eight players and by ALLD. Whether the leading-eight strategy can resist invasion by ALLC or ALLD depends on the differences in the average image, and on the cost and benefit of cooperation. This figure uses the same parameters as in **Fig. 2.2**. The depicted numbers represent average values over the second half of $2 \cdot 10^6$ rounds of the game. Results for all strategies but L7 are robust with respect to starting from a different initial image matrix.

good reputation. This finding suggests that while L3 – L6 might be stable when these strategies are common in the population [OI04, MVC13], they have problems to restrain the payoff of ALLD when defectors are predominant.

Second, leading-eight players may sometimes collectively judge a player of their own kind as bad. In **Fig. 2.2**, such cases are represented by white vertical lines in the upper left square of an image matrix. In **Fig. 2.4** we show that such apparent misjudgments are typically introduced by perception errors. They occur, for example, when a leading-eight donor defects against an ALLC recipient, who is mistakenly considered as bad by the donor. Other leading-eight players who witness this interaction will then collectively assign a bad reputation to the donor – in their eyes, a good recipient has not obtained the help he deserves. This example highlights that under private information, an isolated disagreement about the reputation of some population member can have considerable consequences on the further reputation dynamics.

To gain a better understanding of such cases, we have analytically explored the consequences of a single disagreement in a homogeneous population of leading-eight players (see **SI**, Section 2.5, Section 2.5, for details). There we assume that initially, all players consider each other as good, with the exception of one player who considers a random co-player as bad. Assuming that no further errors occur, we study how likely the popula-

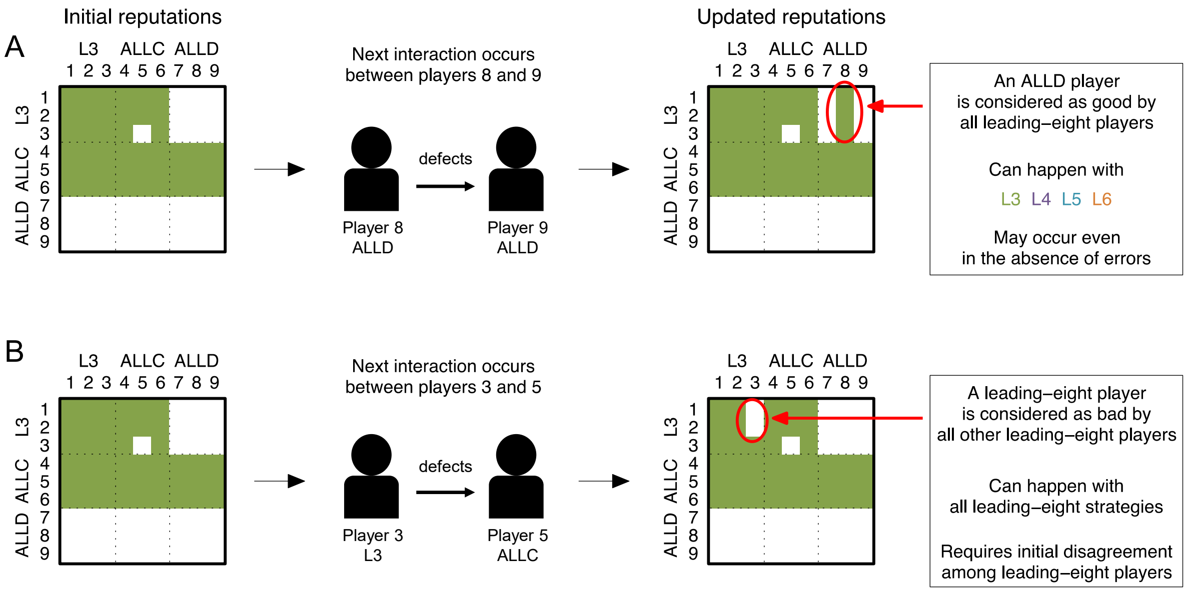

Figure 2.4: **The emergence of apparent misjudgments among leading-eight players.** As the game proceeds, it may occur that an ALLD player is considered as good by all leading-eight players (A), or that a leading-eight player is considered as bad by all other leading-eight players (B). Here we present stylized situations in which these cases occur. (A) For half of the leading-eight strategies (L3 – L6), an ALLD player can acquire a good reputation by defecting against another ALLD player. (B) When information is private and noisy, leading-eight strategies may disagree on the current reputation of a co-player. Here we show a case in which initially, player 3 is the only leading-eight player who deems player 5 as bad. If these two players are chosen for the next interaction, with player 3 as the donor, player 3 defects and acquires a bad reputation among her fellow leading-eight strategists.

tion recovers from this single disagreement (i.e., how likely the population reverts to a state where everyone is considered good), and how long it takes until recovery. While some leading-eight strategies are guaranteed to recover from single disagreements, we find that other strategies may reach an absorbing state where players mutually assign a bad reputation to each other. Moreover, even if recovery occurs, for some strategies it may take a considerable time (**Fig. 2.5**).

Two strategies fare particularly badly: L6 and L8 have the lowest probability to recover from a single disagreement, and they have the longest recovery time. This finding is also reflected in **Fig. 2.2**, which shows that these two strategies are unable to maintain cooperation. L6 eventually assigns random reputations to all co-players, whereas L8 assigns a bad reputation to everyone (**Fig. 2.6**). While L6 ('Stern') has been found to be particularly successful under public information [CSP06, SSP16, SSP18], our results confirm that this strategy is too strict and unforgiving when information is private and noisy [Uch10, US13, OSN17].

2.2.3 Evolutionary dynamics

Next we have explored how likely a leading-eight strategy would evolve when population members can change their strategies over time. We first consider a minimalistic scenario, where players can choose among three strategies only, a leading-eight strategy L_i, ALLC,

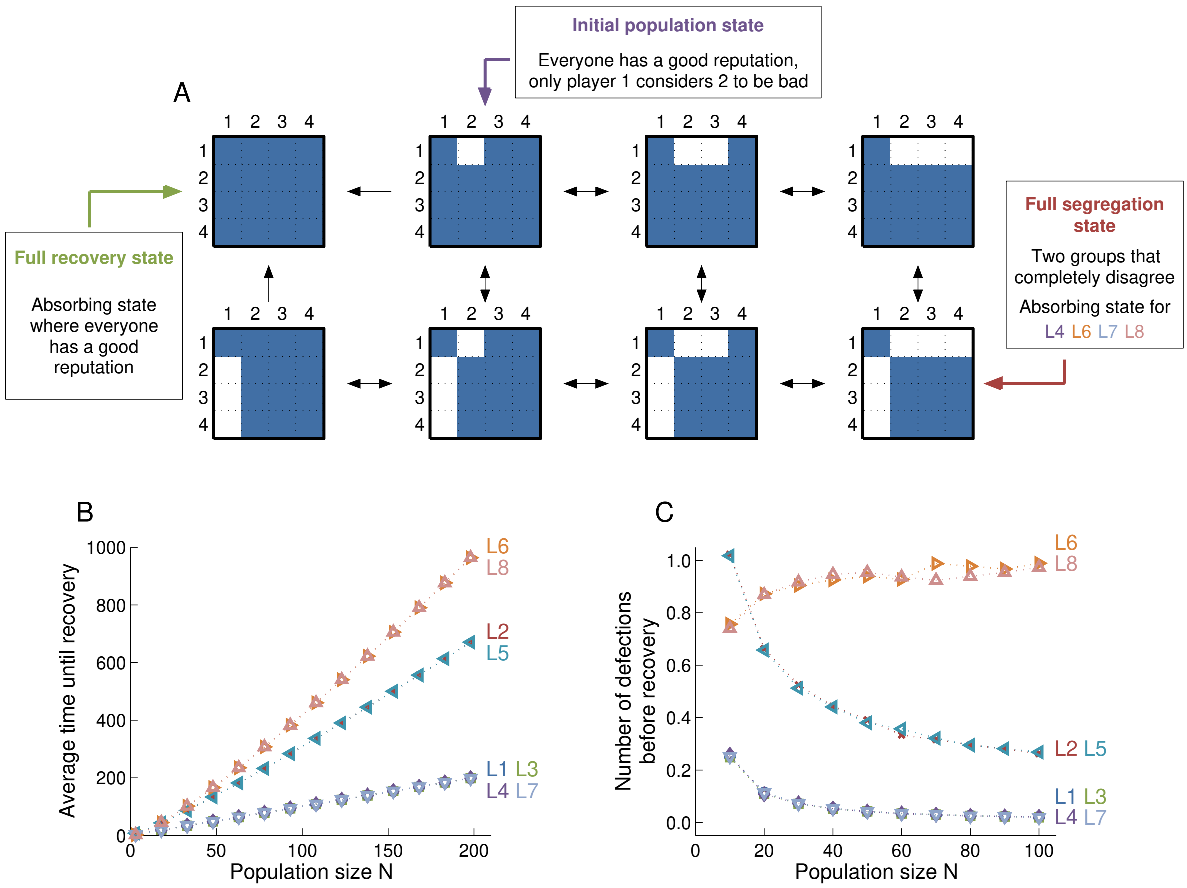

Figure 2.5: **The leading-eight strategies differ in their ability to recover from a single disagreements.** (A) We consider a homogeneous population in which everyone applies the same leading-eight strategy. Initially, all players are assumed to have a good reputation; only player 1 considers player 2 as bad. We derive analytical results for the reputation dynamics under the assumption that no more perception errors occur. We say the population recovers from a single disagreement, if it reaches the state where all players have a good reputation again. This is an absorbing state for all leading-eight strategies. We are interested in how likely the population recovers, and how long it takes until recovery. We find that only for four of the eight strategies, recovery is guaranteed (see Section 2.5.1). (B) With respect to the time until recovery, there are three cases. For four of the leading-eight strategies, recovery occurs quickly, and the recovery time is approximately linear. For all other strategies, including L2, we show that recovery may take substantially more time, being of order $N \log N$. (C) Also with respect to the expected number of defections until recovery, we observe three different qualitative behaviors. For four of the leading-eight strategies, a single perception error typically triggers no further defections, provided the population is sufficiently large. On the other extreme, for L6 and L8 we observe that any initial disagreement triggers on average one further defection.

and ALLD. To model how players adopt new strategies, we consider simple imitation dynamics [TNP06, SP13, RHR+18, HvCN18]. In each time step of the evolutionary process, one player is picked at random. With probability μ (the mutation rate), this player then adopts some random strategy, corresponding to the case of undirected learning. With the remaining probability $1-\mu$, the player randomly chooses a role model from the

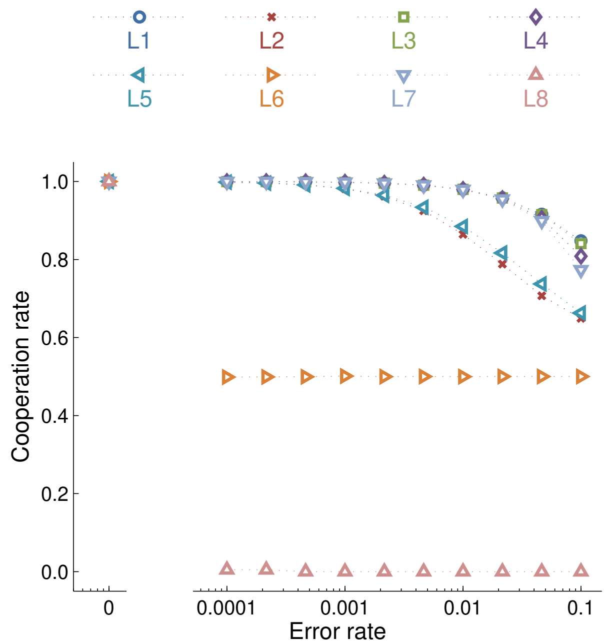

Figure 2.6: **Even rare perception errors render full cooperation impossible under "Stern" (L6) or "Judging" (L8).** We have simulated how often individuals cooperate if everyone in the population applies the same leading-eight strategy. The emerging cooperation rates depend on which leading-eight strategy is chosen and how often perception errors occur. For L6 and L8 we find that even rare perception errors undermine cooperation. Among the other six strategies, L2 and L5 are more susceptible to the noise introduced by perception errors than the other four strategies. Simulations were run with a population size of $N = 50$, assuming complete observation, $q = 1$.

population. The higher the payoff of the role model, the more likely it is that the focal player adopts the role model's strategy. Overall, the two modes of updating, mutation and imitation, give rise to an ergodic process on the space of all population compositions (see **Materials and Methods**, Section 2.4). In the following, we present results for the case when mutations are relatively rare [FI06, WGWT12].

First we have calculated for a fixed benefit-to-cost ratio of $b/c = 5$ how often each strategy is played over the course of evolution, for each of the eight possible scenarios (**Fig. 2.7**). In four cases, the leading-eight strategy is played in less than 1% of the time. These cases correspond to the four leading-eight strategies L3 – L6 that frequently assign a good reputation to ALLD players. For these leading-eight strategies, once everyone in a population has learned to be a defector, players have difficulties to reestablish a cooperative regime (in **Fig. 2.7c–f**; once ALLD is reached, every other strategy has a fixation probability smaller than 0.001). In contrast, the strategy L8 is played in substantial proportions. But in the presence of noise, players with this strategy always defect, because they deem everyone as bad (**Fig. 2.2**).

There are only three scenarios in **Fig. 2.7** that allow for positive cooperation rates. The corresponding leading-eight strategies are L1, L2 ('Consistent Standing'), and L7 ('Staying', [SON17]). For L1 and L7, the evolutionary dynamics takes the form of a rock-scissors-paper cycle [HS98, SS04, CT08, SMJ+14, SPP16]. The leading-eight strategy can be invaded by ALLC, which gives rise to ALLD, which in turn leads back to the

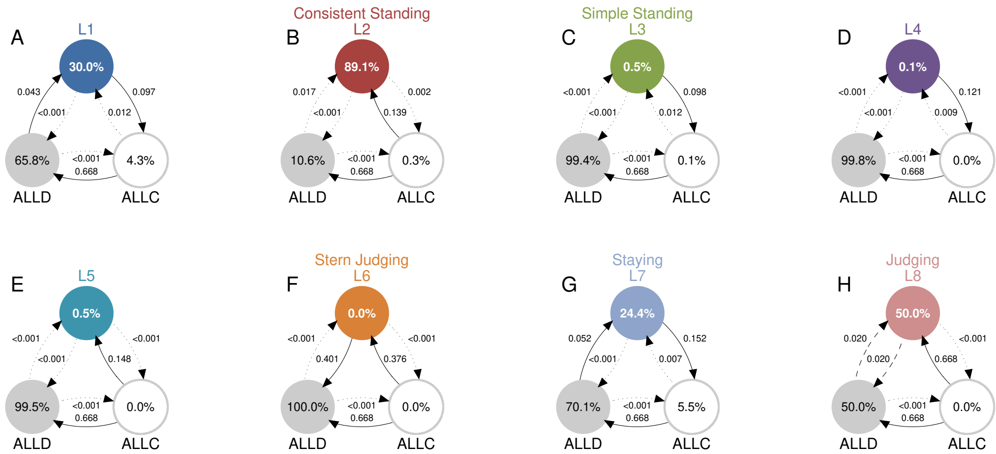

Figure 2.7: **Most of the leading-eight strategies are disfavored in the presence of perception errors.** We have simulated the evolutionary dynamics when each of the leading-eight strategies competes with ALLC and ALLD. These simulations assume that over time, players tend to imitate co-players with more profitable strategies, and that they occasionally explore random strategies (see **Materials and Methods**, Section 2.4). The numbers within the circles represent the abundance of the respective strategy in the selection-mutation equilibrium. The numbers close to the arrows represent the fixation probability of a single mutant into the given resident strategy. We use solid lines for the arrows to depict a fixation probability that exceeds the neutral probability $1/N$, and we use dotted lines if the fixation probability is smaller than $1/N$. In four cases, we find that ALLD is predominant (C - F). In one case (H), the leading-eight strategy coexists with ALLD but without any cooperation. In the remaining cases (A, B, G), we find that L1 and L7 are played with moderate frequencies, but only populations that have access to L2 ('Consistent Standing') settle at the leading-eight strategy. Parameters: Population size $N=50$, benefit $b=5$, cost $c=1$, strength of selection $s=1$, error rate $\varepsilon=0.05$, observation probability $q=0.9$, in the limit of rare mutations, $\mu \to 0$.

leading-eight strategy. Because ALLD is most robust in this cycle, the leading-eight strategies are played in less than a third of the time (**Fig. 2.7a**,**g**).

Only Consistent Standing, L2, is able to compete with ALLC and ALLD in a direct comparison (**Fig. 2.7b**). Under Consistent Standing, there is a unique action in each possible situation that allows a donor to obtain a good standing. For example, when a good donor meets a bad recipient, the donor keeps her good standing by defecting, but loses it by cooperating. Compared to Stern Judging, which has a similar property [SSP18], Consistent Standing incentivizes cooperation more strongly. When two bad players interact, the correct decision according to Consistent Standing is to cooperate, whereas a stern player would defect (**Tab. 1**).

Nevertheless, we find that even when Consistent Standing is common, the average cooperation rate in the population rarely exceeds 65%. To show this, we have repeated the previous evolutionary simulations for the eight scenarios while varying the benefit-to-cost ratio, the error rate, and the observation probability (**Fig. 2.8**).

These simulations confirm that five of the leading-eight strategies cannot maintain any cooperation when competing with ALLC and ALLD. Only for L1, L2, L7, average

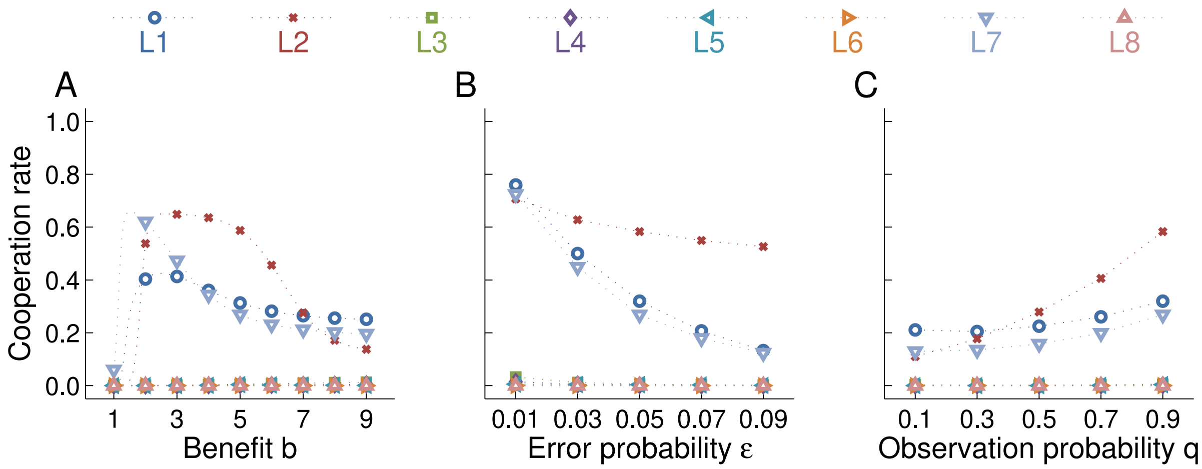

Figure 2.8: **Noise can prevent the evolution of full cooperation even if leading-eight strategies evolve.** We have repeated the evolutionary simulations in **Fig. 2.7**, but varying (A) the benefit of cooperation, (B) the error rate, and (C) the observation probability. The graph shows the average cooperation rate for each scenario in the selection-mutation equilibrium. This cooperation rate depends on how abundant each strategy is in equilibrium, and on how much cooperation each strategy yields against itself in the presence of noise. For five of the eight scenarios, cooperation rates remain low across the considered parameter range. Only the three other leading-eight strategies can persist in the population, but even then cooperation rates typically remain below 70%. We use the same baseline parameters as in **Fig. 2.7**.

cooperation rates are positive, reaching a maximum for intermediate benefit-to-cost ratios (**Fig. 2.8a**). If the benefit-to-cost ratio is too low, we find that each of these leading-eight strategies can be invaded by ALLD, whereas if the ratio is too high, ALLC can invade (**Fig. 2.9**).

In between, Consistent Standing may outperform ALLC and ALLD, but in the presence of noise it does not yield high cooperation rates against itself. Even if all interactions are observed ($q=1$), cooperation rates in a homogeneous L2 population drop below 70% once the error rate exceeds 5% (**Fig. 2.6**). Our analytical results in Section 2.5.1 suggest that while L2 populations always recover from single disagreements, it may take them a substantial time to do so, during which further errors may accumulate (**Fig. 2.5**). As a result, whereas L2 seems most robust when co-evolving with ALLC and ALLD, it is unable to maintain full cooperation. Furthermore, additional simulation results suggest that even if L2 is able to resist invasion by ALLC and ALLD, it may be invaded by mutant strategies that differ in only one bit from L2 (**Fig. 2.10**).

So far, we have assumed that mutations are rare, such that populations are typically homogeneous. Experimental evidence, however, suggests that there is considerable variation in the social norms employed by subjects [EF09, WM00, MSBK01, SS06, BKO05, vAS16, SMUE16]. While some subjects are best classified as unconditional defectors, others act as unconditional cooperators, or use more sophisticated higher-order strategies [SMUE16]. In agreement with these experimental studies, there is theoretical evidence that some leading-eight strategies like L7 may form stable coexistences with ALLC [OSN17]. Thus, we show further figures with evolutionary results for higher mutation rates, in which such

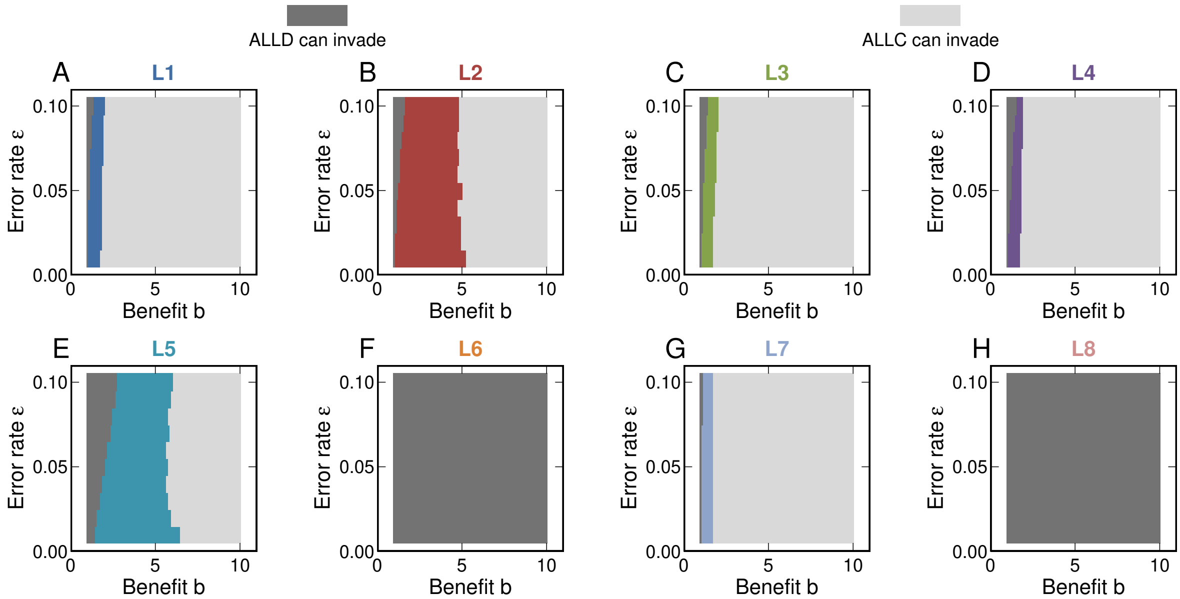

Figure 2.9: **Stability of each leading-eight strategy against invasion by ALLC and ALLD.** We consider a population of $N-1$ players adopting one of the leading-eight strategies, and we explore whether either a single ALLD mutant or an ALLC mutant gains at least the payoff of the residents. To this end, we vary the benefit of cooperation (x-axis) and the probability of perception errors (y-axis). Parameter regions in which ALLD can invade are depicted in dark grey, whereas parameter regions in which ALLC can invade are shown as light grey. Only in the colored region, the respective leading-eight strategy is stable against invasion by either ALLC or ALLD. Except for L2 and L5, we find that most leading-eight strategies are either unstable (F and H), or they only resist invasion in a small subset of the parameter space (A,C,D,G). Parameters: $N=50$, $c=1$, $q=0.9$.

coexistences are possible (**Fig. 2.11– Fig. 2.13**). There we show that in the three cases L1, L2, L7, populations may consist of a mixture of the leading-eight strategy and ALLC for a considerable time. However, in agreement with our rare-mutation results, we find for L1 and L7 that this mixture of leading-eight strategy and ALLC is susceptible to stochastic invasion by ALLD.

2.3 Discussion

Indirect reciprocity explores how cooperation can be maintained when individuals assess and act on each other's reputations. Simple strategies of indirect reciprocity like Image Scoring [NS98b, NS98a] have been suspected to be unstable, because players may abstain from punishing defectors in order to maintain their own good score [LH01]. In contrast, the leading-eight strategies additionally take the context of an interaction into account. They have been considered to be prime candidates for stable norms that maintain cooperation [OI04, OI06]. Corresponding models, however, assume that each pairwise interaction is only witnessed by one observer, who disseminates the outcome of the interaction to all other population members. As a consequence, the resulting opinions within a population will be perfectly synchronized. Even if donors are subject to implementation errors, or if the observer misperceives an interaction, all players will have the same image of the

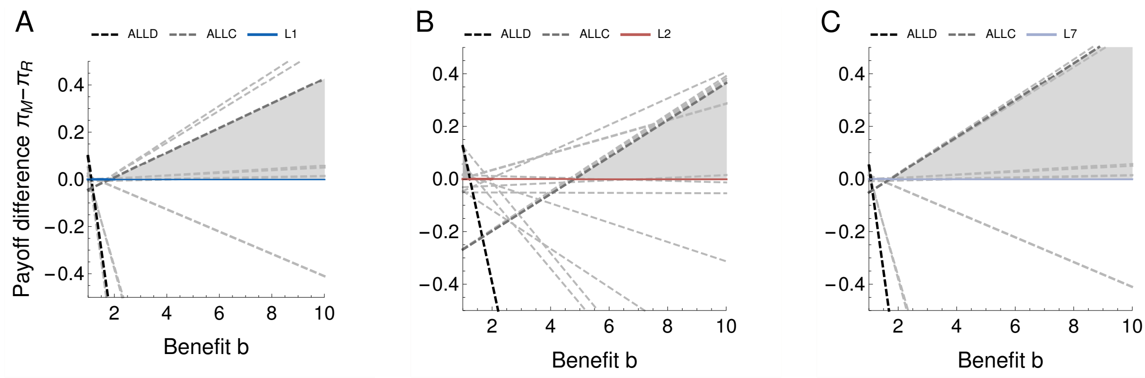

Figure 2.10: **Stability against ALLC and ALLD does not imply stability against all mutant invasions.** We consider a resident population that either applies the leading-eight strategy L1 (A), L2 (B), or L7 (C). For each resident population, we consider 14 possible mutant strategies, including ALLD (black), ALLC (dark grey) and the twelve strategies that differ from the resident in only one bit (light grey). For different benefit values, we plot the payoff advantage $\pi_M - \pi_R$ of a single mutant among $N-1$ residents. If for a given b-value there is a line in the upper half of the panel, the resident strategy can be invaded by the respective mutant. We find that even in the parameter region where a leading-eight strategy can resist invasion by ALLC and ALLD, other mutant strategies may be able to invade. For example, L2 can resist invasion by ALLD for $b \gtrsim 1.5$, and invasion by ALLC if $b \lesssim 5$. In between, for $1.5 \le b \le 5$ there is a different mutant strategy that can invade. This mutant coincides with L2, except that it assesses a good donor who defects against a bad recipient as bad. Similar cases of successful mutants different from ALLC and ALLD also exist for L1 and L7. Parameters: $N=50$, $c=1$, $q=0.9$, $\varepsilon = 0.05$.

donor after the interaction has taken place.

While the assumption of perfectly synchronized reputations is a useful idealization, we believe that it may be too strict in some applications. Subjects often differ in the prior information they have, and even if everyone has access to the same information (as is often the case in online platforms [RZSL06, RvdR12]), individuals differ in how much weight they attribute to different pieces of evidence. As a result, individuals might disagree on each other's reputations. These disagreements can proliferate over time. Herein, we have thus systematically compared the performance of the leading-eight strategies when information is incomplete, private, and noisy. The leading-eight strategies differ in how they are affected by the noise introduced by private perception errors. Strategies like Stern Judging, that have been shown to be highly successful under public information [CSP06, SSP16, SSP18], fail to distinguish between friend and foe when information is private. While we have considered well-mixed populations in which all players are connected, this effect might be even more pronounced when games take place on a network [LHN05, SF07]. If players are only able to observe interactions between players in their immediate neighborhood, network-structured populations may amplify the problem of incomplete information. Pairwise interactions that one player is able to observe may be systematically hidden from his neighbor's view. Thus the study of indirect reciprocity on networks points to an interesting direction for future research.

The individuals in our model are completely independent when forming their beliefs.

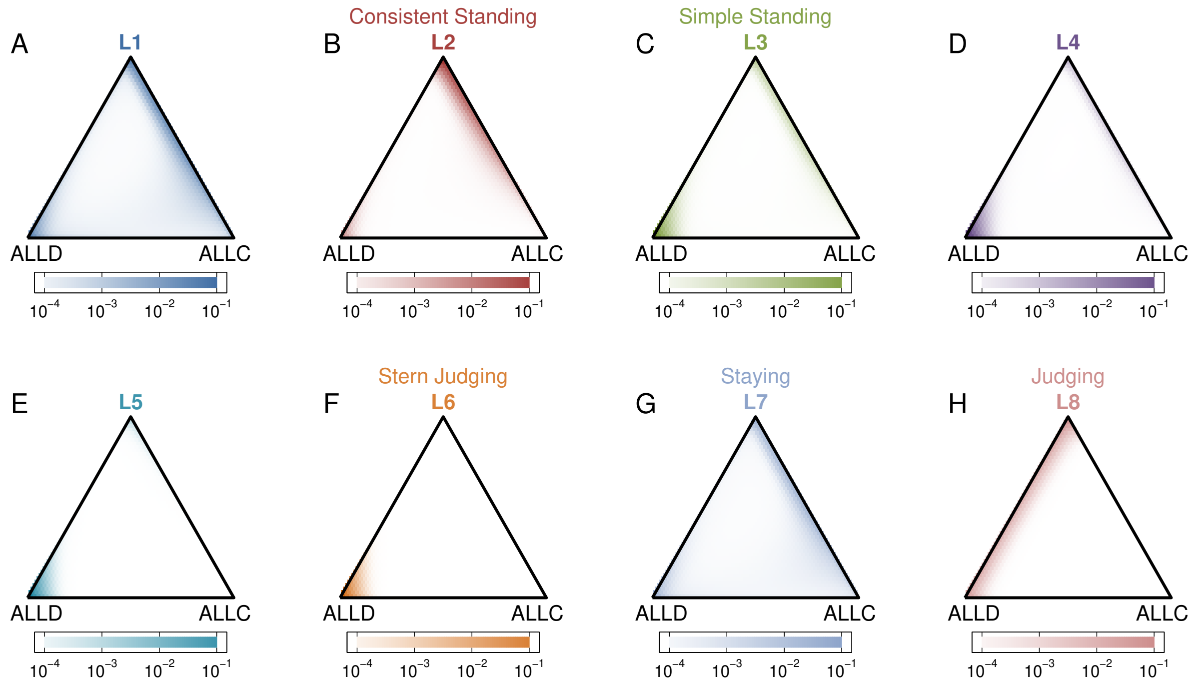

Figure 2.11: **When mutations are sufficiently frequent, three of the leading-eight strategies can maintain cooperation in coexistence with ALLC.** In the main text we have presented evolutionary results in the limit of rare mutations. In this limit, populations are homogeneous most of the time, rendering stable coexistences of multiple strategies impossible. Here, we present results for non-vanishing mutation rates when players can choose between a leading-eight strategy, ALLC, and ALLD. The possible population compositions are represented by a triangle. The corners of the triangle correspond to homogeneous populations, whereas interior points yield the corresponding mixed populations. The colors indicate how often the respective region of the state space is visited by the evolutionary process. (A,B,G) For the considered parameter values, we find that in three cases, a stable coexistence between a leading-eight strategy and ALLC can maintain cooperation for a considerable fraction of time. (C–F) In four cases, we find that populations typically find themselves in the vicinity of ALLD. (H) In the presence of noise, there is neutral drift between L8 and ALLD. Along that edge, no player cooperates. Parameters are the same as in **Fig. 2.7**, but with a strictly positive mutation rate $\mu = 0.01$.

In particular, they are not affected by the opinions of others, swayed by gossip and rumors, or engaged in communication. Experimental evidence suggests that even when all subjects witness the same social interaction, gossip can greatly modify beliefs, and align the subjects' subsequent behaviors [SKSM07]. Seen from this angle, our study highlights the importance of coordination and communication for the stability of indirect reciprocity. Social norms that fail when information is noisy and private may sustain full cooperation when information is mutually shared and discussed.

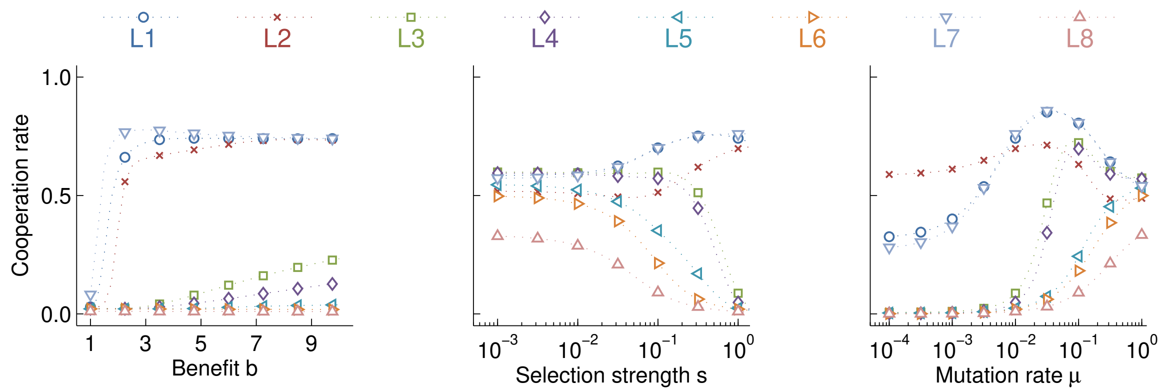

Figure 2.12: **Cooperation is most likely to evolve for high benefit-to-cost ratios and when mutations are sufficiently frequent.** To explore the robustness of our findings, we have systematically varied three key parameters of our model, the benefit b, the selection strength s, and the mutation rate μ. For each combination of parameter values, we have simulated the evolutionary process between ALLC, ALLD, and L_i, for each of the leading-eight strategies L_i. The figure shows the resulting cooperation rates. We recover that for usual b/c-ratios, sufficiently strong selection, and rare mutations, there are only three leading-eight strategies that can maintain some cooperation. Surprisingly, the maximum amount of cooperation is achieved for intermediate mutation rates, $0.01 \leq \mu \leq 0.1$. Here we observe relatively stable coexistences between ALLC and either L1 or L7. Baseline parameters: $b=5$, $s=1$ and $\mu=0.01$; all other parameters are the same as in **Fig. 2.11**.

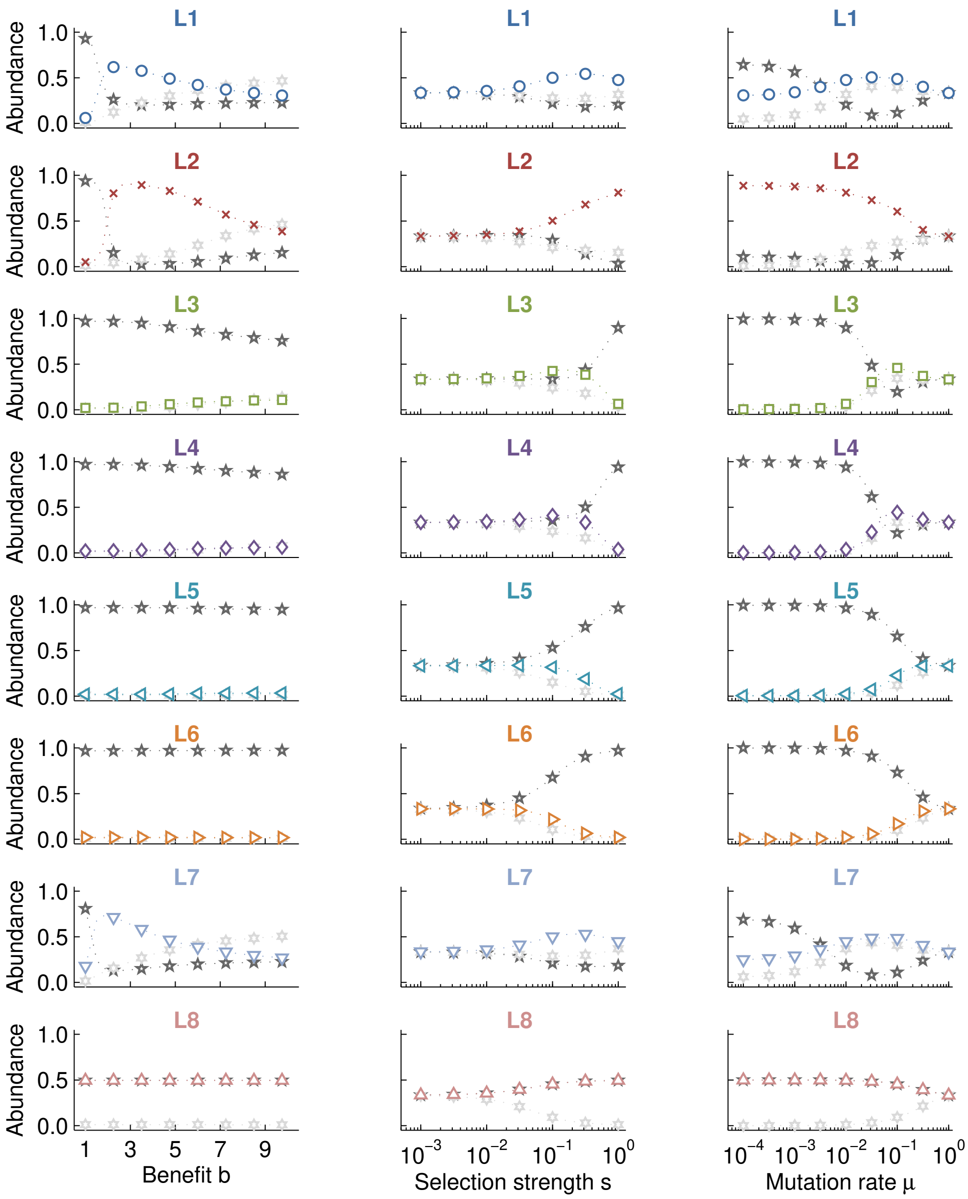

Figure 2.13: **Evolutionary abundance of the leading-eight strategies across different parameter regimes.** This figure considers the same scenario as **Fig. 2.12**, but it depicts the abundance of each strategy in the selection-mutation equilibrium. The abundance of ALLC is depicted in light grey, ALLD is shown in dark grey, and for the leading-eight strategy we use the respective color. In three cases (for L1, L2, L7) we observe that ALLC is played with positive frequency even as the selection strength s increases. In these cases, ALLC typically coexists with a majority of players who apply the leading-eight strategy.

2.4 Materials and Methods

2.4.1 Model setup

We consider N individuals in a well-mixed population. Each player's strategy is given by a pair (α, β). The first component,

$$\alpha = (\alpha_{gCg},\ \alpha_{gCb},\ \alpha_{bCg},\ \alpha_{bCb},\ \alpha_{gDg},\ \alpha_{gDb},\ \alpha_{bDg},\ \alpha_{bDb}) \tag{2.1}$$

corresponds to the player's assessment rule. An entry α_{xAy} is equal to one if the player assigns a good reputation to a donor of reputation x who chooses action A against a recipient with reputation y. Otherwise, if such a donor is considered as bad, the corresponding entry is zero. The second component of the strategy,

$$\beta = (\beta_{gg},\ \beta_{gb},\ \beta_{bg},\ \beta_{bb}). \tag{2.2}$$

gives the player's action rule. An entry β_{xy} is equal to one if the focal player with reputation x cooperates with a recipient with reputation y; otherwise it is zero. The assessment and action rules of the leading-eight strategies are shown in **Tab. 1**. We define ALLC as the strategy with assessment rule $\alpha = (1, \dots, 1)$ and action rule $\beta = (1, \dots, 1)$. ALLD is the strategy with $\alpha = (0, \dots, 0)$ and $\beta = (0, \dots, 0)$.

2.4.2 Reputation dynamics

To simulate the reputation dynamics for players with fixed strategies, we consider the image matrix [Uch10, US13, OSN17] $M(t) = \big(m_{ij}(t)\big)$ of a population at time t. Its entries satisfy $m_{ij}(t) = 1$ if player i deems player j as good at time t, and $m_{ij}(t) = 0$ otherwise. We assume that initially, all players have a good reputation, $m_{ij}(0) = 1$ for all i, j. However, our results are unchanged if the players' initial reputations are assigned randomly. We only get slightly different results if all initial reputations are bad; in that case L7 players are unable to acquire a good reputation over the course of the game (for details, see **SI**, Section 2.5).

In each round t, two players i and j are drawn from the population at random, a donor and a recipient. The donor then decides whether to cooperate. Her choice is uniquely determined by her action rule β and by the reputations she assigns to herself and to the recipient, $m_{ii}(t)$ and $m_{ij}(t)$. The donor and the recipient always observe the donor's decision; all other players independently observe it with probability q. With probability ε, a player who observes the donor's action misperceives it, independent of the other players. All players who observe the interaction update their assessment of the donor according to their assessment rule. This yields the image matrix $M(t+1)$.

We iterate the above elementary process over many rounds (our figures are based on 10^6 rounds or more). Based on these simulations, we can now calculate how often player i considers j to be good on average, and how often player i cooperates with j on average. If the estimated pairwise cooperation rate of i against j is given by $\hat{x}_{ij}$, we define player i's payoff as

$$\hat{\pi}_i = \frac{1}{N-1} \sum_{j \neq i} b\hat{x}_{ji} - c\hat{x}_{ij}, \tag{2.3}$$

i.e. the benefits i gained by receiving cooperation from j ($b\hat{x}_{ji}$), reduced by the costs of i's cooperation towards j ($c\hat{x}_{ij}$), and averaged over all $N-1$ co-players j.

2.4.3 Evolutionary dynamics

On a larger timescale, we assume that players can change their strategies (α, β). To model the strategy dynamics, we consider a pairwise comparison process [TNP06, SP13, RHR+18]. In each time step of this process, one individual is randomly chosen from the population. With probability μ this individual then adopts a random strategy, with all other available strategies having the same probability to be picked. With the remaining probability $1-\mu$ the focal individual i chooses a random role model j from the population. If the players' payoffs according to Eq. [2.3] are $\hat{\pi}_i$ and $\hat{\pi}_j$, player i adopts j's strategy with probability $P(\hat{\pi}_j, \hat{\pi}_i) = \left(1 + \exp\left(-s(\hat{\pi}_j - \hat{\pi}_i)\right)\right)^{-1}$ [ST98]. The parameter $s \geq 0$ is the 'strength of selection'. It measures how strongly imitation events are biased in favor of strategies with higher payoffs. For $s = 0$ we obtain $P(\hat{\pi}_j, \hat{\pi}_i) = 1/2$, and imitation occurs at random. As s increases, payoffs become increasingly relevant when i considers imitating j's strategy.

In the main text, we assume players can only choose between a leading-eight strategy L_i, ALLC, and ALLD. As we show in **Fig. 2.10**, the stability of a leading-eight strategy may be further undermined if additional mutant strategies are available. Moreover, in the main text we only report results when mutations are comparably rare [FI06, WGWT12]. In **Fig. 2.11**– **Fig. 2.13** we show further results for substantial mutation rates. Given the players' payoffs for each possible population composition, the selection-mutation equilibrium can be calculated explicitly. All details are provided in the **SI**, Section 2.5.

2.5 Supplementary Information

We are interested in the evolution of indirect reciprocity under incomplete and noisy information. To this end, we describe the dynamics on two separate timescales. First, we consider the *game dynamics*. Here, we take the players' strategies as given, and we compute how the players' reputations change over time, how often they cooperate, and which payoffs they obtain on average. Second, we describe the *evolutionary dynamics*. Here, we allow players to change their strategies over time, assuming that strategies that yield a comparably high payoff are more likely to be adopted. In the following, we describe these two dynamics in more detail.

2.5.1 Game dynamics

Description of the indirect reciprocity game

We consider a well-mixed population of size N whose members engage in a series of cooperative interactions. In each time step, two members of the population are randomly drawn from the population, a donor and a recipient. The donor can then decide whether or not to pay a cost $c > 0$ to transfer a benefit $b > c$ to the recipient. We interpret paying the cost as cooperation (C) and refusing to do so as defection (D). The donor's decision is partially observable: the donor and the receiver always learn whether or not the donor decided to cooperate, whereas the other population members independently observe it with probability q. Observations may be subject to noise: with probability ε the donor's action is misinterpreted such that a C is taken for a D, or vice versa. All players are equally likely to misperceive an action, independent of whether or not they actively took part in the respective interaction. We assume that there are infinitely many rounds in which players are asked to cooperate, and that the players' payoffs for the indirect reciprocity game are defined as their average payoffs over time (explicit definitions will be provided in the next section).

To make their cooperation decisions, we assume that each player is equipped with a strategy and a private reputation repository. The reputation repository is used to keep track of the reputations of all population members (as illustrated in **Fig. 2.1**). In line with a large body of the previous literature [NS05, Sig10, Sig12], we assume that reputations are binary: players are either considered as 'good' or 'bad'. We represent the state of player i's reputation repository at time t by an N-dimensional vector

$$\mathbf{m}_i(t) = \Big(m_{i1}(t), \ldots, m_{iN}(t)\Big). \tag{2.4}$$

The entries m_{ij} of this vector are either 0 or 1, whereby $m_{ij}(t)=1$ means that player i assigns a good reputation to player j at time t, whereas $m_{ij}(t)=0$ means that i views j as bad. We assume that $\mathbf{m}_i(0) = (1, \ldots, 1)$ for all players i. That is, initially all players consider everyone as good.

A player's strategy then needs to tell the player (i) how to update the entries in her reputation repository after observing a donor's action, and (ii) whether to cooperate if the focal player finds herself in the role of the donor. The first aspect is determined by the player's assessment rule. Assessment rules take the form of an 8-dimensional vector

$$\alpha = (\alpha_{gCg},\ \alpha_{gCb},\ \alpha_{bCg},\ \alpha_{bCb},\ \alpha_{gDg},\ \alpha_{gDb},\ \alpha_{bDg},\ \alpha_{bDb}). \tag{2.5}$$

The entries of α can again take either the value 0 or 1. An entry $\alpha_{xAy}=1$ means that it is regarded as good if a donor with reputation x chooses action A against a recipient with reputation y. Analogously, $\alpha_{xAy}=0$ means that the focal player assigns a bad reputation to the corresponding donor. We note that after any interaction, players only update the donor's reputation; the reputation of the receiver remains unaffected. Assessment rules of the form [5.64] are called third-order assessment rules, since they depend on the donor's action, on the reputation of the donor, and on the reputation of the recipient. If the assessment is independent of the donor's current reputation, the assessment rule is referred to as second-order; and if the assessment is independent of both players' reputations, the assessment rule is referred to as first-order. The two unconditional cases $\alpha=\mathbf{1}:=(1,\ldots,1)$ and $\alpha=\mathbf{0}:=(0,\ldots,0)$, according to which donors are always considered as good or bad, respectively, are zeroth-order assessment rules.

If the focal player finds herself in the role of the donor, her decision is determined by her action rule. Action rules are given by a 4-dimensional vector

$$\beta = (\beta_{gg},\ \beta_{gb},\ \beta_{bg},\ \beta_{bb}). \tag{2.6}$$

An entry $\beta_{xy} = 1$ indicates that a respective donor chooses to cooperate if her own reputation is x and the recipient's reputation is y. An entry $\beta_{xy}=0$ indicates that she defects. A player's strategy $\sigma=(\alpha;\beta)$ is a combination of an assessment rule and an action rule. Since there are 2^8 different assessment rules and 2^4 different action rules, the space of third-order strategies contains $2^{12} = 4,096$ elements.

Herein, we have only considered the dynamics for a small subset of third-order strategies (but the methods that we introduce in the following sections equally apply to all other strategies). The subset we have considered consists of the so-called leading-eight strategies [OI06]. Under the assumption that all relevant information is shared publicly and that all players agree on everyone's reputation, past research has shown that if the whole population adopts any of the leading-eight strategies, the population will be fully cooperative and no other strategy can invade [OI04]. However, the assumption of publicly shared information seems rather stringent; it implies that the reputation repositories of all players agree at all times, $\mathbf{m}_i(t)=\mathbf{m}_j(t)$ for all i, j and t. Instead, we explore the robustness of the leading-eight strategies when information is incomplete, noisy, and private. In that case, different players may hold different beliefs about their co-players, such that for any pair of players i and j there may be times for which $\mathbf{m}_i(t) \neq \mathbf{m}_j(t)$. To probe the robustness of the leading-eight strategies, we let each of these strategies compete with the two unconditional strategies ALLC and ALLD. Herein, we define these two strategies as ALLC= $(\mathbf{1};\mathbf{1})$ and ALLD= $(\mathbf{0};\mathbf{0})$. We note that alternative strategy representations are possible: For example, we could define ALLC as the strategy that deems everyone as bad but cooperates anyway. Using such an alternative definition would not alter our qualitative results.

A formal description of the reputation dynamics

If the strategies $(\alpha^i;\beta^i)$ are given for all players $1 \leq i \leq N$, we can describe the reputation dynamics of the population as a Markov chain. As the state of the Markov chain at time t, we collect all individual reputation repositories into an $N \times N$ matrix

$$M(t) = \Big(m_{ij}(t)\Big). \tag{2.7}$$

We call $M(t)$ the image matrix of the population at time t. Given the players' strategies and the current image matrix $M = M(t)$ we can in principle calculate the transition probability $h_{M,M'}$ that describes how likely we are to find the population in state M' at time $t+1$. Because in every single time-step at most one column of the image matrix can change (the column with respect to the player who was randomly chosen to be the donor), we observe that $h_{M,M'}=0$ if M and M' differ in more than one column. Otherwise, when the image matrices M and M' differ in at most one column, the exact value of $h_{M,M'}$ depends on the players' strategies, on the observation probability q, and on the error probability ε.

We note here that depending on the players' strategies, the Markov chain of the reputation dynamics does not generally need to be ergodic: there are cases where the initial configuration of the image matrix matters for the chain's behavior over time, thus contradicting the irreducibility condition for ergodicity. We find one such case in the leading eight, the strategy L7. According to the assessment rule of L7 (**Tab. 1**), a bad donor can only gain a good reputation by cooperating against a good recipient. However, in the case where there are no good recipients to start with, it is impossible for donors to gain a good standing in the eyes of an L7 observer. We deal with this degenerate case separately further below.

For ergodic reputation dynamics, the average probability to observe a given image matrix M approaches an invariant distribution $\mathbf{v}=(v_M)$ of the Markov chain $H=(h_{M,M'})$ over time, such that $\mathbf{v}H=\mathbf{v}$ and $\sum_M v_M = 1$. The entries v_M give the expected frequency with which we observe the image matrix M over the course of the indirect reciprocity game. Given the invariant distribution $\mathbf{v}$, we can calculate the average probability that player i considers j as good as

$$\bar{m}_{ij} = \sum_M v_M \cdot m_{ij}. \tag{2.8}$$

Moreover, we can calculate the average probability $\bar{x}_{ij}$ with which player i cooperates with player j over the course of the game as

$$\bar{x}_{ij} = \sum_M v_M \beta^i_{m_{ii},m_{ij}}. \tag{2.9}$$

Using this average cooperation probability, we define player i's expected payoff $\bar{\pi}_i$ as

$$\bar{\pi}_i = \frac{1}{N-1} \sum_{j \neq i} b\bar{x}_{ji} - c\bar{x}_{ij}. \tag{2.10}$$

Unfortunately, we note that the above approach quickly becomes computationally infeasible as the population becomes large. For a population of size N, the image matrix has N^2 entries, implying that there are 2^{N^2} possible image matrices.

Instead of calculating numerically exact payoffs according to Eq. [2.10], we have thus simulated the reputation dynamics to obtain estimates $\hat{x}_{ij}$ for the average probability that player i cooperates with j. These estimates are then plugged into Eq. [2.10] to obtain estimated payoffs

$$\hat{\pi}_i = \frac{1}{N-1} \sum_{j \neq i} b\hat{x}_{ji} - c\hat{x}_{ij}. \tag{2.11}$$

By the theory of Markov chains, the estimated payoffs according to [2.11] converge to the true values [2.10] if the simulation is iterated for sufficiently many time steps. Unless noted otherwise, we have simulated the Markov chain for $2 \cdot 10^6$ time steps, using a population of size $N = 50$. The respective simulations have been run with MATLAB; all scripts are provided in the **Appendix**, Section 2.6.

In cases where the Markov chain is not ergodic, the players' average payoffs according to Eq. [2.11] may depend on their initial reputations. All our figures are based on the assumption that all players begin with a good reputation. To explore the robustness of our results, we have re-run the simulations in **Fig. 2.2** for two alternative scenarios. First we have run ten independent simulations in which the initial reputations are assigned randomly. For each of these ten simulations, we have obtained the same result as reported in the main text. Second, we have run a simulation assuming that all players start with a bad reputation. Again, we obtain the same result as in **Fig. 2.2** for seven out of the eight cases. Only for L7 the result is different in this special case; since there are no good recipients in the starting configuration, it is impossible for donors to gain a good standing in the eyes of an L7 observer. Hence, all L7 players keep their initial bad assessment of all co-players under such an initial assignment of reputations. We conclude that for this degenerate case, L7 does not yield any cooperation at all.

Recovery analysis after single disagreements

While we use simulations to explore the general reputation dynamics under noisy private information, we can derive analytical results in the limiting case that while players privately keep track of each others' reputations, their actions are perfectly observable ($q = 1$) and that perception errors are rare ($\varepsilon \to 0$). To this end, we consider a homogeneous population of players who all apply the same leading-eight strategy (as given in **Tab. 1**). We assume that initially, all players perceive everyone as good; only player 1 perceives player 2 as bad, possibly because of an error (**Fig. 2.5a**). That is, the initial image matrix is given by $M(0) = M^0$ with entries

$$m^0_{ij} = \begin{cases} 0 & \text{if } i = 1, j = 2 \\ 1 & \text{otherwise.} \end{cases} \tag{2.12}$$

We assume that in subsequent rounds, no further errors occur and that all individuals observe all co-players' interactions. We say that the population recovers from a single disagreement, if it returns to the state where all players have a good reputation starting from the state M^0. In the following we are interested in the following two quantities, depending on the applied leading-eight strategy L_i:

1. The population's recovery probability ρ_i
2. The expected time until recovery τ_i (conditioned on that the population actually recovers)

The following Proposition simplifies the corresponding analysis.

Proposition 1. *Consider the indirect reciprocity game for a population in which everyone applies the same leading-eight strategy L_i. Moreover, assume that the initial image matrix is $M(0) = M^0$ as defined in [2.12], and let $M(t)$ denote the image matrix at some subsequent time $t > 0$ according to the process with perfect observation and no noise, $q = 1$ and $\varepsilon = 0$. Then $M(t) \in \mathcal{M}$, where $\mathcal{M}$ is the set of all image matrices that satisfy the following three conditions*

$$\text{(i) } m_{ii} = 1 \text{ for all } i, \quad \text{(ii) } m_{ij} = 1 \text{ for all } i, j \geq 2, \quad \text{(iii) } m_{i1} = m_{j1} \text{ for all } i, j \geq 2. \tag{2.13}$$

All proofs are provided in the **Appendix**, Section 2.6. The above Proposition guarantees that in the process with private information, perfect observation, and no noise, (*i*) all players think of themselves as good, (*ii*) the players $2 \leq i, j \leq N$ consider each other as good, and (*iii*) all players $2 \leq i, j \leq N$ have the same opinion about player 1.

Proposition 1 is useful because it allows us to consider a simplified state space when we study the private information setting instead of the public one. Instead of considering the space of all image matrices M, in the following we consider the space of all tuples (r, k) with $r \in \{0, 1\}$ and $k \in \{0, \ldots, N-1\}$. The value of r refers to the reputation of player 1 from the perspective of all other players (due to Proposition 1 (*iii*) all other players agree on player 1's reputation). We use $r = 1$ to indicate that player 1 is perceived as good, and $r = 0$ to indicate that she is perceived as bad. The value of k refers to the number of co-players that player 1 considers as good (due to Proposition 1, we can treat all other players as equivalent). In this reduced state space, the initial state thus corresponds to the pair $(1, N-2)$, and the full recovery state is $(1, N-1)$, see **Fig. 2.5a** for an illustration.

Let $f^i(r, k;\ r', k')$ denote the transition probabilities for the reduced state space; the value of $f^i(r, k;\ r', k')$ corresponds to the probability that a population of L_i players moves from state (r, k) to (r', k') in one round. We can deduce these probabilities as follows:

Transition $(1,k) \to (1,k{+}1)$. This case can only occur if a player $i>1$ is chosen to be the donor who is perceived as bad by player 1. Given that the current state is $(1,k)$, it follows from Proposition 1 that the donor considers everyone as good, and hence she cooperates. If player 1 considers the receiver to be good, this leads her to assign a good reputation to the donor, independent of the applied leading-eight strategy L_i. Otherwise, if player 1 considers the receiver to be bad, the donor only obtains a good reputation for L_1, L_2, L_3, and L_5. Therefore, the corresponding transition probability is

$$f^i(1,k;\ 1,k{+}1) = \begin{cases} \frac{N-k-1}{N} & \text{if } i \in \{1,2,3,5\} \\ \frac{N-k-1}{N}\frac{k+1}{N-1} & \text{if } i \in \{4,6,7,8\}. \end{cases} \tag{2.14}$$

Transition $(1,k) \to (1,k{-}1)$. This case can only occur if a player $i>1$ is randomly chosen to act as the donor who is perceived as good by player 1. Similar to before, player i will always cooperate, which is only considered as bad by player 1 if the receiver is considered as bad by player 1 and if the applied strategy is either L_2, L_5, L_6, or L_8. Therefore, the transition probability is

$$f^i(1,k;\ 1,k{-}1) = \begin{cases} 0 & \text{if } i \in \{1,3,4,7\} \\ \frac{k}{N}\frac{N-k-1}{N-1} & \text{if } i \in \{2,5,6,8\}. \end{cases} \tag{2.15}$$

Transition $(1,k) \to (0,k)$. This case can only occur if player 1 is chosen to be the donor, and if player 1 defects against the receiver (which in turn requires player 1 to consider the receiver as bad). The corresponding transition probability is

$$f^i(1,k;\ 0,k) = \frac{1}{N}\frac{N-k-1}{N-1}. \tag{2.16}$$

Transition $(0,k) \to (0,k{+}1)$. This case requires that a player $i>1$ is chosen to be the donor who is considered as bad by player 1. This donor cooperates, unless the randomly chosen receiver happens to be player 1 (who is bad from the perspective of all other players). Thus, player 1 considers the donor as good after this round unless the receiver is player 1, or the receiver is a group member that is considered as bad by player 1 and the applied leading-eight strategy is L_4, L_6, L_7, or L_8. Hence, we obtain

$$f^i(0,k;\ 0,k{+}1) = \begin{cases} \frac{N-k-1}{N}\frac{N-2}{N-1} & \text{if } i \in \{1,2,3,5\} \\ \frac{N-k-1}{N}\frac{k}{N-1} & \text{if } i \in \{4,6,7,8\}. \end{cases} \tag{2.17}$$

Transition $(0,k) \to (0,k{-}1)$. This case requires that a player $i>1$ is chosen to be the donor who player 1 considers as good. To become bad in player 1's eyes, this donor then either needs to defect against player 1, or he needs to cooperate against a receiver who is considered as bad by player 1 (provided that the applied leading-eight strategy is L_2, L_5, L_6, or L_8). The transition probability becomes

$$f^i(0,k;\ 0,k{-}1) = \begin{cases} \frac{k}{N}\frac{1}{N-1} & \text{if } i \in \{1,3,4,7\} \\ \frac{k}{N}\frac{N-k}{N-1} & \text{if } i \in \{2,5,6,8\}. \end{cases} \tag{2.18}$$

Transition $(0,k) \to (1,k)$. This case requires player 1 to be the donor, and that player 1 cooperates with her co-player. The corresponding probability is

$$f^i(0,k;\ 1,k) = \frac{1}{N}\frac{k}{N-1}. \tag{2.19}$$

All other transitions from (r,k) to (r',k') have transition probability $f^i(r,k;\ r',k') = 0$. We note that for this reduced Markov chain, the recovery state $(1, N-1)$ is absorbing. Once this state is reached, it cannot be left anymore, because $f^i(1,N-1;\ 1,N-2) = 0$ and $f^i(1,N-1;\ 0,N-1) = 0$ for all i. However, for populations that apply one of the four leading-eight strategies L_4, L_6, L_7, L_8, there exists another absorbing state, which is $(0,0)$. This state corresponds to a full segregation state: player 1 considers everyone else as bad, whereas all other players consider player 1 to be bad (**Fig. 2.5a**). Whether the population is able to recover from a single disagreement thus depends on how likely the full segregation state is reached. We obtain the following three cases.

Proposition 2 (Recovery probabilities)**.**
Suppose the population applies the leading-eight strategy L_i.

1. *For $i \in \{1,2,3,5\}$, the recovery probability is $\rho_i = 1$.*
2. *For $i \in \{4,7\}$, the recovery probability satisfies $1 - 2/(N-1)! \leq \rho_i < 1$.*
3. *For $i \in \{6,8\}$, the recovery probability is $\rho_i = 1 - 1/N$.*

With respect to the time it takes the population to recover from a single disagreement, we obtain a similar case distinction.

Proposition 3 (Expected time until recovery)**.**
Suppose the population applies the leading-eight strategy L_i.

1. *For $i \in \{1,3,4,7\}$, the expected time until recovery τ_i is of order $\Theta(N)$.*
2. *For $i \in \{2,5\}$, the expected time until recovery τ_i is of order $\Theta(N \log N)$.*
3. *For $i \in \{6,8\}$, the expected time until recovery is $\tau_i = N{\cdot}H_N - N$, where $H_N = \sum_{n=1}^{N} \frac{1}{n}$ is the N-th harmonic number.*

While some of the results in Proposition 3 only address the asymptotic behavior as the population becomes large, we can numerically compute exact solutions for the expected recovery time for all population sizes, based on the transition probabilities $f^i(r,k;\ r',k')$ specified above. The respective solutions are shown in **Fig. 2.5b**. In the following, we give some informal intuition for these results.

The above results suggest that with respect to recovery from single disagreements, there are four groups of leading-eight strategies. The strategies L_1 and L_3 are most robust: after a single disagreement, they are guaranteed to recover, and the expected recovery time is linear in the population size. These two strategies have in common that a cooperating donor is always perceived as good (even if the donor or the receiver had a bad reputation to start with). Due to this property, a single individual who is perceived as bad can immediately regain a good reputation once it is chosen to act as a donor (which takes an expected time of N rounds).

A similar observation holds for the two strategies L_4 and L_7: a single individual with bad reputation can always regain a good reputation by cooperating, independent of the identity of the receiver. Unlike the first two strategies, however, L_4 and L_7 can end up in the full segregation state. However, this requires a rather specific sequence of events: before player 2 can regain his good reputation in the eyes of player 1, it needs to be the case that player 1 is chosen to act as the donor with respect to player 2 as the recipient. In such a scenario, player 1 would defect, yielding her a bad reputation from the viewpoint of all other players. To end up in the full segregation state, all other players would then have to be chosen to act as a donor against player 1. As the population grows bigger, this sequence of events becomes increasingly unlikely, and the recovery probability quickly approaches 1.

Recovery is guaranteed for the two strategies L_2 and L_5, but in contrast to the first four strategies, recovery may take substantially longer. The longer recovery time is due to the property of these strategies that cooperation against an individual perceived as bad is itself perceived as bad. Thus, if player 2 is randomly chosen to be the recipient of the next indirect reciprocity interaction, the donor is guaranteed to obtain a bad reputation in the eyes of someone (either because the donor is an individual $i \geq 2$, who is then perceived as bad by player 1; or because the donor is player 1, who is then perceived as bad by everyone else).

The above recovery analysis confirms that L_6 and L_8 perform worst in the presence of single disagreements. Recovery is not guaranteed, and even if it occurs, it may take a substantial time until recovery. Both aspects are due to the relative ease with which these two strategies assign a bad reputation to other population members: a donor who cooperates against a receiver who is perceived as bad is guaranteed to end up with a bad reputation (independent of whether the donor was perceived as good or bad before). Under these conditions, bad reputations can proliferate comparably quickly.

2.5.2 Evolutionary dynamics

Description of the evolutionary process

While the previous section has assumed that the players' strategies are fixed, we now describe a simple process that allows us to explore which dynamics arises once players may change their strategies over time. To allow for a transparent treatment, we assume that strategy adaptation occurs on a time scale that is slow compared to the reputation dynamics.

Specifically, we assume that the strategies in the population change according to a simple imitation process [TNP06]. For this process, we again consider a population of N individuals where each individual i may have its own strategy $(\alpha^i;\ \beta^i)$. In each time step of the evolutionary process, one individual i is chosen at random to update her strategy. There are two ways to do so: with probability μ (corresponding to the mutation rate), she adopts a random strategy (with all remaining strategies having the same probability to be picked). With the remaining probability $1-\mu$, she seeks for a role model instead, by randomly choosing another individual j from the population. Suppose player i's payoff is given by $\hat{\pi}_i$ and j's payoff is $\hat{\pi}_j$, where $\hat{\pi}_i$ and $\hat{\pi}_j$ are given by the estimated payoffs according to Eq. [2.11]. We assume that the probability that i adopts j's strategy is given by the Fermi function [ST98],

$$P(\hat{\pi}_j, \hat{\pi}_i) = \frac{1}{1 + \exp\Big(- s(\hat{\pi}_j - \hat{\pi}_i)\Big)}. \tag{2.20}$$

The parameter $s \geq 0$ is called the strength of selection. It determines how much players value payoffs when adopting new strategies. For $s = 0$ we obtain $P(\hat{\pi}_j, \hat{\pi}_i) = 1/2$, and imitation occurs essentially at random. For $s > 0$, the imitation probability $P(\hat{\pi}_j, \hat{\pi}_i)$ is monotonically increasing in the payoff difference $\hat{\pi}_j - \hat{\pi}_i$. That is, the more there is to gain for player i, the more likely she is to abandon her old strategy and to imitate player j's strategy instead.

This elementary updating process, involving mutation and imitation, is then iterated over many time steps. As a result, we obtain an ergodic process on the space of all possible population compositions. This process has again a unique invariant distribution, to which we refer as the selection-mutation equilibrium. The abundance of each strategy in this equilibrium, and the corresponding average cooperation rate, can always be obtained by simulating the above process over a sufficiently long time span. However, in certain special cases, the selection-mutation equilibrium can be computed more efficiently. These more efficient algorithms either require a small or intermediate-sized population in which only a few different strategies are available, or that mutations are rare. In the following two sections we describe these algorithms in more detail.

Here, we have phrased the evolutionary process as an imitation dynamics. However, similar results can be obtained if one assumes that strategies are genetically encoded and inherited in a birth-death process [WBGT15], where the fitness of an individual is given by an exponential function of its payoffs, $\exp(s\hat{\pi}_i)$.

Selection-mutation equilibrium in populations with few strategies

Throughout the main text, we have focused on a population of intermediate size N that has access to at most $k=3$ different strategies. The possible states of the population are elements of the set

$$\Delta_N^k := \Big\{\mathbf{n} = (n_1, \ldots, n_k) \in \mathbb{N}^k \;\Big|\; \sum_{i=1}^{k} n_i = N\Big\}. \tag{2.21}$$

The entries n_i of each vector represent how many players currently apply strategy i. The number of such population compositions is

$$\left|\Delta_N^k\right| = \binom{N+k}{k}. \tag{2.22}$$

The evolutionary process described in Section 2.5.2 defines a Markov chain with state space Δ_N^k. For two population states $\mathbf{n}, \mathbf{n}' \in \Delta_N^k$, the transition probability to move from $\mathbf{n}$ to $\mathbf{n}'$ in one step of the process is given by

$$w_{\mathbf{n},\mathbf{n}'} = \begin{cases} \frac{n_i}{N}\left(\frac{\mu}{k-1} + (1-\mu)\frac{n_j}{N}P(\hat{\pi}_j,\hat{\pi}_i)\right) & \text{if } n_i'=n_i-1,\; n_j'=n_j+1,\; n_l'=n_l \text{ for } l \notin \{i,j\} \\ 1-\sum_{j\neq i}\frac{n_i}{N}\left(\frac{\mu}{k-1} + (1-\mu)\frac{n_j}{N}P(\hat{\pi}_j,\hat{\pi}_i)\right) & \text{if } \;\mathbf{n}=\mathbf{n}' \\ 0 & \text{otherwise.} \end{cases} \tag{2.23}$$

Provided that neither the population size N nor the number of strategies k is prohibitively large, the transition matrix $W=(w_{\mathbf{n},\mathbf{n}'})$ can be computed explicitly. By computing the normalized left eigenvector of W (with respect to eigenvalue 1), we obtain the selection-mutation equilibrium over an evolutionary timescale. We have implemented this algorithm using Matlab; the corresponding code is provided in the **Appendix**, Section 2.6.

Selection-mutation equilibrium in the limit of rare mutations

Alternatively, in arbitrarily large populations with an arbitrary number of available strategies, we can still calculate exact strategy abundances in the selection-mutation equilibrium if mutations are sufficiently rare [FI06, IFN05, WGWT12]. In that case, the population will find itself in a homogeneous state most of the time, in which all individuals adopt the same strategy i. Only occasionally, a mutant strategy j arises. This mutant then either reaches fixation in the population, or it goes extinct. The mutant's fixation probability p_{ij} can be calculated explicitly [NSTF04, TH09],

$$p_{ij} = \frac{1}{1 + \sum_{l'=1}^{N-1} \prod_{l=1}^{l'} \exp\Big(- s(\hat{\pi}_j(l) - \hat{\pi}_i(l))\Big)}. \tag{2.24}$$

Here, $\hat{\pi}_j(l)$ and $\hat{\pi}_i(l)$ refer to the payoffs of a mutant and a resident, provided that the number of mutants in the population is l. Using Eq. [2.24], the selection-mutation equilibrium of the evolutionary process can be computed by considering a reduced Markov chain with k states (corresponding to the homogeneous populations in which every player applies the same strategy). The probability to move from state i to state j is given by $p_{ij}/(k-1)$. The invariant distribution of this reduced Markov chain approximates the invariant distribution of the evolutionary process as μ becomes small [FI06].

2.6 Appendix: Proof of the recovery analysis

Proof of Proposition 1. For the proof, we consider the Markov chain $H = (h_{M,M'})$, as defined in Section 2.5.1, for the limiting case $\varepsilon = 0$ and $q = 1$. We show that the set $\mathcal{M}$ of image matrices that satisfy the three properties in [2.13] is invariant. That is, let $M \in \mathcal{M}$ be arbitrary and suppose $h_{M,M'} > 0$ for some image matrix M'. Then also $M' = (m'_{ij})$ satisfies the three characteristic properties [2.13].

(*i*) $m'_{ii} = 1$ for all i. Since $M \in \mathcal{M}$, initially all players consider themselves as good. All leading-eight strategies have the property that the strategy's action rule prescribes an action that lets a good donor maintain her good reputation in her own eyes, independent of which reputation she assigns to the recipient. Thus all players keep considering themselves as good after one interaction; either they do not need to make a decision (because they were not chosen to act as the donor), or they choose an action they themselves evaluate as good.

(*ii*) and (*iii*) $m'_{ij} = 1$ and $m'_{i1} = m'_{j1}$ for all $i, j \geq 2$. Since $M \in \mathcal{M}$, all players $i, j \geq 2$ initially agree on the reputations of all population members. Because they all apply the same assessment rule and observation errors are excluded, they also agree on how the donor's action in the subsequent interaction needs to be assessed. This shows $m'_{il} = m'_{jl}$ for all $i, j \geq 2$ and all l. Moreover, since all players $i, j \geq 2$ consider each other as good initially, and since their common action rule only lets them choose actions that let them keep their good reputation, we conclude $m'_{ij} = 1$ for $i, j \geq 2$.

□

Proof of Propositions 2 and 3. In the following, we provide the proofs of Propositions 2 and 3 by considering each of the four different cases {L1, L3}, {L2, L5}, {L4, L7}, {L6, L8} individually. For a given leading-eight strategy L_i, we consider the respective Markov chain M_i with $2N$ states $s_{r,k}$. The value of $r \in \{0, 1\}$ corresponds to the reputation

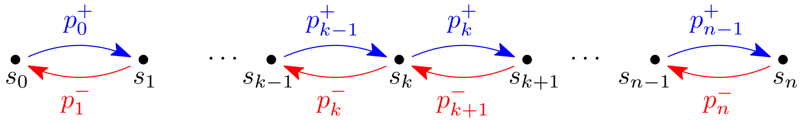

Figure 2.14: **One-dimensional random walk.** A discrete-time Markov chain with $n+1$ states labelled $s_0 \dots, s_n$ arranged in a line and with transition probabilities $p_k^+ : s_k \to s_{k+1}$ and $p_k^- : s_k \to s_{k-1}$.

of player 1 from the perspective of all other players. The value of $k \in \{0, 1, \dots, N-1\}$ corresponds to the number of co-players that player 1 considers as good. Then the recovery probability ρ_i is the probability that a random realization of the process (a *random trace*) starting in a state $s_{1,N-2}$ reaches the absorbing state $s_{1,N-1}$. The expected recovery time τ_i is the expected number of steps until a random trace starting in a state $s_{1,N-2}$ reaches the state $s_{1,N-1}$, conditioned on it doing so. The main idea of the proof is to make use of a coupling argument. That is, to each two-dimensional random walk defined by the Markov chain M_i, we associate a simpler, one-dimensional random walk. For the one-dimensional random walk, the absorption probabilities and conditional absorption times can be calculated explicitly, and they can serve as upper (or lower) bounds for the respective quantities in the two-dimensional random walk. To this end, let us recall the following result.

Proposition 4 (One-dimensional random walk, see Chapter 7.7, proof of Theorem 7.1 in [KT75])**.**
Let M be a discrete-time Markov chain with $n+1$ states $s_0, \dots, s_n$ and transition probabilities $p_i^+ : s_i \to s_{i+1}$ $(i = 0, \dots, n-1)$ and $p_i^- : s_i \to s_{i-1}$ $(i = 1, \dots, n)$, as illustrated in Fig. 2.14.

1. *If both states s_0 and s_n are absorbing (i.e. $p_0^+ = p_n^- = 0$) then the probability $\rho(k)$ of reaching s_n before reaching s_0 when starting at s_k is given by*
$$\rho(k) = \frac{1 + r_1 + r_1 r_2 + \dots + r_1 \dots r_{k-1}}{1 + r_1 + r_1 r_2 + \dots + r_1 \dots r_{n-1}}, \tag{2.25}$$
where $r_i = p_i^- / p_i^+$ for each $i = 1, \dots, n-1$.

2. *If there is no absorbing state except, possibly, s_0 then the expected number of time steps $t_{k,k-1}$ to reach state s_{k-1} from state s_k is given by*
$$t_{k,k-1} = \frac{1}{p_k^-} + \frac{p_k^+}{p_{k+1}^- p_k^-} + \dots + \frac{p_k^+ \dots p_{n-1}^+}{p_n^- \dots p_k^-} = \frac{1}{p_k^-} \left(1 + \sum_{i=1}^{n-k} \prod_{j=1}^{i} \frac{p_{k+j-1}^+}{p_{k+j}^-} \right). \tag{2.26}$$
Moreover, the expected number of time steps $t_{k,l}$ to reach state s_l from state s_k, with $l < k$, is
$$t_{k,l} = t_{k,k-1} + t_{k-1,k-2} + \dots + t_{l+1,l} = \sum_{i=l+1}^{k} t_{i,i-1}. \tag{2.27}$$

We use standard notation $o(\cdot)$ and $\Theta(\cdot)$ for strict asymptotic upper bound and for asymptotically tight bound, ignoring the constant factors. Hence, for example, we have $1/n = o(1)$ and $2n+1 = \Theta(n)$, because for large n we have $1/n \ll 1$ whereas $\frac{2n+1}{n}$ tends to a constant. See Section 1.3 of [Cor09] for a detailed treatment. For asymptotic results, we assume $N \geq 4$.

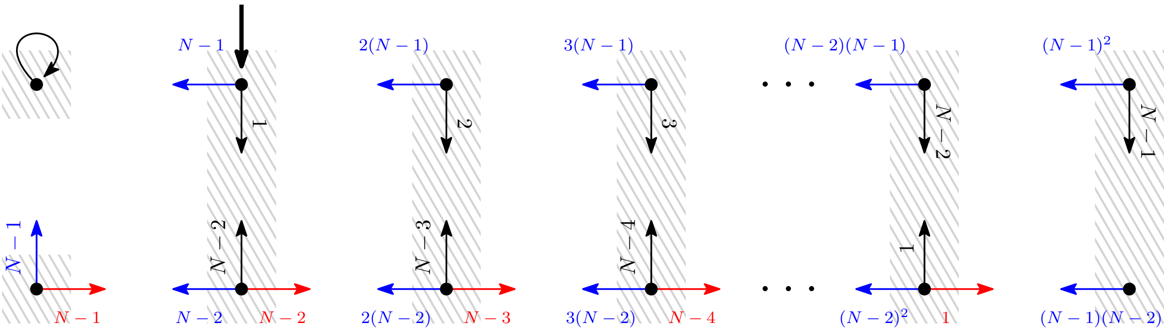

Figure 2.15: Markov chain $M_1 = M_3$. The transition probabilities are normalized by $N(N-1)$. Self-loops in non-absorbing states are not shown.

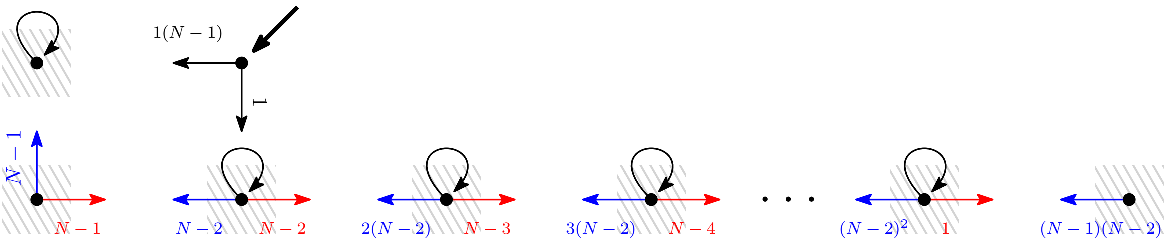

Figure 2.16: Markov chain M_1'. The transition probabilities are normalized by $N(N-1)$. By coupling, $\tau_1' \geq \tau_1$.

Recovery analysis for $L_1 = L_3$. The Markov chain $M_1 = M_3$ is depicted in Fig. 2.15.

- Recovery probability: M_1 has a unique absorbing state $s_{1,N-1}$. Moreover, starting from any other state, there is a positive probability to reach the state $s_{1,N-1}$. Hence, it follows that $\rho_1 = 1$.

- Recovery time: To get a lower bound on the recovery time, we note that reaching $s_{1,N-1}$ requires at least one non-self-looping transition. Starting from the state $s_{1,N-2}$, the expected time until the first non-self-looping transition occurs is $\frac{N(N-1)}{N-1+1} = N-1$. Hence $\tau_1 \geq N-1$.

In the following, we show $\tau_1 \leq N+7$, which establishes the desired $\tau_1 = \Theta(N)$. To this end, consider the Markov chain M_1' obtained from M_1 by erasing states $s_{1,k}$ for $k \leq N-2$, and by replacing transitions $s_{0,k} \to s_{1,k}$ with self-loops (see Fig. 2.16).

First, we argue that $\tau_1 \leq \tau_1'$, where τ_1' is the expected number of steps to reach $s_{1,N-1}$ from $s_{1,N-2}$ in M_1'. To this end, consider an arbitrary trace T in M_1. If T never takes any transition $s_{0,k} \to s_{1,k}$, for $k < N-1$, then we can associate the identical trace T' in M_1' to T. Otherwise, if there is a moment when T moves from $s_{0,k}$ to $s_{1,k}$ for some $k < N-1$, the associated trace T' has a self-loop at $s_{0,k}$. Since for each $k < N-1$ the transition $s_{1,k} \to s_{1,k+1}$ has higher probability than the transition $s_{0,k} \to s_{0,k+1}$, we can then couple traces T and T' such that, from that point on, T is not to the right of or below T' (formally, at any time after that moment, the k-coordinate of T is larger or equal to the k-coordinate of T', and the r-coordinate of T is larger or equal to the r-coordinate of T'). In particular, if T' has reached $s_{1,N-1}$ then so did T, and the inequality $\tau_1 \leq \tau_1'$ follows.

It remains to prove that $\tau_1' \leq N+7$. We use Proposition 4. Let x be the expected number of steps to reach $s_{1,N\text{-}1}$ from $s_{0,N\text{-}2}$ in M_1'. Then $\tau_1' = (N-1) + \frac{N-1}{N} \cdot 0 + \frac{1}{N} \cdot x$. By

Eq. [2.27], x takes the form $x = t_{2,1} + t_{1,0}$. The quantities $t_{2,1}$ and $t_{1,0}$ are calculated using Eq. [2.26]. For the first quantity, we get

$$\begin{aligned} t_{2,1} &= N(N-1)\left(\frac{1}{N-2} + \frac{N-2}{2!(N-2)^2} + \frac{(N-2)(N-3)}{3!(N-2)^3} + \cdots + \frac{(N-2)!}{(N-1)!(N-2)^{N-1}}\right) \\ &\leq \frac{N(N-1)}{N-2}\left(1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{(N-1)!}\right) \leq \frac{N(N-1)}{N-2} \cdot (e-1), \end{aligned}$$

where we have made use of the definition of Euler's number, $e := \sum_{i=0}^{\infty} 1/i!$. For the second quantity, we similarly obtain

$$\begin{aligned} t_{1,0} &= N(N-1)\left(\frac{1}{N-1} + \frac{N-1}{(N-1)(N-2)} + \frac{(N-1)(N-2)}{(N-1)2!(N-2)^2} + \cdots + \frac{(N-2)!}{(N-1)(N-1)!(N-2)^{N-1}}\right) \\ &\leq \frac{N(N-1)}{N-2}\left(1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{(N-1)!}\right) \leq \frac{N(N-1)}{N-2} \cdot e \end{aligned}$$

Hence $x = t_{2,1} + t_{1,0} \leq \frac{N(N-1)}{N-2} \cdot (2e-1)$ and, for $N \geq 4$,

$$\tau_1' = N - 1 + \frac{1}{N} \cdot x \leq N - 1 + \frac{N-1}{N-2} \cdot (2e-1) = N + 2(e-1) + \frac{2e-1}{N-2} < N + 7,$$

as desired.

Recovery analysis for $L_2 = L_5$. The Markov chain $M_2 = M_5$ is depicted in Fig. 2.17.

- Recovery probability: As before, since there is a unique absorbing state that can be reached from any initial state, $\rho_2 = 1$.

- Recovery time: To show that the expected number of steps τ_2 is of the order $\Theta(N \cdot \log N)$, we consider two simpler Markov chains M_2^+ (resp. M_2^-) with expected number of steps τ_2^+ (resp. τ_2^-). We then prove that $\tau_2^+ \geq \tau_2 \geq \tau_2^-$ and that both τ_2^+ and τ_2^- are of the order $\Theta(N \cdot \log N)$.

Intuitively, M_2^+ is obtained by identifying states $s_{1,k}$ and $s_{0,k+1}$ for each $k < N-1$ (that is, we identify all states that have the same number of individuals that are universally considered as good). For the transitions of M_2^+, we take the transition probabilities as in Fig. 2.18(a). In doing so, we decrease the total probability to move either to the left or upwards in each state, while leaving the total probability to move to the right or downwards unchanged. As a consequence, any trace T^+ in M_2^+ can be coupled with a trace T in M_2 such that if T^+ is in a state s_{r^+,k^+} at time t, and T is in a state $s_{r,k}$, then $r^+ + k^+ \leq r + k$. Hence, if T^+ has reached $s_{1,N-1}$ then so did T and the inequality $\tau_2^+ \geq \tau_2$ follows. For M_2^- we proceed analogously: we use the same identification of states, but this time the probability to move left in Fig. 2.18(b) is larger or equal to the corresponding probability to move left or up in Fig. 2.17. As a consequence, $\tau_2 \geq \tau_2^-$.

Finally, we compute τ_2^+, τ_2^-. By Eq. [2.26], we obtain

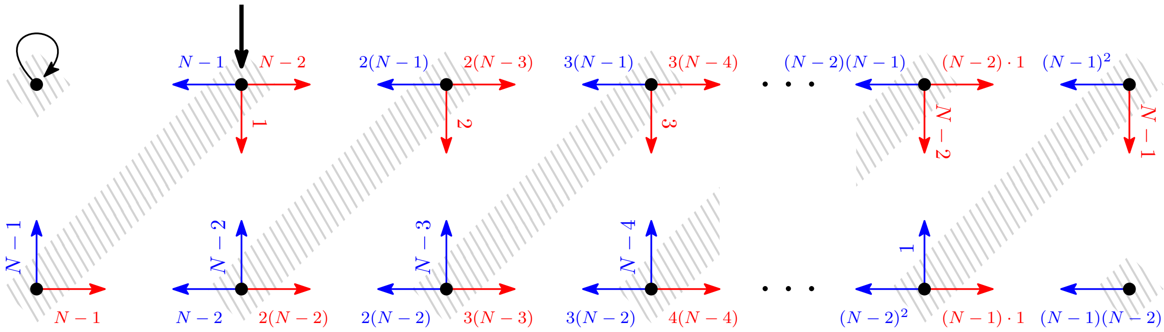

Figure 2.17: Markov chain $M_2 = M_5$. The transition probabilities are normalized by $N(N-1)$. Self-loops in non-absorbing states are not shown.

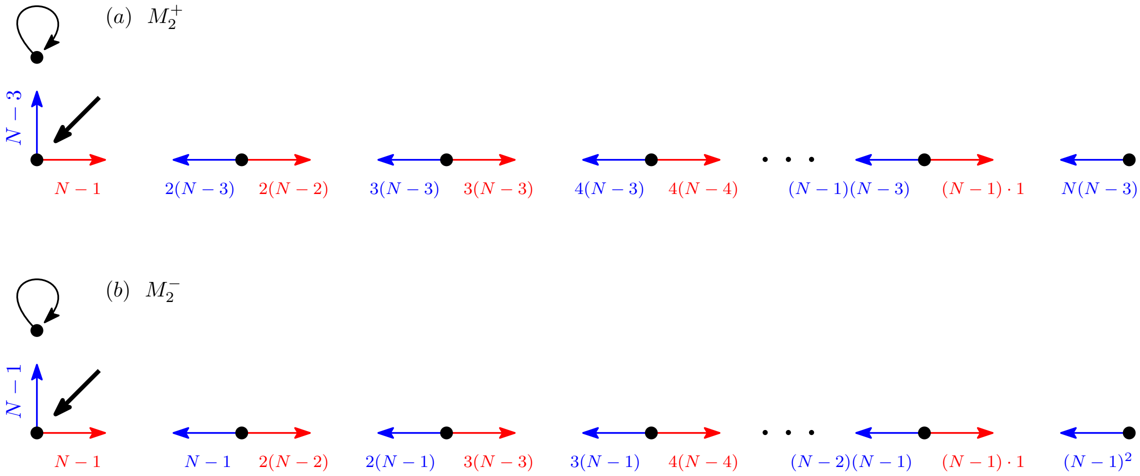

Figure 2.18: Markov chains M_2^+ and M_2^-. The transition probabilities are normalized by $N(N-1)$. By coupling, $\tau_2^+ \geq \tau_2 \geq \tau_2^-$.

$$\begin{aligned}
\tau_2^+ &= N(N-1)\left(\frac{1}{N-3} + \frac{N-1}{2!(N-3)^2} + \frac{1\cdot 2\cdot (N-1)(N-2)}{3!(N-3)^3} + \cdots + \frac{1\cdot 2\cdots (N-1)\cdot (N-1)(N-2)\cdots 1}{N!(N-3)^N}\right. \\
&= \frac{N(N-1)}{N-3}\left(1 + \frac{1}{2}\cdot\frac{N-1}{N-3} + \frac{1}{3}\cdot\frac{(N-1)(N-2)}{(N-3)^2} + \cdots + \frac{1}{N}\cdot\frac{(N-1)!}{(N-3)^{N-1}}\right) \\
&\leq \frac{N(N-1)^2(N-2)}{(N-3)^3}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}\right) = \Theta(N\log N),
\end{aligned}$$

$$\begin{aligned}
\tau_2^- &= N(N-1)\left(\frac{1}{N-1} + \frac{N-1}{2!(N-1)^2} + \frac{1\cdot 2\cdot (N-1)(N-2)}{3!(N-1)^3} + \cdots + \frac{1\cdot 2\cdots (N-1)\cdot (N-1)(N-2)\cdots 1}{N!(N-1)^N}\right. \\
&\geq N\left(1 + \frac{1}{2}\cdot\frac{N-1}{N-1} + \frac{1}{3}\cdot\frac{(N-1)(N-2)}{(N-1)^2} + \cdots + \frac{1}{\sqrt{N}}\cdot\frac{(N-1)\dots(N-\sqrt{N})}{(N-1)^{\sqrt{N}}}\right) \\
&\geq N\left(1 + \frac{1}{2}\frac{N-1}{N-1} + \frac{1}{3}\frac{N-3}{N-1} + \cdots + \frac{1}{\sqrt{N}}\frac{N-\frac{1}{2}N}{N-1}\right) \geq N\cdot\frac{1}{2}\cdot H_{\sqrt{N}} \geq \frac{1}{4}N\log N = \Theta(N\log N),
\end{aligned}$$

where we only took the sum of the first $\sqrt{N}$ terms and used that $1+2+\cdots+\sqrt{N} \approx \frac{1}{2}N$ and that $\log\sqrt{N} = \frac{1}{2}\log(N)$.

Recovery analysis for $L_4 = L_7$. The Markov chain $M_4 = M_7$ is depicted in Fig. 2.19.

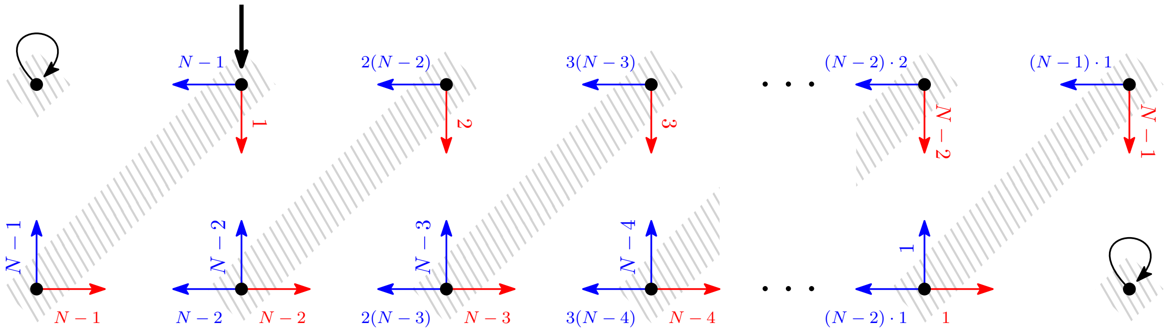

Figure 2.19: Markov chain $M_4 = M_7$. The transition probabilities are normalized by $N(N-1)$. Self-loops in non-absorbing states are not shown.

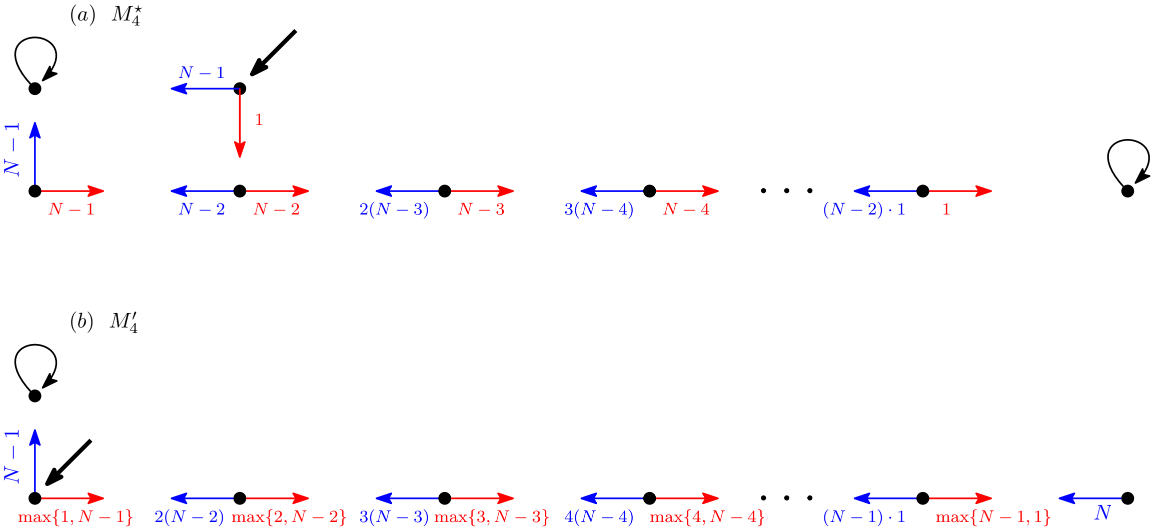

Figure 2.20: Markov chains $M_4^\star$ and M_4'. The transition probabilities are normalized by $N(N-1)$. By coupling, $\tau_4 \leq \tau_4'$.

- Recovery probability: M_4 has two absorbing states, $s_{0,0}$ and $s_{1,N-1}$, and both states can be reached with positive probability from any other state. Nevertheless, we show in the following that ρ_4 quickly approaches 1 as the population size increases, because $\rho_4 \geq 1 - 2/(N-1)!$.

To obtain this lower bound for ρ_4, we consider the Markov chain $M_4^\star$ that is obtained from M_4 by erasing states $s_{1,k}$ for $k < N-1$, and replacing the corresponding transitions of the form $s_{0,k} \to s_{1,k}$ by self-loops. Formally, we define $M_4^\star$ to be as in Fig. 2.20(a). For the coupling, note that any time a trace T in M_4 takes a transition of the form $s_{0,k} \to s_{1,k}$, the corresponding trace $T^\star$ in $M_4^\star$ waits until (if ever) T reaches state $s_{r,k}$ with $r = 0$ again. At that point, T is to the left of $T^\star$ or in the same state. Hence if $T^\star$ ever reaches $s_{1,N-1}$ then so did T and we have $\rho_4 \geq \rho_4'$.

Denoting by x the probability of reaching $s_{0,0}$ from $s_{0,N-2}$ (before reaching $s_{1,N-1}$), we have $1 - \rho_4' = \frac{N-1}{N} \cdot 0 + \frac{1}{N} \cdot x = x/N$ and by Proposition 4 we compute

$$x = \frac{1+1}{1+1+1!+2!+3!+\cdots+(N-2)!} \leq \frac{2}{(N-2)!}$$

implying that $\rho_4 \geq \rho_4' = 1 - x/N \geq 1 - 2/(N-1)!$ as desired.

- Recovery time: since reaching $s_{1,N-1}$ requires taking at least one non-self-looping transition, we immediately get $\tau_4 \geq \frac{N(N-1)}{N-1+1} = N - 1$. In the following, we focus on the upper bound. To that end, we define a simpler Markov chain M_4'. Intuitively, M_4' is obtained by identifying states $s_{1,k}$ and $s_{0,k+1}$ for each $k < N-2$. Since, for each $k = 1, \dots, N-2$, the Markov chain M_4 has the property that

$$\mathbb{P}[s_{1,k-1} \to s_{1,k}] = \mathbb{P}[s_{0,k} \to s_{1,k}] + \mathbb{P}[s_{0,k} \to s_{0,k+1}],$$

the transition probabilities to the left are equal for each pair of identified states. To get an upper bound for τ_4, for the transition probabilities to the right we take the larger of $\mathbb{P}[s_{1,k} \to s_{0,k}]$ and $\mathbb{P}[s_{0,k} \to s_{0,k-1}]$. Formally, we define M_4' as in Fig. 2.20(b). Note that we have defined $\mathbb{P}[s_{0,0} \to s_{0,1}] = N$, such that M_4' has only one absorbing state.
Next let us show that $\tau_4 \le \tau_4'$. By construction, this is clear for traces that don't reach state $s_{0,0}$. For those that do, we need to show that the expected number s of steps to reach $s_{1,N-1}$ from $s_{0,0}$ in M_4' satisfies $s \ge \tau_4$. This follows from $\mathbb{P}[s_{0,0} \to s_{0,1}]$ in M_4' is equal to $\mathbb{P}[s_{1,N-2} \to s_{1,N-1}] + \mathbb{P}[s_{1,N-2} \to s_{0,N-2}]$ in M_4: once the trace in M_4' transitioned from $s_{0,0}$, the coupled trace in M_4 that started at $s_{1,N-2}$ either converged directly or transitioned to $s_{0,N-2}$ which is to the left of $s_{0,0}$.
Finally, we compute τ_4' using Eq. [2.26] again:

$$\begin{aligned}
\tau_4' &= N(N-1)\left(\frac{1}{N-1} + \frac{1}{2!(N-2)} + \frac{1}{3!(N-3)} + \dots + \frac{1}{(N/2)!(N-N/2)} + \underbrace{X_1 + X_2 + \dots + X_{N/2}}_{X_i \le \frac{1}{(N/2)!}}\right) \\
&\le \frac{N(N-1)}{N/2}\left(1 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{(N/2)!}\right) + \underbrace{N(N-1) \cdot \frac{N}{2} \cdot \frac{1}{(N/2)!}}_{o(1)} \\
&\le 2(N-1) \cdot (e-1) + o(1) = \Theta(N).
\end{aligned}$$

where we used that each of the last $N/2$ terms in the first line is less than $1/(N/2)!$ and $\sum_{i=0}^{\infty} 1/i! = e$.

Recovery analysis for $L_6 = L_8$. The Markov chain $M_6 = M_8$ is depicted in Fig. 2.21.

- Recovery probability: Identifying the states $s_{1,k}$ and $s_{0,k+1}$ for $k < N-1$, we obtain an equivalent Markov chain M_6' depicted in Fig. 2.22. Proposition 4 applied to M_6' immediately implies

$$\rho_6 = 1 - \frac{1}{\underbrace{1 + 1 + \dots + 1}_{N\times}} = 1 - \frac{1}{N}.$$

- Recovery time: For M_6' we observe that $p_k^+ = p_k^- = \frac{k(N-k)}{N(N-1)}$. Thus, up to rescaling, the Markov chain M_6' is equivalent to the Moran process under neutral drift [see e.g. AT09, Section 4]. Denoting by t^a the expected time until the process in M_6' reaches an absorbing state, and by t^b the conditional time given that it reaches the non-recovering state $s_{0,0}$, equations [18] and [19] from [AT09] rescale to

$$t^a = (N-1) \cdot H_{N-1} \qquad \text{and} \qquad t^b = (N-1)^2.$$

The probability of absorbing at $s_{1,N-1}$ is $\rho_6 = 1 - \frac{1}{N}$. Therefore, $t^a = (1 - \frac{1}{N}) \cdot \tau_6 + \frac{1}{N} \cdot t^b$, or

$$\tau_6 = \frac{N \cdot t^a - t^b}{N-1} = N \cdot H_{N-1} - (N-1) = N \cdot H_N - N,$$

as desired.

□

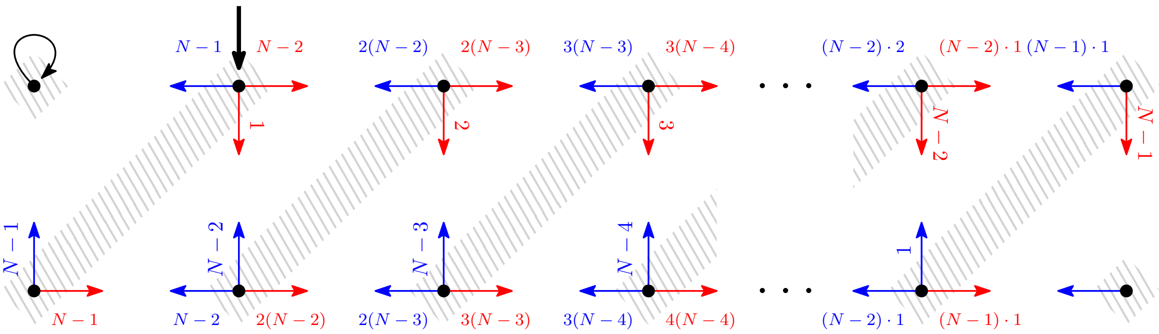

Figure 2.21: Markov chain $M_6 = M_8$. The transition probabilities are normalized by $N(N-1)$. Self-loops in non-absorbing states are not shown.

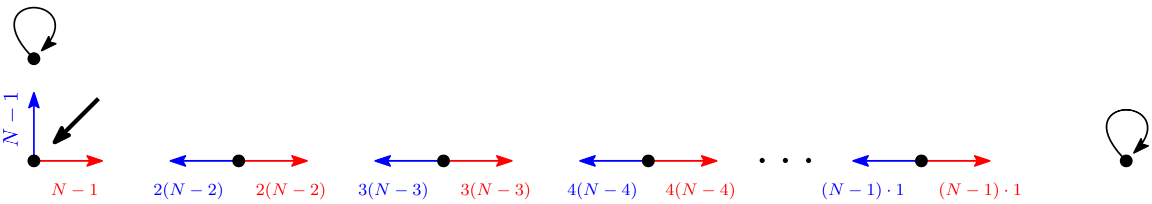

Figure 2.22: Markov chain M'_6 that is equivalent to $M_6 = M_8$. The transition probabilities are normalized by $N(N-1)$. We have $\rho_6 = \rho'_6$ and $\tau_6 = \tau'_6$.

CHAPTER 3

The evolution of indirect reciprocity under action and assessment generosity

Indirect reciprocity is a mechanism for the evolution of cooperation based on social norms. This mechanism requires that individuals in a population observe and judge each other's behaviors. Individuals with a good reputation are more likely to receive help from others. Previous work suggests that indirect reciprocity is only effective when all relevant information is reliable and publicly available. Otherwise, individuals may disagree on how to assess others, even if they all apply the same social norm. Such disagreements can lead to a breakdown of cooperation. Here we explore whether the predominantly studied 'leading eight' social norms of indirect reciprocity can be made more robust by equipping them with an element of generosity. To this end, we distinguish between two kinds of generosity. According to assessment generosity, individuals occasionally assign a good reputation to group members who would usually be regarded as bad. According to action generosity, individuals occasionally cooperate with group members with whom they would usually defect. Using individual-based simulations, we show that the two kinds of generosity have a very different effect on the resulting reputation dynamics. Assessment generosity tends to add to the overall noise and allows defectors to invade. In contrast, a limited amount of action generosity can be beneficial in a few cases. However, even when action generosity is beneficial, the respective simulations do not result in full cooperation. Our results suggest that while generosity can favor cooperation when individuals use the most simple strategies of reciprocity, it is disadvantageous when individuals use more complex social norms.

3.1 Introduction

People tend to put in time and effort to achieve and maintain a good standing with those that surround them [Ale87, JHTM11, Feh04, BKO05]. This human concern for a good reputation can be an important motivation to engage in cooperative behavior: Once somebody's reputation is at stake, people have more of an incentive to engage in costly acts of helping others [WB02, Sug86, SKM04, Kan92, MSK02, CGLF$^+$15]. This reputation-based mechanism for the evolution of cooperation is referred to as indirect reciprocity [NS05, Sig10, Sig12, Oka20]. In contrast to direct reciprocity [AH81, HCN18, GvV18, GK21], which is based on repeated interactions between the same players, indirect reciprocity does not require any two individuals to interact with one another more than once. Instead it only requires population members to continually assess each others' actions, and to act based on these reputations. Whereas direct reciprocity is based on individual memories of interactions with a given group member, indirect reciprocity builds upon a group's collective memories.

In order to analyze the process by which reputations evolve, the literature on indirect reciprocity considers certain social dilemma situations [NS05, Sig10, Sig12, Oka20]. In the simplest case, each dilemma situation involves only two population members, who are referred to as 'donor' and 'recipient', respectively. The donor is asked whether or not to pay a personal cost to provide help to the recipient. The two possible actions of either offering help or refusing help are interpreted as cooperation and defection, respectively. While the recipient does not make any decisions, the donor's action is observed by other population members. These observers then decide how the donor's reputation should be updated in light of their behavior.

How observers update reputations, and how donors decide whom to help, depends on the social norm applied in the population. These social norms consist of two components. First, the social norm's assessment rule specifies which behaviors of the donor should improve the donor's reputation, and which behaviors should be condemned. Second, the social norm's action rule tells the donor whether or not to help the recipient; this decision may in turn depend on the recipient's and the donor's current reputation. A well-known example of a social norm is 'Image Scoring' [NS98b, NS98a, Ber11]. According to the assessment rule of image scoring, a donor's reputation should improve every time the donor cooperates, and it should deteriorate every time the donor defects. According to its action rule, a donor should only cooperate with those recipients who are sufficiently well-reputed. Importantly, the social norm of Image Scoring has the property that assessments only depend on whether or not the donor cooperated. Social norms with this property are sometimes referred to as 'first-order' norms. While this basic principle of first-order norms may appear intuitive, Image Scoring has been shown to be unstable [PB03, BS04, LH01]. The reason for this instability is that individuals are required to defect against an ill-reputed group member; yet by doing so, they harm their own reputation. To overcome this inconsistency, it has been argued that stable norms of indirect reciprocity need to be sufficiently complex to differentiate between justified and unjustified acts of defection [PB03, BS04].

To identify such stable cooperative norms, the landmark papers by Ohtsuki and Iwasa [OI04, OI06] consider an even simpler decision situation. In their setup, reputations are required to be binary, such that individuals can either be 'good' or 'bad'. For this binary model of reputations, they allow for up to all 'third-order' assessment rules. In addition to the donor's action, third-order assessments may also depend on the current reputation of the donor and the recipient. For example, while helping a good recipient may be assessed as good, the very same action towards a bad recipient may be assessed as bad. With an exhaustive search on the space of all third-order norms, Ohtsuki and Iwasa identified eight successful norms that can maintain cooperation, which they coined the "leading eight". All leading eight norms agree that a cooperative action towards a good recipient should yield a good reputation, whereas a defecting action towards a good recipient should yield a bad reputation. The norms differ, however, in how they evaluate interactions with a bad recipient. While some norms find it acceptable if help is provided to a bad recipient, other norms like 'Stern Judging' [PSC06] condemn such behaviors completely.

The findings of Ohtsuki and Iwasa have been hugely influential for the further development of the field [BS05, OIN09, PSC06, SON17, SSP18, XGH19], yet they are based on a number of important assumptions. One of their key assumptions is that reputations are assigned publicly. This means that after any interaction between a donor and recipient, it is a central authority that assigns an updated reputation to the donor; this publicly assigned reputation is then adopted by the entire population. As a consequence, the views of different individuals are always perfectly correlated: if a given group member is considered to be good by one individual, they are also considered to be good by anyone else. The assumption of public information greatly facilitates the mathematical analysis of indirect reciprocity. However, it cannot capture scenarios in which individuals make their own judgments, based on their own private information. When individuals make their own judgments, they may start to disagree about which reputation they assign to a given group member. Such disagreements can arise, for example, when individuals differ in which interactions they observe, or when they occasionally misinterpret a given interaction. Once such disagreements arise, they can further proliferate, because individuals now also perceive all future interactions of that group member differently [Uch10, US13, OSN17]. For private, scarce, and incomplete information, individual-based simulations thus suggest that the leading-eight social norms no longer effectively promote the evolution of cooperation[HST$^+$18].

This finding naturally calls for mechanisms to mitigate the effect of noisy environments and private information on indirect reciprocity [KH21, RSP19]. Here we explore the effect of a particular mechanism, generosity. The value of generosity has previously been stressed in the literature on direct reciprocity [Mol85, NS92, MVCS12, SP13, HWTN14]. For example, by occasionally cooperating against defectors, the strategy 'Generous Tit-for-Tat' can stabilize full cooperation [Mol85, NS92], even though the classical Tit-for-Tat strategy cannot. Similarly, the analogous first-order social norm 'Generous Scoring' has been shown to be stable in the context of indirect reciprocity [SCHN21], even though classical Image Scoring is not. In both cases, there is the same intuition for why a certain degree of generosity is favorable. By becoming more generous, populations of reciprocators are more likely to sustain cooperation among themselves even in noisy environments. This in turn makes them more robust against invasion by unconditional cooperators. At the same time, however, reciprocators must not be too forgiving, for otherwise they can be invaded by unconditional defectors. The optimal degree of generosity thus needs to strike a balance between being sufficiently forgiving to correct errors, and being sufficiently strict to avoid exploitation.

In the following, we propose a framework to incorporate an element of generosity into the leading eight social norms. To this end, we distinguish between two kinds of generosity. The first, assessment generosity, makes individuals more generous when assigning reputations to other group members. In situations in which they would usually assign a bad reputation to a given group member, they may occasionally assign a good reputation instead. The second, action generosity, makes individuals more generous in their decisions whom to help as a donor. In situations in which they would usually defect, they may occasionally cooperate instead. Although these two kinds of generosity may seem similar, our results suggest that their effect on the resulting reputation dynamics is very different. In particular, while action generosity can sometimes enhance cooperation, assessment generosity is always detrimental. Moreover, even under action generosity, we do not observe the evolution of high cooperation rates when environments are noisy. Our results suggest that complex social norms of indirect reciprocity are not compatible with individual acts of generosity. Unless generosity is the result of a coordinated effort among all population members, generosity merely acts as another seed of disagreement.

3.2 Results

3.2.1 A theoretical framework of action and assessment generosity

To incorporate generosity into higher-order social norms of indirect reciprocity, we consider a well-mixed population of fixed size N. In every round, two players are randomly chosen to engage in an interaction. One of the players is randomly assigned to be the donor, whereas the other player is assigned to be the recipient. The donor chooses whether to confer a benefit b to the recipient, at an own cost of c, with $0<c<b$. Other members of the population independently observe the donor's decision with probability q. We assume that information is private: every observer keeps track of others' reputations by updating their personal reputation repository, and there is no shared judgment of observed actions (see **Methods**, Section 3.4). Players thus individually update and keep track of their opinion about others based on their observations. These observations are subject to noise: observers can misperceive an action with a probability ε, and mistakenly interpret e.g. cooperation as defection.

How observers update others' reputation based on what they observe, and how they subsequently act towards others, is governed by their social norm. We broadly interpret social norms as rules that tell individuals how they should behave in social interactions. In our context, we consider norms that consist of two components: an assessment rule and an action rule [Oka20]. The assessment rule prescribes how reputations are updated, based on the observed actions and the possible context of an interaction. The action rule governs a player's behavior when it is their turn to decide whether to confer a benefit to a co-player. Following Ohtsuki and Iwasa [OI04, OI06], we assume that reputations are binary (good or bad), and social norms are at most third order. To allow for generosity, we consider modified versions of the deterministic leading eight social norms $L1 - L8$ (see **Figure 3.1a**). Players with these modified versions always cooperate if the original leading eight prescribe to do so. Similarly, they always assign a good reputation in cases where the original version would. However, in cases where the original social norm would assign a bad reputation, the modified version instead assigns a good reputation with probability g_1 (**Figure 3.1b,c**). Analogously, in cases where the original leading eight norm prescribes defection, the modified version instead cooperates with probability g_2 (**Figure 3.1d,e**). We refer to g_1 and g_2 as the probabilities that players engage in assessment generosity and action generosity, respectively. In the limiting case with $g_1=g_2=0$, we recover the original leading eight.

We note that our framework makes two assumptions about the stochasticity we introduce. For one, players are only forgiving, but not spiteful: players never defect when the original leading eight rule prescribes cooperation, and they never assign a bad reputation to a player they are supposed to regard as good. This means that only when the original norm prescribes a negative assessment or defection, there is a probability to positively assess (g_1) or cooperate (g_2) instead. Another simplifying assumption is that players use

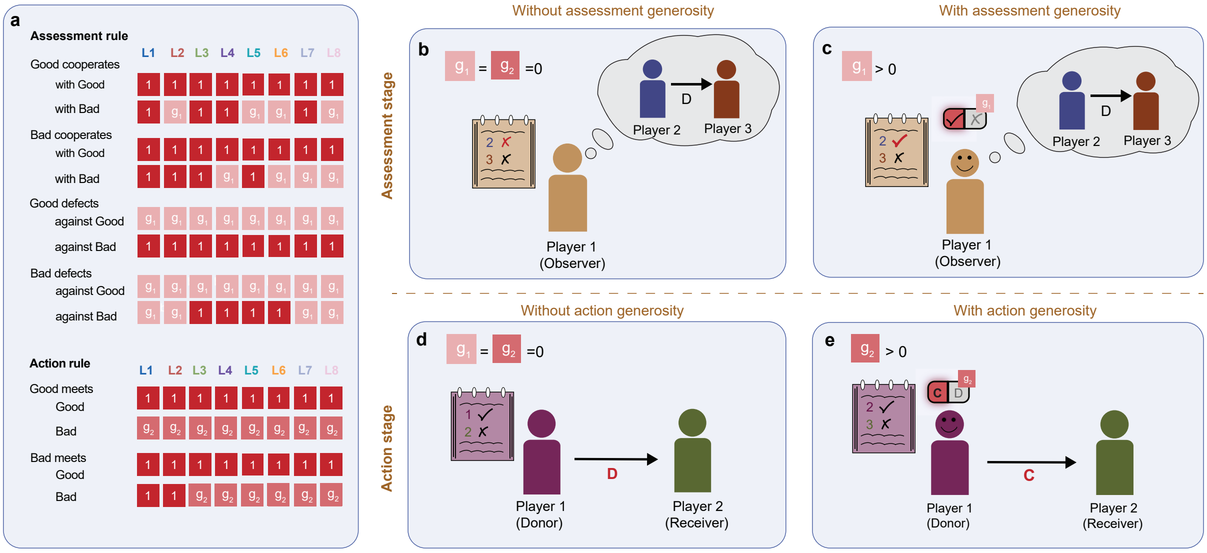

Figure 3.1: **The leading eight social norms with assessment and action generosity.** **a,** To sustain cooperation based on indirect reciprocity, Ohtsuki and Iwasa[OI06] suggested a set of eight social norms. Each social norm consists of an assessment rule and an action rule. The assessment rule determines with which probability observers assign a good reputation to a given donor. This assessment depends on the donor's action (cooperation or defection) and on the reputations of the donor and the recipient (good or bad). The action rule determines with which probability a donor cooperates with a given recipient. Again, this choice depends on the reputation of the donor and the recipient. The original leading eight social norms are deterministic, such that all probabilities are either zero or one. In our framework, we introduce stochasticity by allowing individuals to be generous. We distinguish two kinds of generosity. **b,c,** Assessment generosity means that every time individuals usually assign a bad reputation, they instead assign a good reputation with probability g_1. **d,e,** Action generosity means that every time individuals usually defect, they instead cooperate with probability g_2.

the same probability g_1 for all instances in which assessment generosity can be applied (similarly, they use the same probability g_2 for all instances in which action generosity can be applied). For example, an observer with social norm Stern Judging (or $L6$) makes no distinction between a player who cooperates with a bad recipient and a player who defects with a good recipient. In both cases, the observer generously assigns a good reputation with probability g_1.

3.2.2 Reputation dynamics under assessment generosity

In the following, we first explore each kind of generosity in isolation. We start by considering the effects of assessment generosity. To this end, we first take the players' norms as given. Different players may adopt different norms, but each player's social norm is fixed in time. We can then describe how the players' reputations change over time with a so-called image matrix [Uch10] $M(t)=\big(m_{ij}(t)\big)$. This image matrix represents a collection of the players' reputation repositories. An entry $m_{ij}(t)=1$ indicates that player i assigns a good reputation to j at time t. Similarly, an entry $m_{ij}(t)=0$ indicates that i regards j as bad. After every round of the game, $M(t)$ is updated. The updated image matrix depends on which players are chosen to be the donor and the recipient, on the donor's action, on who observes the interaction, and on the social norms applied by the observers. For example, if an individual j is regarded as good by all players, but j defects against some other good individual, then some of the entries of the image matrix may change from $m_{ij}(t)=1$ to $m_{ij}(t+1)=0$.

Let us illustrate the concept of image matrices with an example. To this end, we consider a population that in equal proportions applies one of the social norms L_1, $ALLC$, $ALLD$. Here, $ALLC$ is the (trivial) social norm that assigns a good reputation to all behaviors, and that prescribes to cooperate with everyone. Similarly, $ALLD$ is the norm that uniformly assigns bad reputations and that always prescribes to defect. We consider four scenarios, depending on whether or not information is noisy, and depending on whether or not L_1 displays some assessment generosity (**Figure 3.2a**). In the baseline case of no noise and no generosity, we observe that L_1 perfectly distinguishes between different players. Eventually, all L_1 players assign a good reputation to each other and to all unconditional cooperators, whereas they assign a bad reputation to all defectors. Once we allow for perception errors, however, there is no longer a perfect match between the players' norms and their reputations. For instance, there is now a 7.5% chance that L_1 players regard an $ALLD$ opponent as good despite their constant defection. Similarly, L_1 players now consider each other as bad with an average probability of 11.6% (**Figure 3.2b**). Given that the error rate is only 5%, this increase in bad assessments cannot be explained by the direct effect of errors alone. Instead, the original disagreements caused by errors trigger further disagreements over time[HST$^+$18]. It is exactly this excess in bad assessments that generosity might help to reduce.

Surprisingly, we find that assessment generosity rather has the opposite effect. It further increases the mismatch between the players' social norms and their reputations (**Figure 3.2a**, right panels). Even in the absence of errors, generosity increase the likelihood that L_1 players regard each other as bad (from 0% to 4.1%). Similarly, it increases the likelihood that L_1 players have a positive image of $ALLD$ opponents (7.2% instead of 0%). This mismatch becomes even worse when errors and generosity act simultaneously (**Figure 3.2a,b**, lower right panels). Now, 12.5% of L_1 players regard each other as bad, while they assign a good reputation to defectors in 13.9% of the cases. In particular, independent of whether or not there are errors, we observe that generosity undermines the accuracy of L_1 to assign correct reputations: (*i*) With and without errors, generosity reduces the likelihood that members of the L_1 subpopulation regard each other as good; (*ii*) With and without errors, generosity increases the likelihood that L_1 players assign a good reputation to defectors. We observe the same negative trend for all other leading eight norms (as another example, the case of L_7 is depicted in **Figure 3.2c,d**).

To gain some intuition for these results, **Figure 3.3a** considers a stylized example with three L_1 players and one $ALLD$ player. Due to assessment generosity, some L_1 players may assign a good reputation to the defector (**Figure 3.3a**, second panel). This can have two negative consequences for the relative performance of L_1: (*i*) Generous players provide help to the defector (**Figure 3.3a**, second and third panel); (*ii*) They assign bad reputations to fellow L_1 players who do not show the same generosity (**Figure 3.3a**, fourth panel). In this way, uncoordinated generosity can itself act as a seed of disagreement between generally cooperative players. Both of these effects of assessment generosity seem to undermine the robustness of the leading-eight norms in mixed populations, rather than enhancing it.

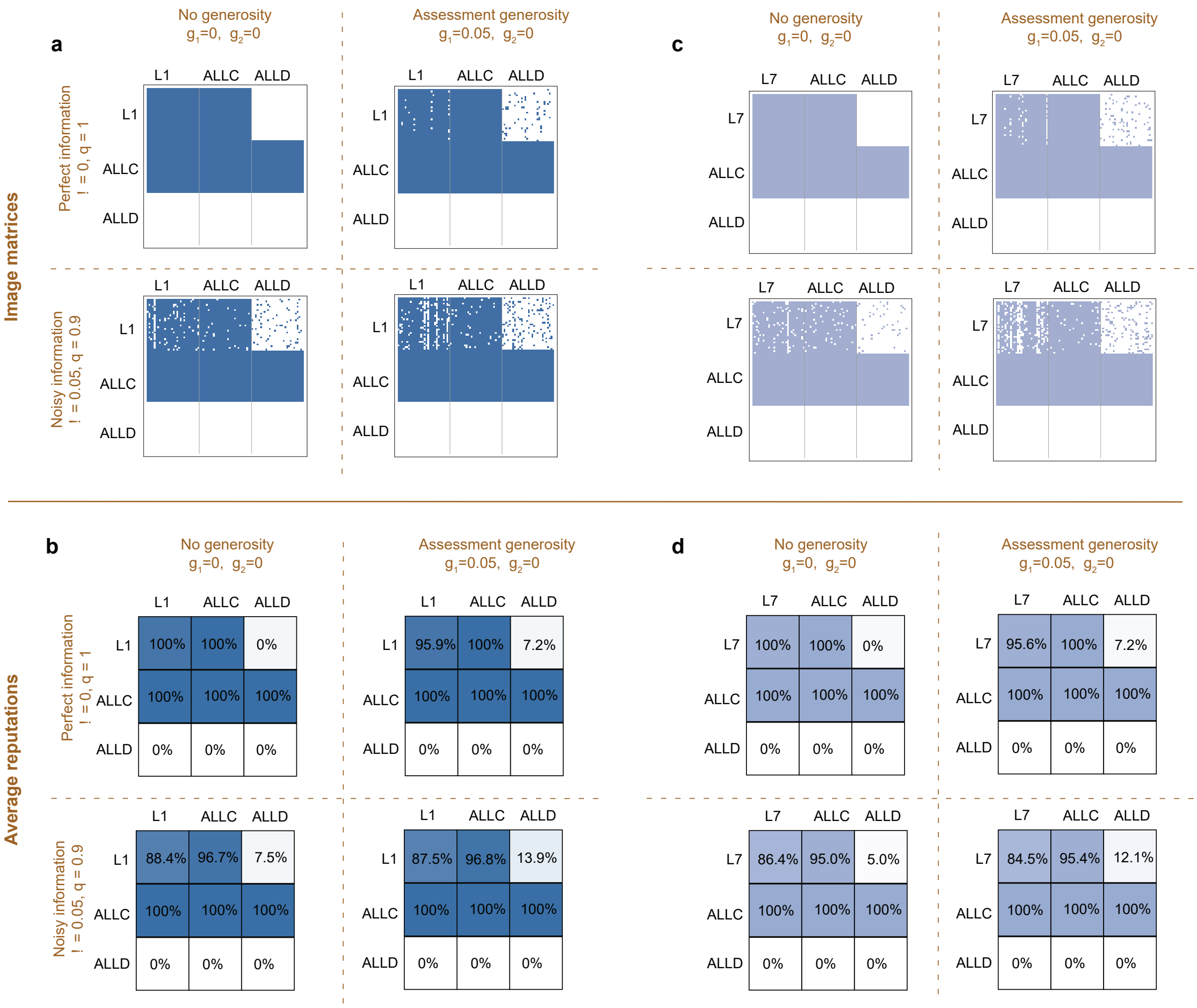

Figure 3.2: **The effect of assessment generosity on the dynamics of reputations.** **a,** Image matrices are representations of how players assess each other at any given time. To depict these image matrices graphically, a colored dot means that the corresponding row player attributes a good image to the corresponding column player. Here, we show snapshots of such image matrices when players either use the leading-eight social norm L_1, $ALLC$, or $ALLD$ (in equal proportions). We consider four scenarios. These scenarios differ in whether information is perfect or noisy, and in whether or not L_1 players are generous. When information is perfect and there is no generosity, we observe that the reputation assignments of different L_1 players are perfectly correlated. If one L_1 player assigns a good reputation to some other group member, then so does every other L_1 player. In contrast, the presence of either noise or generosity introduces disagreements among L_1 players. **b,** Here, we show the average image players have of one another. Generosity makes L_1 players perceive each other less favorably, and it makes them perceive $ALLD$ players more favorably, irrespective of whether information is perfect or noisy. **c,d,** We observe similar patterns for all other leading eight social norms. Here we illustrate the competition between L_7, $ALLC$, and $ALLD$. Parameters: We use a population of size $N = 90$, an error rate of either $\varepsilon = 0$ or $\varepsilon = 0.05$, and a generosity probability of either $g_1 = 0$ or $g_1 = 0.05$. Simulations are run for $2 \cdot 10^6$ iterations, and the initial image matrix assumes a good reputation for all players.

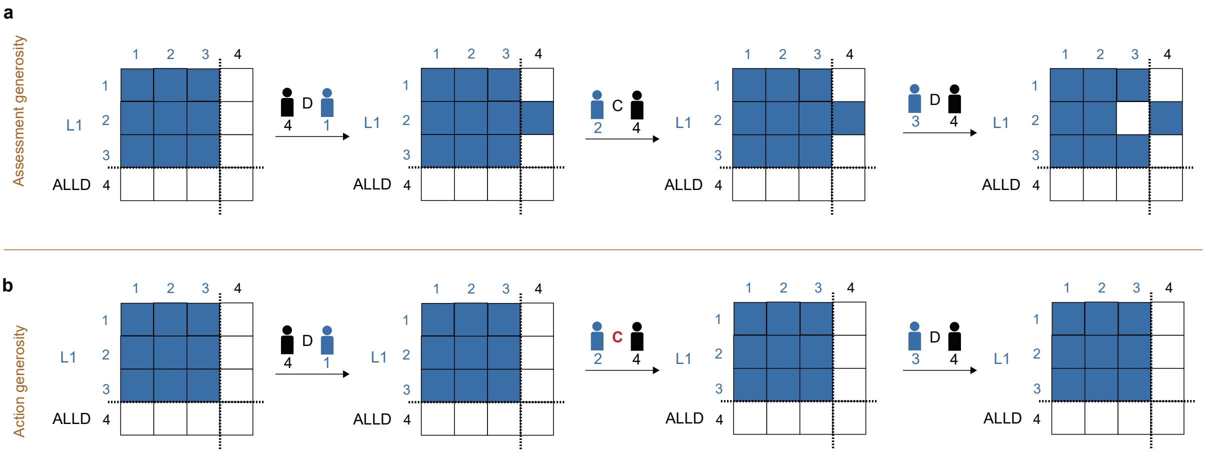

Figure 3.3: **The mechanisms of assessment and action generosity differ. a,** Assessment generosity g_1 works like an additional private error rate, and seeds disagreements that can later proliferate in a population using any generous leading eight strategy, even when information is not noisy or incomplete. We show an example situation where such a disagreement arises in a population of 3 generous players and one ALLD player. When a generous player deems the ALLD player as good despite having observed or received a defection from her, he is setting up ALLD for higher reward and the generous strategy for in-fighting later on. **b,** Action generosity means more frequent cooperation with players deemed bad. This does not change the reputation dynamics for L1 or L7, as these two strategies either never judge any cooperative action as bad or do not change their opinion about a cooperative player in the first place. However, action generosity still ends up giving benefits to defectors, who can then exploit the leading eight players' generosity. Intuitively, action generosity can be compared to a "public error" - this way it either actively harms in strategies where cooperation with bad players is judged as bad, or harms on the level of strategy evolution.

3.2.3 Reputation dynamics under action generosity

In a similar way, we can also explore the isolated effects of action generosity. To this end, we again consider populations that consist in equal proportions of L_1, $ALLC$, and $ALLD$, and we compare the outcomes with and without errors, and with and without action generosity (**Figure 3.4a**). In contrast to assessment generosity, we observe that action generosity does not compromise the perceptions that L_1 players have of each other. In the absence of errors, all L_1 players regard each other as good, with or without action generosity (**Figure 3.4b**, upper two panels). Once there are perception errors, action generosity can even improve L_1's self-perception. Instead of regarding each other as bad with an average probability of 11.6% (in the case of errors but no generosity), L_1 players assign each other a bad reputation in 10.9% of the cases (with action generosity). Similarly, the chance to misperceive $ALLD$ players as good is slightly diminished. L_1 players regard 7.5% of their $ALLD$ opponents as good without generosity, compared to 7.4% with generosity (note, however, that under action generosity, an L_1 player may cooperate with $ALLD$ players even when they have a bad reputation).

To gain some intuition for the differences between assessment generosity and action generosity, we revisit our previous stylized example with three L_1 players and one $ALLD$ player (**Figure 3.3b**). Similar to assessment generosity, we again observe that action generosity may lead L_1 players to cooperate with defectors (**Figure 3.3b**, second and

Figure 3.4: **The effect of action generosity on the reputation dynamics.** We consider the same basic setup as in **Figure 3.2**, except that leading eight players now use action generosity instead of assessment generosity. **a,b,** When the population consists of L_1, *ALLC*, and *ALLD*, action generosity results in more appropriate judgments. In the presence of errors, generosity makes L_1 players more likely to perceive each other as good (89.1% instead of 88.4%). Similarly, it makes them slightly less likely to perceive defectors as good (7.4% instead of 7.5%). Importantly, however, L_1 players with action generosity occasionally cooperate with *ALLD* opponents even if they regard them as bad. **c,d,** We observe the same patterns for other leading eight norms. Here we again depict the case of L_7. Parameters are the same as in **Figure 3.2**.

third panel). However, action generosity no longer seeds additional disagreements among the L_1 players (**Figure 3.3a**, fourth panel). The main difference between assessment and action generosity is that instances of action generosity become common knowledge. Cooperating with a defector is a decision that others can see, whereas assigning a good reputation to a defector is a private decision which can lead to future misunderstandings.

These results suggest that with respect to the robustness of L_1, action generosity results in a trade-off. On the one hand, action generosity makes it more likely that L_1 players cooperate among each other when information is noisy. On the other hand, however, action generosity also increases the chance that help is given to undeserving defectors.

Similar trade-offs can also be observed for other leading-eight norms (**Figure 3.4c,d** shows the example of L_7). Whether the overall effect of this trade-off is favorable to the robustness of indirect reciprocity depends on the overall composition of the population. In populations in which leading eight players are abundant, the positive effects of increased forgiveness may outweigh the negative effects of providing help to free riders. In an evolutionary context, this composition of the population can change over time, depending on the relative success of each social norm. How generosity affects this evolutionary dynamics is what we explore in the following.

3.2.4 Evolutionary dynamics under assessment and action generosity

While we have assumed in the previous sections that the population members use fixed social norms, here we consider how the abundance of different social norms may change over time. We assume that the evolution of social norms happens on a separate, much larger timescale than the updating of reputations. As a result, we may assume that by the time social norms change, the players' reputations have reached a stationary state (as depicted by the average image matrices in **Figure 3.2b,d** and **Figure 3.4b,d**). Given these average images, we can compute the cooperation rate $\hat{x}_{ij}$ with which a player i provides help to player j. Using these cooperation rates, we compute player i's average payoff as $\pi_i = \frac{1}{N-1}\sum_{j\neq i} b\hat{x}_{ji} - c\hat{x}_{ij}$. If a social norm allows a player to have a relatively high payoff, such a social norm is more likely to spread in a population on an evolutionary timescale.

To describe how successful social norms spread in a population, we consider a simple imitation dynamics based on pairwise comparison[TPN07, WBGT15]. In every time step of the evolutionary process, some player i is randomly selected from the population to revise their social norm. With probability μ, player i picks a new norm uniformly at random. With probability $1-\mu$, he instead chooses a role model j randomly from the remaining population. In that case, player i adopts j's social norm with a probability given by the Fermi function[Blu93, ST98, TNP06], $P(\pi_i, \pi_j) = (1+\exp[-s(\pi_j - \pi_i)])^{-1}$. Here, the parameter $s \geq 0$ describes the strength of selection, which measures how relevant payoffs are for updating the norms. For $s = 0$, updating happens at random. For increasing s, social norms with higher payoffs are increasingly likely to be imitated. The resulting stochastic process is ergodic, and gives rise to a stationary distribution, called selection-mutation equilibrium. The average cooperation rate or payoff in the population can be computed from there by weighting the payoffs of the individual norms with their abundance in equilibrium. In the following, we assume mutations are sufficiently rare, such that populations are homogenous most of the time[FI06, WGWT12, McA15]. Only occasionally, a mutant social norm arises. This mutant norm is then either adopted by the entire population or it goes extinct before the next mutant appears. To furthermore keep the system as simple as possible, we study the evolutionary competition between three social norms only: players can choose between $ALLC$, $ALLD$, and one of the leading eight social norms (similar to what we studied in the previous sections and again similar to much of the previous work in the field[NS98a, PB03, BS04, BS06, Ber11, SSP16, RSP19]). For all details, see **Methods**, Section 3.4.

Figure 3.5a shows the resulting dynamics for the leading eight norm L_1 in the baseline case without any generosity[HST+18]. Overall, L_1 is only played in 30% of the cases, as compared to 4.3% for $ALLC$, and 65.8% for $ALLD$. While L_1 is relatively robust against direct invasion by $ALLD$, it is vulnerable to indirect invasion by $ALLC$. Overall, the dynamics is similar to a rock-paper-scissors cycle: $ALLC$ mutants are favored to invade into L_1, $ALLD$ is favored to invade into $ALLC$, and L_1 can invade $ALLD$ in turn. However, due to the relative robustness of $ALLD$, players only cooperate in a minority of cases. Adding a moderate amount of either assessment generosity (**Figure 3.5b**) or action generosity (**Figure 3.5c**) leaves the qualitative dynamics unchanged. However, both kinds of generosity make L_1 slightly more robust against invasion by $ALLC$, but less likely to invade into $ALLD$, as one may expect. The overall effect is slightly negative in the case of action generosity (the share of defectors increases from 65.8% to 67.0%), and substantially negative in the case of assessment generosity (where the share of defectors increases to 78.6%). For the given parameter values, we thus conclude that neither form of generosity favors the evolution of the social norm L_1.

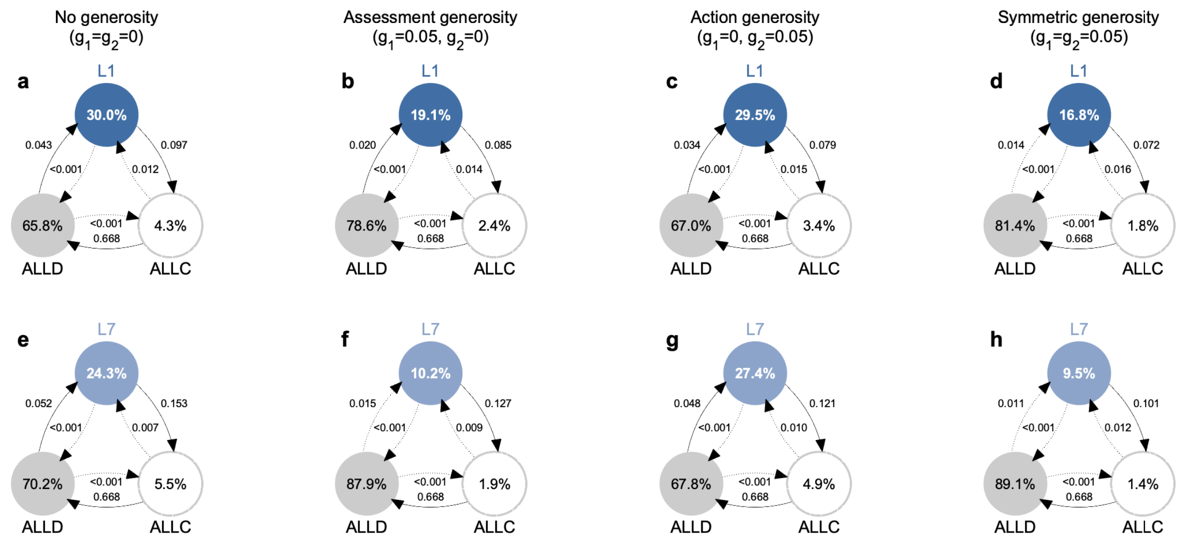

Figure 3.5: **Evolution of the leading eight under assessment and action generosity.** We simulate the evolutionary dynamics when players can choose among three different norms, a leading eight norm, $ALLC$, and $ALLD$. Social norms spread in the population according to a pairwise comparison process[TPN07], such hat norms of players with high payoffs are more likely to spread. Here we depict results for the limit of rare mutations, such that populations are homogeneous most of the time[FI06, WGWT12, McA15]. Numbers in circles show how often each social norm is adopted on average. Arrows indicate how likely other social norms can invade a given resident population. Solid arrows indicate that the respective transition is more likely to occur than expected under neutrality, whereas dotted arrows indicate that the respective transition is comparably unlikely. We consider four scenarios, depending on whether leading eight players exhibit no generosity, only assessment generosity, only action generosity, or both variants of generosity. **a–d,** The norm L_1 is most abundant without any generosity. **e–h,** The norm L_7 is most abundant with action generosity. However, even in that case, it is played in less than 30% of time. Parameters: $N\!=\!50$, $\varepsilon\!=\!0.05$, $b\!=\!5$, $c\!=\!1$, $q\!=\!0.9$, using a strength of selection of $s\!=\!1$.

This conclusion, however, can change for other leading eight social norms. To illustrate this point, **Figure 3.5e–h** shows the resulting dynamics when $ALLC$ and $ALLD$ compete with the alternative norm L_7. The overall dynamics is similar to the case of L_1, with evolution leading in a rock-paper-scissors cycle from L_7 to $ALLC$ to $ALLD$ and back to L_7. Also the impact of generosity is similar. On the one hand, it reduces the risk of L_7

with respect to indirect invasions by $ALLC$; but on the other hand, it also reduces L_7's ability to invade into $ALLD$. However, for L_7 we observe that the net result of these two opposing effects can be positive. In the case of action generosity, the share of L_7 increases to 27.4% (compared to 24.3% in the baseline scenario without any generosity). We note however that even in this case, in which generosity is favorable, the population is still most likely to settle at unconditional defection.

To explore how robust these patterns are, we have repeated these simulations for all other leading eight social norms, and we have systematically varied how likely the leading eight are to engage in either assessment or action generosity (**Figure 3.6**). These simulations exhibit the following regularities: (*i*) In the presence of noise, almost all of the leading eight norms have problems to evolve, irrespective of how generous they are. The only social norm that is able to reach an abundance of more than 50% is L_2 (called 'consistent standing'[HST+18]). However, this social norm is most successful without generosity, when $g_1 = g_2 = 0$. (*ii*) In those cases in which the leading eight norm is played by a sizable fraction of the population (L_1, L_2, L_7), assessment generosity has a systematically negative effect on overall cooperation. In contrast, action generosity can promote cooperation, but only in the case of L_7, for which the optimal degree of generosity is $g_2 \approx 8\%$. However, even in that case, cooperation rates remain comparably low. For all other leading eight norms (L_3-L_6, L_8), cooperation fails to evolve altogether.

We observe similar results as we vary the error rate ε, the benefit b of cooperation, and the observation rate q (**Figure 3.7**). In all cases, cooperation is unlikely to evolve.

Moreover, generosity seems to harm the evolution of cooperation rather than enhancing it (**Figure 3.8**), despite its positive effect in homogeneous populations (**Figure 3.9**).

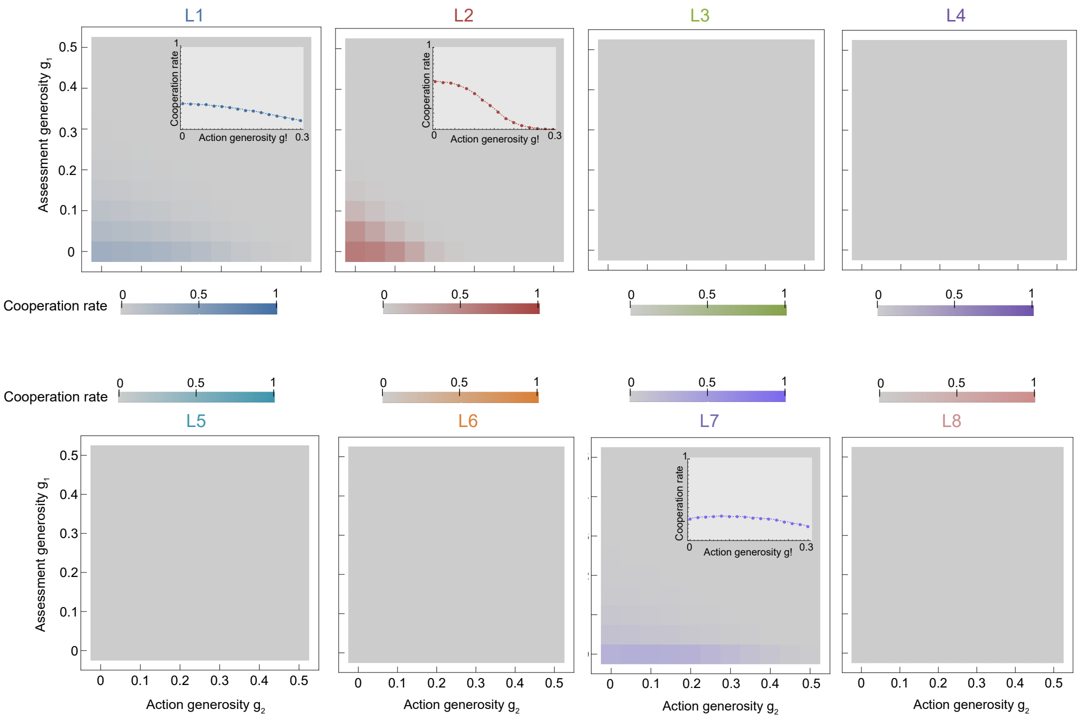

Figure 3.6: **A systematic analysis of the effect of generosity on cooperation.** For this figure, we repeat the evolutionary simulations shown in **Figure 3.5** for all leading eight strategies. To explore the impact of generosity, we systematically vary how likely leading eight players exhibit assessment generosity (g_1, y-axis) and how likely the exhibit action generosity (g_2, x-axis). The color indicates the average cooperation rate of the population, according to the selection-mutation equilibrium of the evolutionary process (see **Methods**, Section 3.4). In particular, grey indicates the absence of cooperation. We find that the five social norms $L_3 - L_6$ and L_8, which fail to evolve in the baseline deterministic model ($g_1 = g_2 = 0$), do not evolve in any generous form either, no matter in which combination. The strategies L_1 and L_2 occasionally evolve, but they do not benefit from either form of generosity. Here, the achieved cooperation rate has its maximum in the origin. Only for L_7, the cooperation rate becomes maximal for a positive amount of action generosity, with $g_2 \approx 0.08$. The inserts show the cooperation rate as a function of g_2 for $g_1 = 0$. Parameters are the same as in **Figure 3.5**.

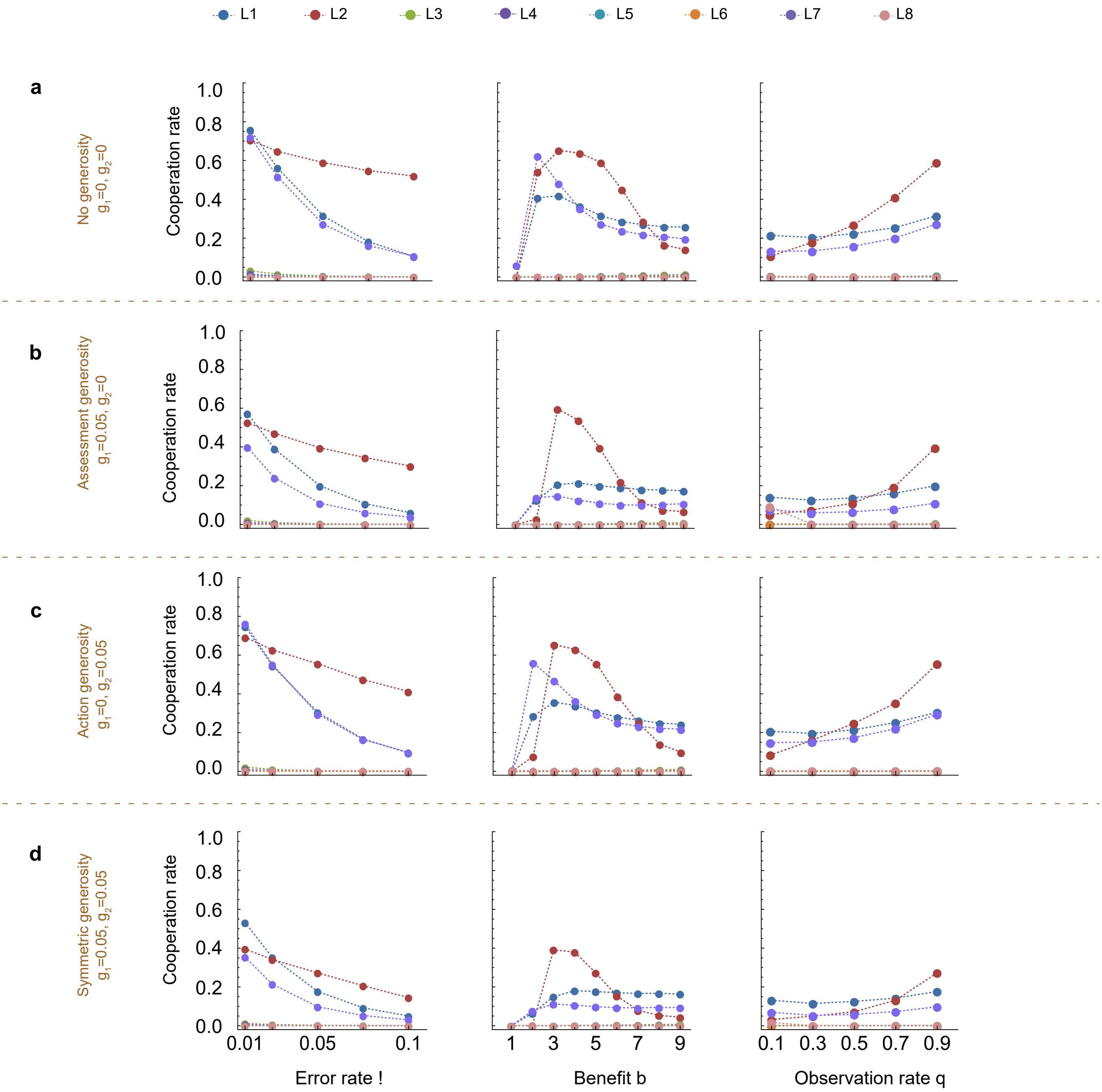

Figure 3.7: **Generosity does not enhance the evolution of cooperation even when we vary parameter values.** We vary the noise on observations ε, the benefit-to-cost ratio b/c, with $c = 1$, and observation probability q. All other parameters remain constant at the values of **Figure 3.2**. In each scenario, we plot the average cooperation rate of each individual leading eight strategy when they compete against ALLD and ALLC. This again enables us to compare the generous L_i (**b-d**) in three variants with their baseline counterparts (**a**). **b,** For assessment generosity only, we find that compared to the baseline, the cooperation rate of the leading eight with generosity is reduced for all values of ε, b/c and q. The qualitative shape of the curves remains the same as in the case of no generosity. **c,** In the case of action generosity only, the negative effect on cooperation rates is not as large as with assessment generosity. Yet, action generosity still fails to enhance cooperation, and is also detrimental for some parameter ranges, especially higher values of ε and b/c. **d,** For the leading eight with both kinds of generosity, cooperation rates are lowest across all parameter ranges and scenarios. Again, the shape of the curve is identical to its deterministic baseline in **a**, but cooperation cannot evolve in the same way as in the baseline, where it is already limited.

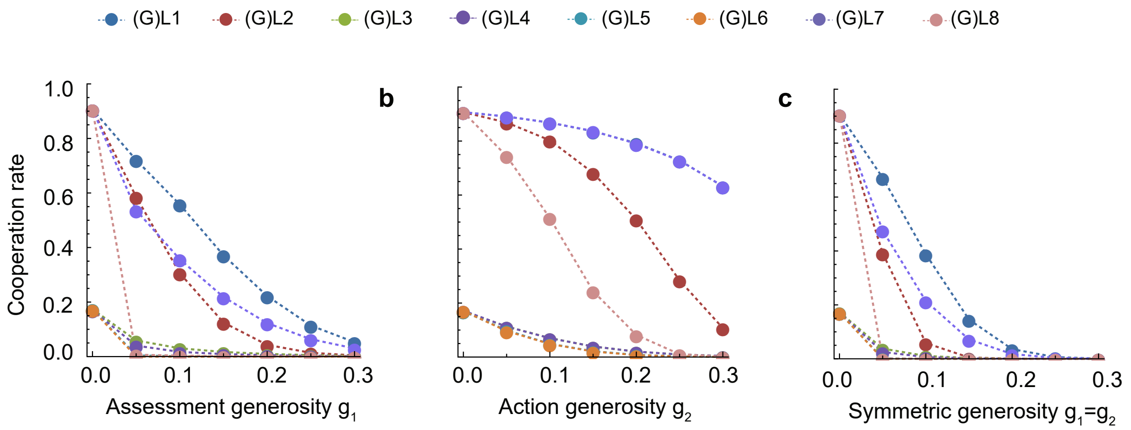

Figure 3.8: **Generosity in itself inhibits the evolution of cooperation.** We consider an error-free scenario ($\varepsilon = 0$) with perfect observation ($q = 1$), hereon called a “perfect information scenario”, and calculate the cooperation rate in equilibrium as a function of generosity. **a,** In the case of assessment generosity only as well as when both variants of generosity are at play (**c**), there is a quick decline of cooperation for all generous leading eight as generosity increases. L1 does a bit better than the rest, but suffers the same losses once generosity is past 1%. **b,** When we consider only action generosity, the picture is slightly different: L1 and L7 can keep up a higher level of cooperation with a rate of around 85% until a probability of forgiveness at around 0.25. L2 also fares better in this setup than in the two other scenarios, but shows a decline in performance earlier on, at around $g_2 = 0.15$. It thus stands to reason that generosity in itself introduces disagreements and opportunities for ALLD to gain an advantage. Parameters: $N = 50, \varepsilon = 0, q = 1$. $b = 5$, $c = 1$, $s = 1$, in the limit of rare mutations $\mu \to 0$.

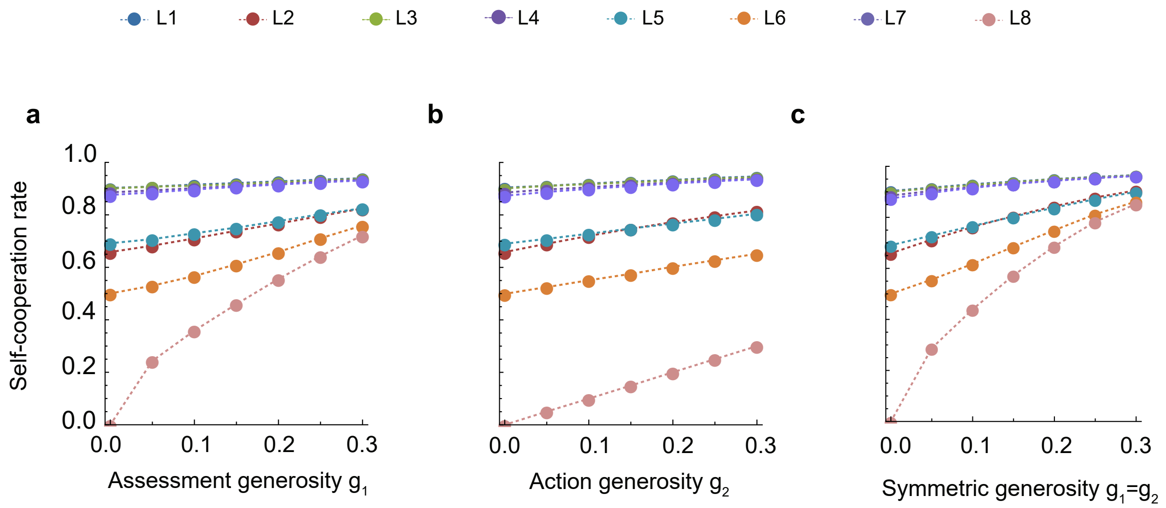

Figure 3.9: **The self-cooperation rate in homogeneous populations of generous leading eight players benefits from generosity. a-c,** We consider homogeneous populations consisting of $N = 50$ players using generous leading eight strategies in a noisy environment, and calculate their strategy's cooperation rate against itself. We find that as generosity increases, so does self-cooperation. This is true for assessment generosity only (**a**), action generosity only (**b**), as well as symmetric generosity ($g_1 = g_2$) (**c**). This result suggests that only when strategy evolution is at play and the generous strategies have to compete against other social norms, forgiveness is a hindrance. Single generous strategies in isolation fare better in noisy environments than their deterministic variants. This is in line with the intuition that forgiveness helps balance out noise-related misinterpretations of cooperation as defection. Parameters: $N = 50, \varepsilon = 0.05, q = 0.9$.

3.3 Discussion

Indirect reciprocity is a mechanism for cooperation that translates the principle of direct reciprocity to a population level [NS05, Sig10, Sig12, Oka20]. When individuals use direct reciprocity, they ask how cooperative others have been in direct encounters. In contrast, when individuals use indirect reciprocity, they ask how others have behaved in general, in interactions with third parties. Once individuals take into account third party interactions, they also need to develop a sense of which behaviors should be regarded as bad, and which consequences bad reputations should have. In this way, indirect reciprocity has become the subfield of evolutionary game theory that explores the evolution of moral behavior [Ale87]. To explore these questions, Ohtsuki and Iwasa [OI04, OI06] have suggested eight social norms that can maintain cooperation. Their work, however, rests on the assumption that reputations are assigned publicly. In their framework it is a central authority that takes note of everyone's behavior, and that assigns a good or a bad reputation in turn. In contrast, when population members make their own judgments and when information is noisy, Ohtsuki and Iwasa's leading eight become less effective in sustaining cooperation [Uch10, US13, OSN17, HST^{+}18]. Noise can introduce initial disagreements, such that different individuals assign different reputations to a given group member. Such disagreements can proliferate, because they also affect how future interactions of that group member are interpreted. As a result, minor initial disagreements can sometimes lead to a complete separation of the population. In such cases, the population fragments into distinct subpopulations that assign good reputations within their community and bad reputations to all outgroup members[OSI13].

Here, we have explored whether the leading eight social norms can be made more robust if they are equipped with an element of generosity. We distinguish two ways of being generous. When individuals engage in assessment generosity, they are more likely to assign a good reputation to someone they would usually regard as bad. When individuals engage in action generosity, they are more likely to cooperate with someone against whom they would usually defect. Although these two manifestations of generosity may follow a similar motive, our results suggest that they have different consequences. Assessment generosity leads to a private reassessment of somebody's reputation. Unless this reassessment is mutually agreed upon, it can seed additional disagreements within a population. These disagreements can further undermine the robustness of the leading eight norms. In contrast, action generosity leads to a public display of cooperative behavior. While such generous acts may be misguided towards undeserving free riders, they do not generate the mutual disagreements that assessment generosity is susceptible to. As a result, we find that the effect of assessment generosity on cooperation is always negative, whereas action generosity can sometimes enhance cooperation. But even then, the leading eight norms have problems to sustain full cooperation in noisy environments.

We note here that the case of fully private, individual observations and reputation tracking constitutes an extreme case just as much as fully public information. We assume that for a more realistic model, the truth lies somewhere in the middle [OYU20]. For example, gossip and communication help offset and mediate disagreements. Still, for a full understanding of the dynamics of indirect reciprocity, it should be useful to first explore the extreme cases. If not for practical reasons, these cases are still relevant because of their theoretical merit as boundary cases.

The failure of generosity to enhance cooperation may seem surprising. After all, previous work on direct reciprocity has shown that generosity is a powerful mechanism to reestablish mutual cooperation in the presence of noise [Mol85, NS92, MVCS12, SP13, HWTN14]. Also among first-order social norms of indirect reciprocity, generosity can help to stabilize cooperation[SCHN21]. Our study can shed light on why generosity is conducive to cooperation in those cases, while it is detrimental in the case of the leading eight norms. For the leading eight norms, generosity entails a two-fold cost. The first cost is the immediate cost c that comes with any cooperative act. The second cost is the risk of losing one's reputation by providing help to someone others regard as bad. In contrast, Generous Tit-for-Tat and Generous Scoring do not suffer from such a two-fold cost. Although the direct act of cooperating with someone is still costly, cooperators always obtain a good reputation. The optimal level of generosity in this case can be calculated by requiring that the positive reputation effect matches the negative immediate cost of cooperation [SCHN21].

To model the impact of generosity, we make a number of simplifying assumptions. For one, we have considered a well-mixed population. While the assumption of well-mixed populations is common in the indirect reciprocity literature [NS05, Sig12, NS98b, OI04, OI06, HST^{+}18, Sig10] interesting effects may arise once a population's network structure is taken into account. In the context of indirect reciprocity, this population structure may not only determine with whom players interact with, but also whose interactions they are able to observe. While population structure can enhance the evolution of cooperation in general [LHN05, ON08, ON07, SF07], it remains to be shown how it affects the dynamics of indirect reciprocity with and without generosity.

In addition, our analysis is based on computer simulations of the dynamics in finite populations. Instead, recent work has shown that analytical results are feasible in infinite populations if the players' observations can be assumed to be independent [OSN18, RSP19]. While the dynamics in finite populations is perhaps more relevant for applications, an analytical treatment of generosity is another potential venue for future research.

Furthermore, we have assumed an "aligned" version of forgiveness: players have a single probability value each for generosity in assessments (g_1) and actions (g_2), respectively. They do not differentiate between cases within one component of their social norm, i.e. do not apply less stochasticity in scenarios they deem less "forgivable". Note that a non aligned version of generosity with different probability of forgiveness for different observations or actions would certainly be more realistic. For example, it might well be that cooperation with a person deemed as bad should be subject to forgiveness more often than defection against a good person. However, this also intuitively demands a higher cognitive capacity from the players, who then need to prioritize some assessments and actions over others, and need to decide what they think are unforgivable offenses. That is, they need a more involved sense of morality. This points to a direction of interesting future work, which can take into account the importance that players assign to different scenarios they might encounter rather than applying an element of randomness uniformly to their actions and assessments. Exploring such "non-aligned" generosity component-wise would aid our understanding of which aspects of norms need to truly be set in stone, and which aspects deserve more flexibility in assessment and action. Our study with its negative result thus is only a first step towards understanding the effect of generosity for social norms.

Finally, previous work has shown that in more complex social situations, apologies are a necessary first step before forgiveness can occur. Without them, individuals lack discernment of what constitutes a justified or unjustified defection [HMPS11, MVHPL15]. This makes our results intuitive from a sociology perspective: without a person being able to observe the intent behind an action, forgiveness does not help in re-establishing cooperation. It thus would potentially be worthwhile to explore the use of apology strategies in indirect reciprocity, similar to [MVHPL15], which analyzes direct reciprocity dynamics involving apology and forgiveness. Our study thus further serves to highlight the importance of shared information and coordination in a society when more complex social norms are applied. Simply acting in good faith every once in a while cannot balance out the intricate dynamics that result from disagreements in a population.

3.4 Methods

3.4.1 Reputation dynamics

We first describe how the reputations of players change over time. To this end, we assume social norms to be fixed for each individual. We record the reputations that players assign to one another in one image matrix $M(t)$, which is updated in every step of the dynamics. An entry $m_{ij} = 1$ ($m_{ij} = 0$) means that player i regards player j as good (bad), respectively. In the following, we always start with all entries $m_{ij} = 1$, i.e. in a state where everyone assigns a good reputation to everyone else, no matter the social norm. In every round, a donor and recipient are chosen at random, and the donor can choose whether to confer a benefit b to the recipient at own cost c. Her action is dependent on her norm as well as the reputation she assigns to herself and the recipient. This action is observed by her co-players with individual probability q. This observation is subject to misperception with probability ε. The players who observe the interaction then update the donor's reputation according to their assessment rules. These updates are recorded in the matrix $M(t+1)$. We can iterate this process over many rounds and calculate the average image players have of each other. Specifically, if players have interacted for T rounds in total, the average image that player i's assigns to player j is defined as $\frac{1}{T}\sum_{t=1}^{T} m_{ij}(t)$. The resulting average cooperation frequency $\hat{x}_{ij}$ enables us to calculate players' payoffs for a fixed social norm as

$$\pi_i = \frac{1}{N-1} \sum_{j \neq i} b\hat{x}_{ji} - c\hat{x}_{ij}. \tag{3.1}$$

We illustrate our approach in **Figure 3.2** and **Figure 3.4**. In both cases, the upper panels (**a,c**) show snapshots of the image matrix at time $T = 2 \cdot 10^6$. The lower panels (**b,d**) show the respective average images. In **Figure 3.9**, we further present the results of reputation dynamics in a homogeneous population of leading eight players, by showing the average cooperation rate of each leading eight norm against itself.

We want to point out that in general, "good" and "bad" are merely labels with no inherent meanings. In particular, for each of the leading eight strategies, one could also define "mirror strategies". According to these mirror strategies, acceptable behavior would be rewarded with a "bad" reputation, and players would be more likely to cooperate with "bad" individuals. If we are also to allow for such mirror strategies, it might be less clear how "assessment generosity" should be defined. After all, for the mirrors, assessment generosity should imply that forgiveness makes it more likely that players now assign a "bad" reputation to the respective player. However, in this study, we are only concerned with the leading eight strategies as defined by Ohtsuki and Iwasa [OI04]. In their definition, the terms "good" and "bad" already have some intrinsic meaning. Here, "good" is considered to mean the reputation label that, according to the leading eight's action rule, makes it more likely for the respective player to receive cooperation. In our model, players thus unambiguously prefer to be labeled as "good". In that case, assessment generosity is straightforward to define. It means that players are more likely to assign good reputation to their peers.

As one of our results, we show that action generosity may be more conducive to the evolution of cooperation than assessment generosity. This may appear unsurprising, because action generosity immediately affects cooperation. In case of action generosity, each instance of generosity directly translates into an act of cooperation. In contrast, in the case of assessment generosity, each instance of generosity only affects the respective player's reputation. Whether this improved reputation translates into more received cooperation depends on the player's further interactions. However, for many of the presented results, we actually do not explore the impact of generosity on the eventual cooperation rate. Rather, we explore how easily generosity can induce disagreements between leading eight players (**Figure 3.2** and **Figure 3.4**), or how it affects the evolution of leading eight strategies (**Figure 3.5**). Already on this level, we observe that action generosity is more favorable than assessment generosity. Furthermore, we show that even if action generosity by itself increases cooperation, we find that in most cases it actually has a negative effect on cooperation.

We also emphasize that our analysis technique is not limited to the restricted setup where only three social norms compete. The approach can be naturally extended by allowing population members to choose among additional norms. However, computational complexity increases rapidly in the number of considered social norms. Also, our results show that already when competing with only these two very simple other strategies, the leading eight norms are unstable. Thus, adding further social norms is unlikely to increase the stability of the leading eight norms and similarly unlikely to affect our conclusions.

3.4.2 Evolutionary dynamics

In a next step, we explore which social norms the individuals themselves choose to adopt when their norms are no longer fixed. To describe this process formally, we assume players change their norms over a separate, longer time scale. On this longer time scale, we assume individuals adopt new social norms based on a pairwise comparison process [TPN07]. According to this process, in every evolutionary time step, one player is randomly chosen from the population. With probability μ, with μ the mutation rate, this player picks a new social norm at random from the respective set of available norms. In this work, players can choose from three social norms, a given leading eight norm, $ALLC$, and $ALLD$. Meanwhile, with $1-\mu$, the focal player randomly chooses a role model from the population. If the focal player's payoff according to Eq. (3.1) is given by π_i and the role model's payoff is π_j, then the focal player adopts the role model's norm with probability $P(\pi_i, \pi_j) = (1+\exp[-s(\pi_j - \pi_i)])^{-1}$. The parameter s is called the strength of selection. When s is small, imitation occurs largely at random. For larger s, however, players are most likely to imitate those role models with a higher payoff.

We note that in evolutionary game theory, imitation processes are often used as a standard model to describe the spread of strategies in a population. For this model to be sensible, it is necessary to assume however that players are able to infer their co-players' strategies from their observed behaviors. This is straightforward for simple games, where strategies correspond to the action that a player chooses. It is less straightforward in games of indirect reciprocity. Here, individuals use more complex norms that are more difficult to learn by imitation. However, it may still be argued that players may be able to learn each other's norms. For example, it is not unusual to assume that people discuss their world views and moral guidelines with others. This is what we implicitly assume for this study. Instead of imitation, one could also consider an alternative model by assuming that social norms spread through a birth-death process, i.e. that parents pass on their own social norms to their children. However, for the function we use to model imitation, the imitation process is equivalent to a birth-death process with exponential fitness mapping [WBGT15]. Thus, the results would not change.

This evolutionary process based on mutations and imitation is ergodic. Hence, it gives rise to a unique stationary distribution, which we refer to as the selection-mutation equilibrium. This equilibrium reflects how often each of the available social norms is adopted over time. In this work, we use the limit of rare mutations, which assumes that populations are homogeneous most of the time. When a mutation arises, it either fixes in the population or goes extinct before the next mutant appears. We can calculate this fixation probability of a mutant with social norm M into a resident population with social norm R explicitly as [TH09]

$$\rho_{MR} = \frac{1}{1 + \sum_{i=1}^{n-1} \prod_{k=1}^{i} e^{-\beta\,(\pi_M(k) - \pi_R(k))}}. \tag{3.2}$$

Here, $\pi_M(k)$ and $\pi_R(k)$ are the respective payoffs of mutants (M) and residents (R) when k individuals in the population employ the mutant norm. This means that we can describe the evolution of the social norms between three available norms in the rare mutation limit as a Markov chain with three states. These three states correspond to the respective homogeneous populations, i.e. all players using $ALLC$, all using $ALLD$ or all players using a leading eight norm. Given the pairwise fixation probabilities according to Eq. 5.31, the respective transition matrix of this evolutionary Markov chain is given by

$$W = \begin{pmatrix} 1 - \frac{1}{2}(\rho_{LC} + \rho_{LD}) & \frac{1}{2}\rho_{LC} & \frac{1}{2}\rho_{LD} \\ \frac{1}{2}\rho_{CL} & 1 - \frac{1}{2}(\rho_{CL} + \rho_{CD}) & \frac{1}{2}\rho_{CD} \\ \frac{1}{2}\rho_{DL} & \frac{1}{2}\rho_{DC} & 1 - \frac{1}{2}(\rho_{DL} + \rho_{DC}) \end{pmatrix}. \tag{3.3}$$

The stationary distribution of this transition matrix is the selection-mutation equilibrium of the process for rare mutations[FI06]. Given this equilibrium, we can compute how often players cooperate on average by taking the average cooperation rate of each homogeneous population, and multiplying it by how often we are to observe the respective homogeneous population in equilibrium.

We use this approach in **Figure 3.5** – **Figure 3.8**, where we first simulated the reputation dynamics for all possible population compositions, (n_L, n_C, n_D), with $N = n_L + n_C + n_D = 50$ and 10^6 steps each. Here, n_L stands for the number of players using a (generous) leading eight norm, n_C for the number of players using $ALLC$, and n_D for the number of $ALLD$ players. Payoffs are computed with Eq. 3.1, as explained in the subsection on the reputation dynamics.

Specific methods employed for the figures. **Figure 3.2** shows the results of reputation dynamics in a population of $N = 90$ players. The population composition is as follows: 1/3 uses $ALLD$, 1/3 uses $ALLC$, and 1/3 uses a leading eight norm. We consider the cases of L_1 (**a,b**) and L_7 (**c,d**). We additionally differentiate between four parameter setups. For these setups we first distinguish whether information is perfect or noisy (i.e., whether $\varepsilon = 0$ and $q = 1$, or $\varepsilon = 0.05$ and $q = 0.9$). In addition, we distinguish whether or not players engage in assessment generosity (i.e., whether $g_1 = 0.05$ or $g_1 = 0$, while $g_2 = 0$ throughout).

Figure 3.4 uses exactly the same method as **Figure 3.2**, but considers the case that players use action generosity ($g_1 = 0, g_2 = 0.05$) instead of assessment generosity.

In **Figure 3.5**, we show the abundance of $ALLC$, $ALLD$ and either L_1 or L_7 in the selection-mutation equilibrium. We compare a scenario without generosity ($g_1 = g_2 = 0$) to scenarios with assessment generosity only ($g_1 = 0.05, g_2 = 0$), action generosity only ($g_1 = 0, g_2 = 0.05$), and symmetric generosity ($g_1 = g_2 = 0.05$). Parameters are $b = 5, c = 1, s = 1, \varepsilon = 0.05, q = 0.9$.

In **Figure 3.6**, we systematically explore the effect of generosity on cooperation in the selection-mutation equilibrium for all leading eight norms. We use the same baseline parameters as in **Figure 3.5**, while varying both assessment (y-axis) and action (x-axis) generosity in steps of 0.05 between 0 and 0.5. To visualize the impact of action generosity on the three norms L_1, L_2, and L_7 in more detail, we also ran more fine-grained simulations where we varied g_2 in steps of 0.01, while $g_1 = 0$. These results are presented in the inset panels **a, b, g**.

For **Figure 3.7**, we again employed the same method as in **Figure 3.5** and **Figure 3.6**. Here, we rerun the simulations for varying benefit b , noise ε , and observation probability q. We again compare **a,** the baseline without generosity ($g_1 = g_2 = 0$) to the three scenarios of **b,** assessment generosity only ($g_1 = 0.05, g_2 = 0$), **c,** action generosity only ($g_1 = 0, g_2 = 0.05$), and **d,** both kinds of generosity ($g_1 = g_2 = 0.05$). The other parameters remain the same as in the previous figures.

Figure 3.8 explores the effect of generosity in a perfect information scenario, for an error rate of $\varepsilon = 0$ and an observation probability of $q = 1$. We once more consider three scenarios: assessment generosity only with $g_1 = 0.05$ (panel **a**), action generosity only with $g_2 = 0.05$ (panel **b**), and symmetric generosity, with $g_1 = g_2 = 0.05$ (panel **c**). Methods employed are as in the previous evolutionary figures **Figure 3.5**–**Figure 3.7**. All other parameters remain the same.

For **Figure 3.9**, we vary generosity levels in all three scenarios (assessment generosity only, action generosity only, symmetric generosity) when considering reputation dynamics in a homogeneous population of leading eight players with noisy environment ($\varepsilon = 0.05$, $q = 0.9$). We plot the average cooperation rate of each employed leading eight norm against itself.

CHAPTER 4

Quantitative assessment can stabilize cooperation via indirect reciprocity under imperfect information

Indirect reciprocity describes the role of reputation and social norms for the evolution of cooperation. This mechanism assumes that when individuals interact with someone, they are subject to observation and judgement by their peers. Helping someone at own cost increases individuals' reputation, which benefits helpers in future interactions and therefore can sustain cooperative behavior. Previous literature has identified eight highly successful social norms, the "leading eight". However, to be effective, these norms require all relevant information to be reliable and public. Otherwise, when information is private, disagreements resulting from noise can proliferate and prevent cooperation from being sustained. Studies on the performance of the leading eight usually assume that reputations are binary: players can be thought to be either good or bad. In contrast, here we study the case where reputations are more nuanced. With this "quantitative" assessment, every individual privately keeps track of others' integer reputation scores, incrementing and decrementing these scores according to observations and norm-specific rules. To make an overall judgement of a peer, an individual then compares the corresponding score with a threshold. Our work suggests that such quantitative assessment can act as a powerful stabilizer of cooperation for four of the leading eight strategies when they compete with defectors and cooperators. Keeping track of reputations in a more nuanced way has error-correcting properties, thus increasing the resilience against the effects of noise by enabling recovery from disagreements. For cooperation to be reliably sustained by complex social norms, such comparably more sophisticated assessment and a broader definition of goodness thus seems to be necessary in order to overcome the challenge of noisy environments with incomplete information.

4.1 Introduction

Reputation-based social dynamics are a fundamental feature of the human experience [Ale87, JHTM11, Feh04, BKO05, CGLF+15]. When reputation becomes a sufficiently valuable commodity and actions of individuals can be observed and judged by other members of their community, investing in a good reputation by engaging in costly acts of helping others pays off [WB02, Sug86, SKM04, Kan92, MSK02]. Those that help are more likely to receive such help themselves, making it possible for cooperative behavior to emerge. The framework of indirect reciprocity in evolutionary game theory describes this mechanism for the evolution of cooperation, which is based on shared moral systems in a population [NS05, Now06, Sig10, Sig12, Oka20]. Its success results from population members assessing others' actions towards third parties and then acting according to the resulting reputations. It does not require members to meet more than once. Cooperation becomes the preferable option in interactions simply because it increases public standing, and thus increases the likelihood of an individual receiving benefits from third-party population members in the future.

Simple social dilemmas such as the donation game are a useful model for indirect reciprocity[NS05, Sig10, Sig12, Oka20]. In this game, a donor chooses whether or not to confer a benefit to a recipient at own cost, i.e. whether to cooperate or to defect. Other population members can observe the donor's action, and use this observation to update their assessment of the donor's reputation. This behavior in the population is governed by social norms, assumed to consist of two components [OFP95]. One component is an assessment rule prescribing how to update reputations based on observations. It specifies which behaviors of the donor should improve the donor's reputation, and which behaviors should be condemned, and can depend on various contextual factors. The second part of a social norm is an action rule that tells the donor whether to cooperate or defect in a specific situation. The donor's decision may also depend on his and the recipients' reputation.

Much of the literature on indirect reciprocity has dealt with the question of which social norms excel most at sustaining cooperation, and how complex they need to be [NS98a, NS98b, SSP18, OI06, LH01, NS05, PB03]. Early work mainly focused on simple norms such as "Image Scoring" [NS98b, NS98a, BS05, WM00]. When players use this norm, donors' reputations, i.e. their "score", increase when they cooperate, and decrease when they defect. The action rule then prescribes to cooperate with those that have a sufficiently high reputation score. Such norms that only take into account the donor's action when making an assessment are referred to as "first-order norms". Image Scoring, while intuitive, however has been shown to be unstable under most circumstances [PB03, BS04, LH01, Oht04]. For one, it has been argued that in the long run a player will have no incentive in the first place to care about his interaction partner's score, since it has no bearing on their own success: there is no adaptive motivation to use the norm. Another main reason for the instability of Image Scoring is that punishing defectors by not cooperating with them comes at a high cost to the punisher: their own reputation drops, thus harming them in future interactions. The binary version of Image Scoring, which only considers "good" and "bad" reputations, has additionally been suggested to be highly susceptible to errors. Only variants of the norm that have some tolerance for single defections are considered to be able to sustain cooperation [Ber11, BG16]. More generally, the literature argues that successful norms need to be able to differentiate between justified and unjustified defections, and thus need to be more complex by at least taking into account the recipient's reputation ("second-order norms") [PB03, BS04, Sug86], or taking into account both the donor's as well as the recipient's reputation ("third-order norms") [SSP18, OI04]. For example, an ill-reputed donor helping another ill-reputed recipient may be judged as bad, whereas a well-reputed donor helping another well-reputed recipient may be judged as good.

In order to systematically explore which norms of indirect reciprocity are stable, Ohtsuki and Iwasa implemented an exhaustive search for social norms up to third order that are evolutionarily stable and can maintain cooperation in a population. Their landmark papers [OI04, OI06] employ some crucial assumptions. For one, the focus lies on deterministic assessment and action rules: given the donor's and recipient's current reputation, there is no ambiguity in how an observer updates a donor's reputation, or how a donor acts towards a recipient. Furthermore, reputations are binary: individuals are assessed either as "good" or "bad". Finally, reputations are public and synchronized: all members of the population update their assessment of donors in the same way, such that two different players never disagree on the assessment of a third party's action. This can be interpreted as a central authority assigning updated reputations. Ohtsuki and Iwasa identified eight particularly successful social norms of either second and third order, termed the "leading eight". These norms are able to maintain full cooperation and resist the invasion of defectors if the entire population employs them. The assessment and action rules of the leading eight norms share some commonalities, such that a donor with a good reputation who cooperates with a receiver with a good reputation should always obtain a good reputation (and thus should cooperate in this case), and that a defection against a receiver with good reputation should result in a bad reputation. The norms however employ different assessment of actions towards bad recipients. For example, while some norms always reward cooperation no matter the recipient's reputation, others like "Stern Judging" disincentivize helping an ill-reputed recipient.

The leading eight norms have been an important subject of study in indirect reciprocity [PSC06, SSP16, BS05, OIN09, SON17, SSP18, XGH19]. The premise that information is public and synchronized naturally facilitates analysis of the dynamics in the population. In more realistic scenarios, however, this assumption is somewhat unrealistic: individuals do not always agree on others' reputations, nor do they always have access to the same information. Previous work has shown that once players can make their own judgements in private, the success of the leading eight is not a given anymore [Uch10, US13, OSN17]. Rather, under private and noisy information, disagreements arise both due to some individuals not observing a certain interaction, or occasionally misinterpreting an action [HST+18]. Some of the most successful leading eight strategies under public information, like Stern Judging, then have the problem that such differences in opinion can have a lasting negative effect [OYU20]. Proliferation of these disagreements can lead to the emergence of separate communities within the population who consider each other as bad, creating a deadlock. This suggests that the leading eight norms therefore are not effective in letting cooperation emerge when information is private, noisy, and scarce.

Given this finding, the question of how to boost the performance of the leading eight in a noisy environment remains open. Previous approaches have ranged from finding new, potentially simpler and less problematic strategies of indirect reciprocity [OYS+18, SON17, CFW20, BG16, Ber11, SCHN21] to adding elements of empathy or "pleasing" behaviors to higher-order norms in order to overcome the negative effects of noise and private information [RSP19, KH21]. On the other hand, attempts to stabilize higher order assessment by introducing elements of "generosity" in analogy to direct reciprocity have been show to fail [SSHC21]. While stochastic strategies have been shown to have error-correcting properties that help stabilize cooperation in repeated games between the same pair of individuals, this notion of stochasticity does not transfer to the leading eight. Dropping the assumption of the leading eight being deterministic norms, and having their action and/or assessment rule include a probability of acting in a “forgiving” manner does not improve cooperation rates. Rather, if players using the leading eight occasionally cooperate instead of defect, and in particular if they occasionally assess actions as good that they would usually assess as bad, cooperation can even be further reduced.

In contrast, we show in the present work that dropping a different original assumption on the leading eight, namely the assumption of binary reputations, can mitigate the problems that arise through private and noisy information. To do so, we introduce “quantitative assessment” into higher order strategies, where reputations can take values beyond “good” and “bad”. This is a natural premise in many real-life situations, and is in line with the previous intuition that equipping players using the simpler Image Scoring norm with a larger range of reputation scores and threshold strategies can aid in staving off the effects of errors [BS04, NS98b, Oht04, LFS99]. More refined reputation scores can act as a buffer for errors, increasing the tolerance for a small number of actions that are regarded as bad by an observer [BG16, Ber11], without having negative consequences for the donor’s reputation and immediate future. Furthermore, work that investigates the effect of taking reputation and actions to be continuous variables on cooperation rates of the leading eight norms under private, noisy and incomplete information shows positive results using a local stability analysis [LMB21]. Replacing binary reputations and actions with probabilistic mixtures enables conditions for which cooperation can evolve. These findings and the promise they hold for more nuanced reputation systems are another indicator that the leading eight can benefit from quantitative, if discrete-valued, assessment.

In the following, we present a framework that incorporates quantitative assessment into the leading eight norms. In line with previous work, we consider players using deterministic rules for both assessment and actions. However, individuals now keep track of their fellow population members’ reputations in the form of integer scores, taking values from a fixed range. For an observer, positively assessing a donor’s action then means to increment the donor’s score, while a negative assessment translates into decrementing the score. Individuals also need to make an overall judgement of their co-players. This means that individuals must translate others’ integer reputation scores into a corresponding binary label, giving the context for the assessment and action rule of the social norm. For example, when it is a player’s turn to be the donor, he first needs to judge his co-player and himself as either good or bad. To do so, this player computes the appropriate labels by comparing their own and their co-player’s score to a threshold. A larger range of reputation scores therefore implies a greater insensitivity of overall judgment to single assessments. Our agent-based simulations suggest that this assessment model effectively works to correct disagreements between players using the same norm. Hence, players using a leading eight norm can more easily single out defectors while achieving a good self-image. This makes it more likely for four of the leading eight norms to be favored in evolution, and leads to high cooperation rates even at significant noise levels for these four norms. Our results thus hint at the power of more nuanced opinions and tolerance when interacting in a community using complex social norms. In a noisy environment with scarce information, rash judgments can irrevocably harm cooperation; it is instead a healthy dose of tolerance that can help protect reputations.

4.2 Results

4.2.1 Game dynamics

Similar to previous work, we consider a well-mixed population of fixed size N. The members of this population, who we refer to as players, are continuously engaged in pairwise interactions. In every round of this series, two players, a donor and a recipient, are drawn at random from this population for an interaction. The donor chooses whether or not to confer a benefit b to the recipient at his own cost $c < b$, i.e. chooses between cooperation and defection. Meanwhile, the recipient does not respond with an action of their own. The donor's choice is independently observed by other members of the population with probability q. They use these observations to privately update their opinion of the donor, without publicly sharing their judgment. Every player individually tracks the reputations of all other members of the population. We furthermore assume that players' observations are subject to noise: with probability ε, an observer mistakes a defection for cooperation, or a cooperative action for a defection. Given that information is thus assumed to be private and noisy, it is important to note that when two observers differ in their initial assessment of a given donor, they may also disagree on the donor's updated reputation, even if both observe the same interaction.

Every player is equipped with a social norm (corresponding to their "strategy") that governs their behavior, both in terms of judging actions as well as choosing an action towards a recipient. Social norms thus consist of two rules, an assessment and an action rule [Oka20]. The assessment rule prescribes how to update a donor's reputation after observing their action and taking this action's context into account. The action rule prescribes whether to cooperate or defect in a given situation.

In line with Ohtsuki and Iwasa's original work [OI04, OI06], we consider norms of at most third order. This means that the context that is taken into account for assessment can be both the donor's previous reputation and the recipient's previous reputation. For the action rule that decides whether to cooperate, the donor's own reputation and the recipient's reputation make up the necessary context. In our framework, reputations are not binary; instead players assign integer reputation scores r to each other that can take a wider range of values on a scale from a lower limit V to an upper limit A. That is, we equip higher order norms with quantitative assessment (**Fig. 4.1a**). While this leaves the considered leading eight norms $L1 - L8$ formally unchanged, the interpretation of the rules making up the norm and therefore the resulting reputation dynamics differ from the baseline model [HST+18].

We note that there are two components to our model of quantitative assessment. For one, the integer reputation scores are used by the players to track changes in opinion about others in a fine-grained manner. We assume that if an action is judged as good based on an observer's assessment rule, the donor's score in the eyes of that observer is increased by one, whereas if the action is judged as bad, the donor's score is decreased by one (**Fig. 4.1b**). On the other hand, players need to be able to translate these nuanced scores into an overall judgment of a player in order to use the leading eight norms as defined in previous work. This overall judgement must result in a binary label to become the input bits for players' assessment and in particular action rules (**Fig. 4.1c**). To make the transformation unambiguous, we assume that players compare the reputation score of a given player with a threshold S that separates scores into good and bad, similar to previous work on non-binary Image Scoring norms. An individual with a score equal to

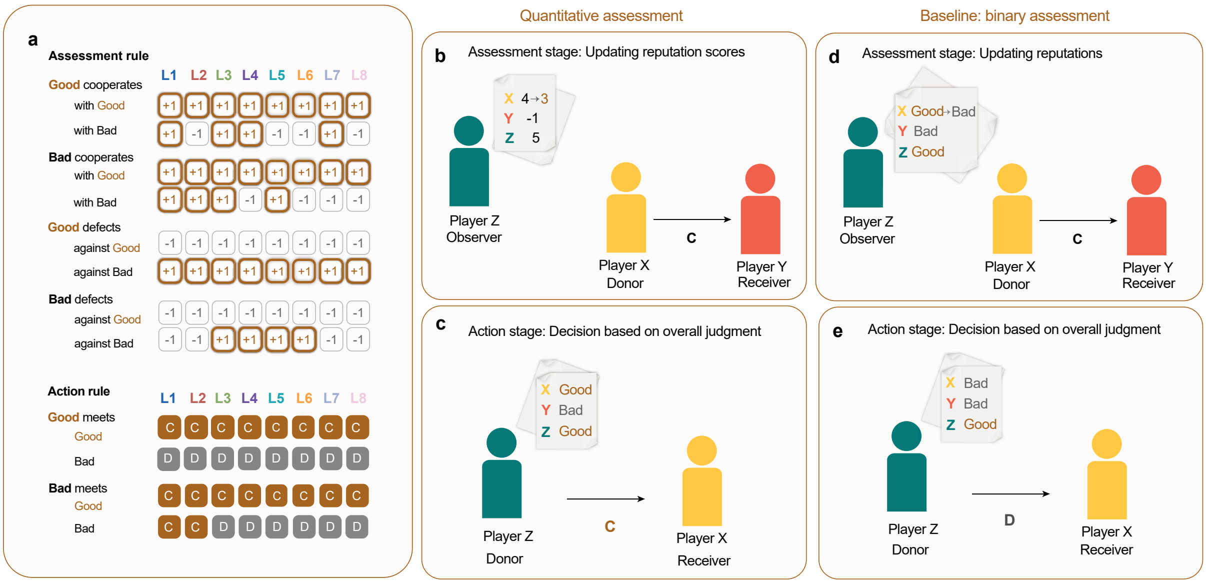

Figure 4.1: **The leading eight norms with quantitative assessment. a,** We consider the standard deterministic form of the leading eight norms [OI04, OI06]. Each norm consists of an assessment rule that determines how an observer updates a donor's reputation, and an action rule that governs players' behavior when they are chosen to be the donor. The assessment rule takes the context of an observed action into account: how an observer judges a donor depends on the donors' and recipient's reputation. Similarly, the action rule uses the current donor's and recipient's reputation to decide whether the donor should cooperate with the recipient. In contrast to the original baseline model, we now interpret a positive assessment of an action as an increment of +1 to the donor's reputation score, and a negative assessment as a decrement of -1. **b,** The observing Player Z assesses Player X's action of cooperating with a bad player as bad, such that he decrements X's previous score by one. **c,** When it is Player Z's turn to be the donor, he translates his and Player X's reputation score into a binary label of "good" or "bad". Since both his and Player X's score are above the threshold S, he judges both himself and Player X as good, and cooperates. **d,e** In the baseline model using binary assessment, the same starting scenario ends differently: Player Z changes his view of Player X from good to bad after Player X cooperates with Player Y, and therefore defects against Player X.

or above the threshold is judged as "good" overall, whereas a score below the threshold leads to an individual being judged as overall "bad". This means that the more refined the reputation scores, the less sensitive the overall label to single events, since this implies more shades of "good" and "bad". It is these overall reputation labels that are then used by players in order to correctly assess an observed action and to decide whether to cooperate with a specific recipient. We note that when we set $V = 0$, $A = 1$, $S = 1$, we recover the original baseline model of the leading eight with binary assessment (**Fig. 4.1d,e**).

In the following, we will make some assumptions about our framework. For one, all players use the same frame of reference for their assessment, such that everyone agrees on the assessment scale, i.e. the range of reputation scores. Hence, each individual uses the same fixed values for V, A, and S throughout. No player has a more nuanced view of their environment than another, and no player judges others' scores with a different measure than another. Furthermore, we define the possible range of reputation scores to be symmetric, with $A = R$, $V = -R$, and a threshold of $S = 0$ in the middle. We thus denote a particular setting of players' frame of reference with R. This underlying assumption naturally implies that there are slightly more possible ranks for a "good" player than a "bad" one, as a player with $r = S$ will be judged as "good". However, to test this bias, we have also run simulations where a player with $r = S$ will be judged as "bad", and have found no significant qualitative difference in results as long as players start with a good reputation.

4.2.2 Analyzing reputation dynamics

We start testing our framework by first considering players' norms as fixed, if possibly different. Individuals play the donation game over many rounds and update others' reputation scores after every interaction. We can describe this change in players' reputation scores with time-dependent image matrices [Uch10] $M(t) = r_{ij}(t)$, with $r \in [-R, R]$, representing the collection of players' reputation score repositories. These matrices thus record at any point in time which scores players assign to each other (**Fig. 4.2a, left side**). In every round of the game, $M(t)$ is updated according to the assessments of those players that have observed the donor's action, with their observation subject to noise ε. Noise in the environment implies that every observer has an independent probability ε to misperceive the donor's action. If player i assesses his observation of player j's action as good, $r_{ij}(t+1) = r_{ij}(t) + 1$, otherwise, $r_{ij}(t+1) = r_{ij}(t) - 1$. In case the reputation score is already at the maximum R (minimum $-R$), it simply keeps its value when the action is assessed as good (bad). To make the overall judgment of player j's and k's reputation, i compares their reputation scores with the threshold S, and labels them "good" when $r_{ij} \geq S$ and "bad" when $r_{ij} < S$. This acts like a second, less refined layer of the reputation dynamics (**Fig. 4.2a**, right side). For example, if player i judges donor j to be good and recipient k to be good, they will assess a defection of j against k as bad and decrease j's score by one.

We illustrate this concept with an example. We consider a population consisting in equal proportions of players using one of three social norms, a leading eight norm L_i, ALLD and ALLC (**Fig. 4.2b**). ALLD is the norm that prescribes to assign a bad reputation to everyone else and unconditionally defect, whereas ALLC prescribes to assign a good reputation to everyone else and unconditionally cooperate. For this population, we explore the reputation dynamics when information is noisy and scarce. In **Fig. 4.2c–j** , we present two snapshots of image matrices at the same timestep for each of the norms $L_1 - L_8$, where players use an assessment scale from -5 to 5. The first snapshot visualizes the integer reputation scores that players assign to each other at this particular time. The other snapshot represents the overall judgment that emerges when players compare others' scores with the given threshold ($S = 0$). The resulting binary label is then an indicator of how more refined assessment can help protect reputations from the negative effects of disagreements. We visualize the difference between the impact of the baseline binary assessment and quantitative assessment on reputation dynamics in **Fig. 4.3**. In this figure, we present the average images the three norms L_i, ALLC and ALLD have of each other, for each of the L_i.

We find that, compared to the baseline model, quantitative assessment helps all leading eight norms assign significantly more accurate reputations to population members. It clearly improves not only the self-image of each leading eight norm, but also enhances their ability to distinguish between the two other strategies they compete with. In particular,

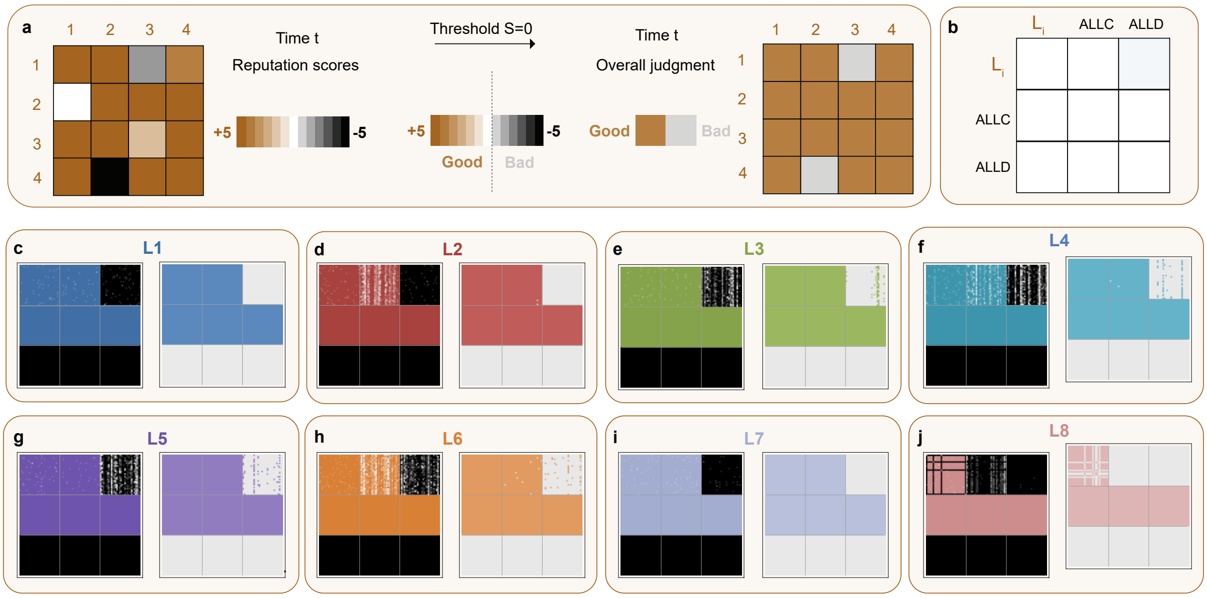

Figure 4.2: **Quantitative assessment and reputation dynamics. a,** Image matrices are representations of how players assess each other at any given time. We assume that every player keeps track of each population member's reputation score, with scores in the interval $[-5, +5]$. To depict these image matrices graphically, we use colored dots, with the intensity of the color corresponding to the score: for example, a white dot means means that the corresponding row player attributes a score of $r = 0$ to the corresponding column player (left side). On the other hand, players also make an overall judgment of others, in order to be able to use their assessment and action rules. To do so, they compare the scores to a threshold $S = 0$, resulting in a binary labeling of "good" and "bad". To visualize this second, less refined layer of the reputation dynamics, we use a matrix with colored and grey dots (right side). **b,** We show image matrices when players either use a leading-eight social norm L_i, *ALLC*, or *ALLD* (in equal proportions). **c – j,** We show the snapshots at $T = 2 * 10^6$ of players' reputation scores and binary labels they translate into for all leading eight norms. We see that for $L1$ (**c**) and $L7$ (**i**), the reputation assignments of different L_i players are perfectly correlated. They assign only good reputations to each other and ALLC, while they only assign bad reputations to ALLD. The picture is very similar for $L2$ (**d**). For all other norms, there are disagreements among the L_i players, where they can also perceive ALLD players favorably. We note that $L8$ does not perceive any ALLC player as good; however, this is one of two very stable states in the reputation dynamics. The other stable state sees $L8$ players have a favorable opinion of ALLC players. Parameters: We use a population of size $N = 90$, an error rate of $\varepsilon = 0.05$, and an observation probability $q = 0.9$. The frame of reference is $R = 5$, such that the interval for reputation scores is $r \in [-5, 5]$. Threshold $S = 0$. Simulations are run for $2 \cdot 10^6$ iterations, and the initial image matrix assumes a good reputation for all players.

singling out defectors becomes much easier to do. We find that $L1$ (**Fig. 4.3a**) and $L7$ (**Fig. 4.3g**) excel most at correct reputation assignment despite the presence of noise. The norm $L2$ (**Fig. 4.3b**) also does exceptionally well, even if it sometimes judges an ALLC player as bad (which however also happens without noise). Even $L3 - L6$ (**Fig. 4.3c–f**), which have particular trouble to accurately label defectors as bad in the baseline model, now do much better in singling out ALLD players in comparison to the binary assessment

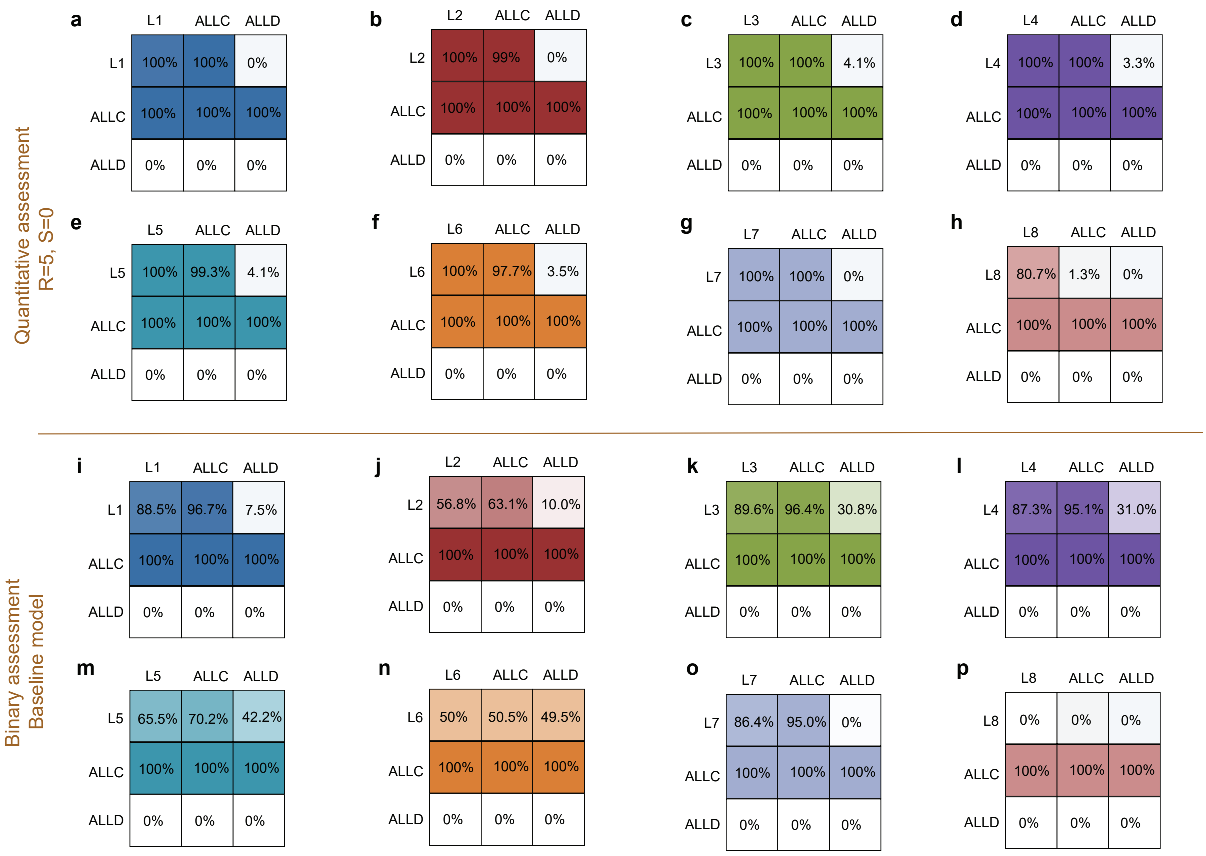

Figure 4.3: **Quantitative assessment significantly improves the accuracy of reputation assignments by leading-eight players.** We show the average overall judgments that players with frame of reference R make of each other when comparing others' reputation scores with the threshold (**a–h**). As basis for comparison to the baseline model, we use the average images that players have of each other when they use the standard binary assessment (**i–p**). We observe that quantitative assessment and more nuanced reputations lead to a significant improvement of the accuracy with which players assign each other images. All leading eight norms achieve a perfectly correlated good self image, as opposed to the baseline model, where only $L1$ (**i**) and $L7$ (**o**) achieve a self image of more than 80% good. Players using quantitative assessment also do much better in judging ALLD as bad, and with the exception of the (less stable) $L8$ (**h, p**), also manage to assess ALLC as close to 100% good. This hints at the power of a more refined reputation dynamics. Parameters are the same as in Fig.2.

model. Finally, $L8$ (**Fig. 4.3h**), which eventually judges everyone as bad under binary assessment, can judge its own kind as good much more easily when players use quantitative assessment, and accurately singles out defectors. Unconditional cooperators however are often assigned a bad reputation, due to $L8$ judging both good and bad players as bad when they cooperate with another bad individual. We note here that the averages for $L8$ however might not be fully accurate, as the reputation dynamics do not easily stabilize for this norm. In fact, the dynamics between $L8$, ALLD and ALLC have two very stable configurations: one where the norm judges itself and $ALLC$ as good, and one where it judges itself as around $70 - 80\%$ good and $ALLC$ as bad. Given the averages we have computed so far, the latter outcome is the significantly more predominant one and thus is shown and discussed here. To verify this, we are also running a longer experiment for $L8$.

From this, we can see that, compared with the binary assessment common in the literature, our results so far suggest that quantitative assessment indeed can stabilize reputations in a population with fixed composition. In fact, it can correct disagreements introduced by noise, such that populations quickly recover from erroneous reputation assignments and misjudgments. The outcomes thus come close to a setting with perfect information. However, this finding does not yet let us make a conclusive statement about how well the leading eight fare in evolution where population composition can change over time, and whether cooperation effectively emerges.

4.2.3 Evolutionary dynamics

We now aim to understand the effect of quantitative assessment when players' norms are not fixed, such that their abundance can change over time. We explore how likely it is for a leading eight strategy to evolve, and what cooperation rates are achieved in the population. Similar to our previous setup and to what is common in the literature on indirect reciprocity, we again consider a minimalistic scenario where players can choose from only three norms: a leading eight norm L_i, ALLD and ALLC. We assume that the evolution of social norms happens on a longer timescale that is separate from the reputation dynamics. This implies that the reputation dynamics have reached stationarity by the time that social norms change. Iterating the elementary process of reputation updating, we can not only calculate how often on average player i considers player j to be good, but also how often on average they cooperate with j. With the estimated pairwise cooperation rate $\hat{x}_{ij}$, with which player i helps player j, we can define the payoff of player i using a fixed strategy as $\pi_i = \frac{1}{N-1}\sum_{j\neq i} b\hat{x}_{ji} - c\hat{x}_{ij}$. To model how players adopt new strategies, we then consider simple imitation dynamics [TPN07, WBGT15, SSHC21]. In every timestep of the evolutionary process, a player i is picked uniformly at random to revise their norm. With probability μ they pick a random new norm. With probability $1-\mu$, they randomly choose a role model j to imitate wth a probability $P(\pi_i, \pi_j)$ depending on the difference between the two players' payoffs. This probability takes the form of the Fermi function $P(\pi_i, \pi_j) = (1+\exp[-s(\pi_j - \pi_i)])^{-1}$. The parameter $s \geq 0$ describes the strength of selection, which measures how relevant payoffs are for updating strategies. For $s = 0$, updating happens at random, and as the parameter increases, norms with higher payoffs are more likely to be imitated. We note that for imitation processes to be a reasonable model of strategies spreading, we implicitly assume that people discuss their world views and moral guidelines with others.

The resulting process is ergodic due to the possibility of random mutations in every timestep. Thus, it gives rise to a unique stationary distribution, called selection-mutation equilibrium. This distribution represents the abundance of each strategy in the long run. To calculate the average cooperation rate in the population, the payoffs of the individual strategies are then weighted with this equilibrium abundance. In the following, we will assume that mutations are rare [FI06, WGWT12], which implies that populations are homogeneous most of the time: a new mutant only arises when the previous mutant has either gone extinct or has fixed in the population. We can calculate the fixation probability of a mutant into a resident population with social norm R explicitly [TH09], cf. also Chapter 2.

Fig. 4.4 visualizes the evolutionary dynamics between each leading eight strategy, ALLC and ALLD. We find that in four cases, for $L3 - L6$ (Fig. 4.4c –f), the leading eight norm does not evolve. Once players have learned to use ALLD, there is very little chance of

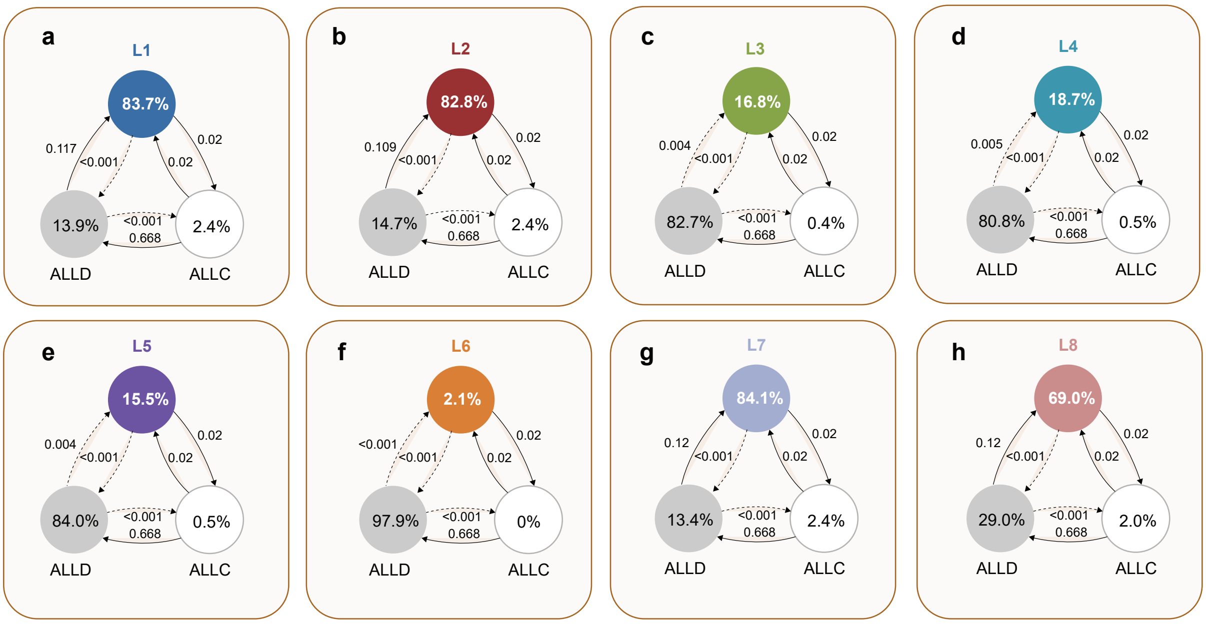

Figure 4.4: **Evolution of the leading eight when assessment is more refined.** We show the results of simulating evolutionary dynamics when players can choose among three different norms: a leading eight norm, *ALLC*, and *ALLD*. We assume that the spread of social norms is described by a pairwise comparison process[TPN07], such that norms of players with high payoffs are more likely to be successful. Here, we use the limit of rare mutations, such that populations are homogeneous most of the time[FI06, WGWT12, McA15]. Numbers in circles show how often each social norm is adopted on average. Arrows indicate fixation probabilities, i.e. how likely it is for other social norms to invade a given resident population. Solid arrows indicate that the respective transition is more likely to occur than expected under neutrality, whereas dotted arrows indicate that the respective transition is comparably unlikely. We see that four of the eight considered norms, $L1$ (**a**), $L2$ (**b**), $L7$ (**g**), and $L8$ (**h**) achieve significant abundance in equilibrium, with $L1$, $L2$ and $L7$ played over 80% of the time. The remaining four norms do not evolve in significant proportions, and the respective dynamics strongly favor ALLD. Parameters: $R = 5$, $S = 0$, $N\!=\!50$, $\varepsilon\!=\!0.05$, $b\!=\!5$, $c\!=\!1$, $q\!=\!0.9$, using a strength of selection of $s\!=\!1$.

reestablishing cooperation. On the other hand, three of the leading eight, $L1, L2, L7$ (**Fig. 4.4a,b,g**) are more than 80% abundant in equilibrium, which means that their evolution is strongly favored.

$L8$ is also played almost 70% over the course of evolution. This stands in stark contrast to the baseline model, where only $L2$ achieves an equilibrium abundance of that magnitude (**Fig. 4.5b**), and where most leading eight strategies do not evolve in meaningful proportions. We can see that this difference is mainly due to a significant probability of $L1, L2, L7$, and $L8$ invading ALLD (**Fig. 4.4a,b,c,g**). ALLC and each leading eight norm are approximately neutral with respect to each other (their fixation probability into one another is $\approx 1/N$ each), and ALLC is invaded by ALLD. Thus, the fixation probability of the leading eight norm into ALLD becomes the deciding factor. This also suggests that the power of quantitative assessment lies mainly in its ability to correct errors in the norm's self image by effectively reducing noise in the population, hence enabling a more accurate judgment of defectors. We can see this in the behavior of $L8$ (**Fig. 4.5h**), which

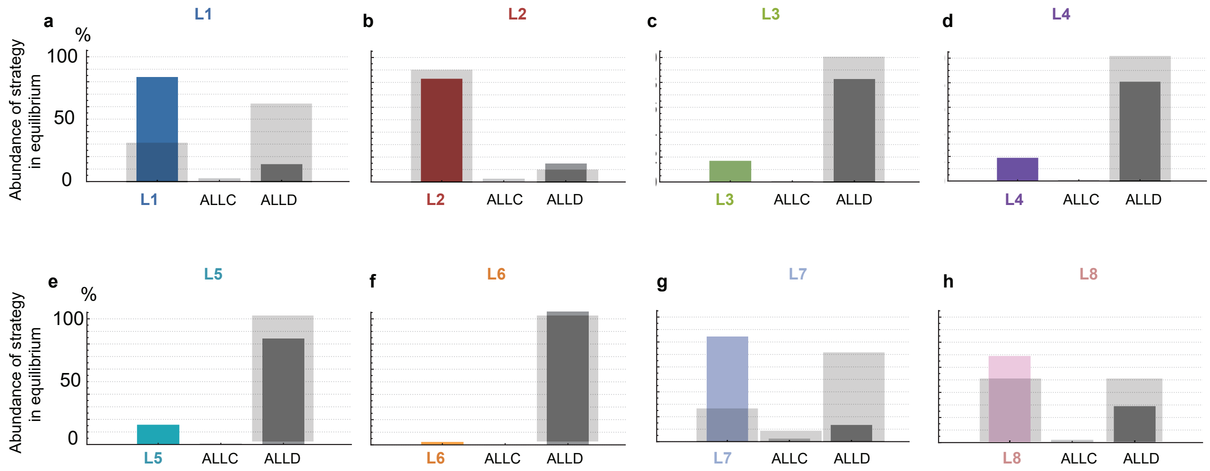

Figure 4.5: **Four of the leading eight evolve in significant proportions for quantitative assessment.** We compare the abundance of the leading eight strategies in selection-mutation equilibrium between the case of quantitative assessment and the baseline model. We use the same evolutionary process and setup as in Fig.3 and present the changes in how often each norm is played on average. Colored bars represent the abundance in equilibrium under quantitative assessment, while the light grey bars in the background of each panel represent the results in the baseline model. We find that four of the eight strategies now evolve much more readily (**a,b,g,h**) than in the baseline model, and are played in significant proportions. The three remaining strategies, which do not evolve at all in the baseline model, only do slightly better due to still being outcompeted by ALLD. Parameters are the same as in Fig.3.

is less abundant compared to e.g. $L7$ due to the fact that it does not always judge itself as good in the presence of ALLD.

Meanwhile, $L3, L4, L5, L6$ are unsuccessful in evolution, as they have difficulties labeling defectors as bad already in a non noisy environment, and therefore do not profit as much from the error-correcting quality of quantitative assessment. This strongly hints at the conclusion that for a higher order norm to be stable under noisy and private information, it must at least negatively assess a bad player defecting against another bad player.

Fig. 4.6 then visualizes how the eight norms' abundance in equilibrium translates into cooperation rates. We see that compared to the baseline model (**Fig. 4.6d–f**), seven of eight norms lead to increased cooperation, with only $L6$ completely failing to evolve cooperative behavior.

$L1, L2, L7$ give rates of almost 90%, whereas $L8$, which leads to no cooperation at all in the baseline model, also is significantly boosted. These findings are highly robust when we vary parameters (**Fig. 4.6a–c**), most remarkably even when we vary the error rate (**Fig. 4.6a**). Even at a noise level of $\varepsilon = 0.1$, cooperation evolves and is maintained at over 80% for $L1, L2, L7$, while it also does not fall below 60% for $L8$. This is due to both the high abundance of these four leading eight norms when players use quantitative assessment, and to an increased self-cooperation rate in homogeneous populations even when errors are more frequent. Quantitative assessment thus changes the behavior of the leading eight norms' cooperation rates significantly.

Finally, we explore the effect of the size of the assessment scale, i.e. the range of the reputation scores, on cooperation (**Fig. 4.7**). The case of two possible ranks (reputation

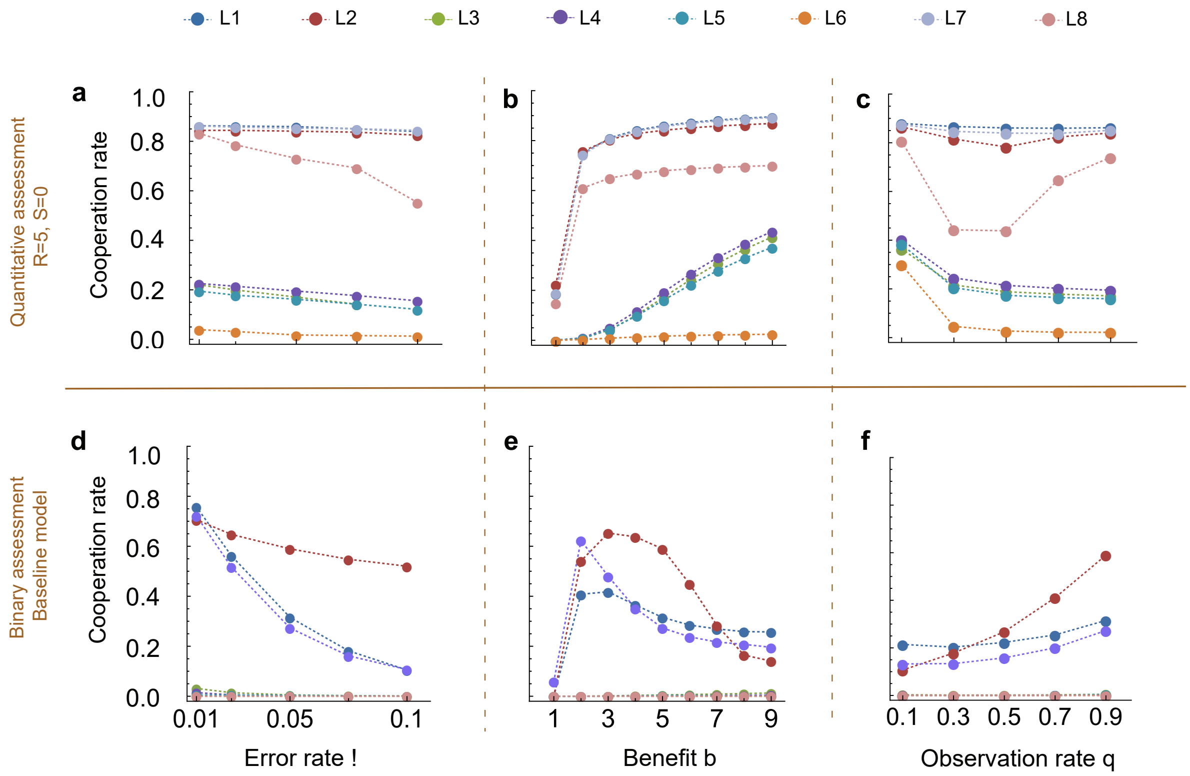

Figure 4.6: **Quantitative assessment has a profound impact on cooperation rates.** We vary the noise on observations ε, the benefit-to-cost ratio b/c, with $c = 1$, and observation probability q. All other parameters remain constant at the values of Fig. 3. In each scenario, we plot the average cooperation rate of each individual leading eight norm when they compete against ALLD and ALLC, according to the selection-mutation equilibrium of the evolutionary process. We can compare the results when players use refined assessment with $R = 5$ ($\mathbf{a} - \mathbf{c}$) with the outcome of binary assessment in the baseline model[HST$^+$18] ($\mathbf{d} - \mathbf{f}$). **a,** Under quantitative assessment, cooperation rates of $L1$, $L2$, and $L7$ remain at around 85% even when the error rate ε increases to 0.1. The generally more unstable $L8$ is more affected by the increased noise, but still remains above 50% even at $\varepsilon = 0.1$. **b,** Increasing the benefit of cooperation b leads to an increase in cooperation rate for all eight considered norms in contrast to the baseline. **c,** When we increase the observation probability q, the behavior of the leading eight norms' cooperation rates is also markedly different from the baseline. $L1, L7$ are barely affected while $L2$ and $L8$ exhibit nonlinearity for intermediate values of q.

values) corresponds to binary assessment. Notably, we find that for $L1, L2, L7, L8$, cooperation rate non-monotonically increases with the number of possible reputation scores (**Fig. 4.7a**). In fact, choosing an assessment scale smaller than taking $R = 5$, as previously used, leads to even more cooperation, especially for the more unstable $L8$ with its strict assessment rule. **Fig. 4.7b** suggests that an intermediate number of levels gives the four successful social norms a higher abundance, due to a higher probability of invading $ALLD$.

The norm $L2$ is also a special case: here, going from binary assessment to quantitative assessment with $R = 1$ harms the cooperation rate, as abundance drops down to only 46% from 89%. These observations suggest that there is a tradeoff between the size of the assessment interval and the increased self-cooperation rate going towards 100% that

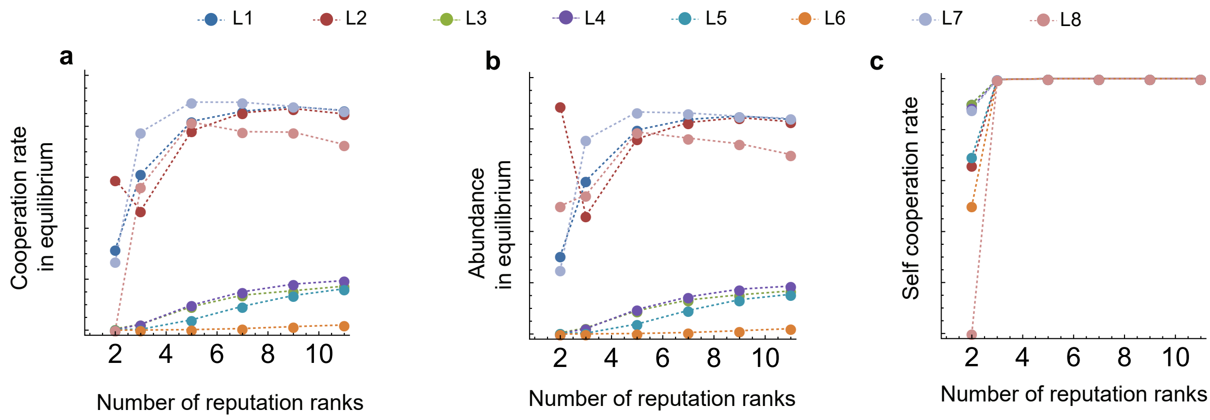

Figure 4.7: **Varying the frame of reference for quantitative assessment.** For this figure, we repeat the evolutionary simulations shown in Fig.3, and vary the frame of reference R. That is, we explore the impact of the number of possible reputation ranks on cooperation, including the case of binary assessment with two reputation ranks. **a,** We show the cooperation rate in equilibrium for the leading eight norms as the number of reputation ranks increases. We note that for the four successful norms $L1, L2, L7, L8$, the largest frame of reference does not correspond to the highest cooperation rate. An intermediate number of ranks is the most beneficial. $L2$ also exhibits a drop in cooperation rate from binary assessment to $R = 1$ (i.e. 3 reputation ranks). The behavior of the cooperation rates is mainly determined by the behavior of the equilibrium abundance of the eight norms as the frame of reference varies (**b**). Meanwhile, self-cooperation rates quickly increase to 1 as the frame of reference increases (**c**), which implies that leading-eight players have a perfectly correlated image of each other once assessment is more nuanced. Parameters are the same as in Fig.3.

comes with more fine grained reputation scores (**Fig. 4.7c**). Intuitively, an increased range for reputation scores can act as a buffer that not only protects good reputations from misjudgments, but can also make it harder to quickly and durably correct mistakes. In turn, this can lead to sunk costs when players cooperate too often with defectors. This becomes a particular issue for $L2$, as it is the only norm to actually evolve in a large proportion in the baseline model, but also is sensitive due to its negative assessment of good players' cooperation with bad players. Therefore, the net effect is negative when going from binary assessment to quantitative assessment with $R = 1$. The other norms do not exhibit this drop in cooperation rates when moving to $R = 1$, as they profit far more from the increased error correction, such that the negative effect of larger reputation intervals only becomes apparent later on.

This shows that changing the inner workings of the reputation updating has a profound effect on the overall dynamics by effectively minimizing the effect of noise.

4.3 Discussion

Indirect reciprocity explores how people form good reputations, and how social norms evolve [JHTM11, Feh04, BKO05, Now06]. It approaches fundamental moral questions, for example, how a good person should act and how goodness should be defined in the first place, from the viewpoint of evolutionary biology [Ale87]. A crucial question in indirect reciprocity is how social norms can sustain cooperation when individuals observe others' interactions with third parties and judge them according to their own moral standard and idea of goodness [Sig10, Sig12, OI04]. Previous work has suggested that to reliably maintain cooperative behavior, a social norm needs to be sufficiently complex: simple rules like Image Scoring, which tie the idea of "good" and "bad" individuals merely to their actions without considering the context of these actions, have been shown to be unstable in a population [PB03, BS04, LH01]. Ohtsuki and Iwasa proposed the so-called "leading eight" social norms as the most successful ones to maintain stable cooperation [OI04, OI06]; these norms contextualize actions with the binary reputations of the interaction partners, and are therefore able to differentiate between justified defections and unjustified defections. Yet, they also were discovered to be unstable in the presence of noise when players do not share information between them and privately keep track of others' reputations [HST+18]. Disagreements between individuals using the same norm can arise in this case, fragmenting the population into subgroups and breaking down cooperation.

Here, we have proposed a way to overcome this challenge by using a more refined assessment system in the leading eight strategies, where individuals privately keep track of others' reputation in the form of integer scores that can take a range of values. Positive assessment of an individual's action by an observer then translates into a score increment, whereas negative assessment decrements the acting individual's score. This implies that "goodness" is reinterpreted in a much broader way. Whereas with binary assessment, there is only one "good" and one "bad" label coinciding with the assessment of an action, individuals now use a threshold to evaluate a peer's reputation score and thus judge them. Similar ideas have also been proposed in research on the Image Scoring norm to reduce its vulnerability to errors [Ber11, BG16, NS98b]. Our results now suggest that this error correcting capability works very strongly in favor of the leading eight strategies. Disagreements between individuals using the same strategy no longer are the same threat to cooperation. Rather, quantitative assessment acts as a buffer for misjudgments due to noise and lapses in observation. This protects reputations and enables norms to have both a stable good self image as well as a stable negative image of defectors. Overall, this leads to significant improvements in cooperation rates for four of the leading eight norms, even when noise in information increases. In contrast to a quantitative version of image scoring [NS98b], a quantitative version of the leading eight cannot tempt players to simply try and maintain their own score just at the threshold regardless of their partners' reputations: given that observers' assessments take their private image of the interaction partners into account, players cannot calculate a minimum effort to only just remain "good" in the eyes of others.

Ohtsuki and Iwasa's landmark papers identified certain properties of a higher order norm that need to be fulfilled for a norm to be successful in evolution. Cooperation must be maintained by assessing cooperation of good players with other good players as good. Furthermore, defectors must be recognized, meaning that defection against a good player must be assessed as bad. There must also be a way to punish defectors, by assessing a defection of a good player against a bad player as good. Finally, forgiveness must be possible: the cooperation of a bad player with a good player should restore their reputation. All other bits of the norms are left flexible. Our results meanwhile suggest one more criterion that should be added for the case of private and noisy information: norms cannot be "gullible" for cooperation to be maintained. This means that they should never assess a bad player defecting against another bad player as good (making them easily deceived by false shows of solidarity), which fixes another bit of the assessment rule. In fact, we have seen that those four of the leading eight that are gullible do not evolve even in the simple setting we have explored.

It might come as little surprise that nuanced reputations and tolerance for a few negative experiences with others help to resolve disagreements and maintain cooperation, whereas a setting where individuals have binary opinions about each other and are very sensitive to single events ends up favoring defectors. This result is intuitive from a psychology perspective: So-called dichotomous thinking[Osh09] - thinking in simple terms of binary opposition instead of seeing shades of grey - is assumed to be beneficial for quick decision making and taking control of situations, but at the same time has been found to be a cognitive distortion correlated with personality disorders [NM07, BDF15, YD05, Osh12]. This bias is particularly prevalent in Cluster B and C disorders such as borderline personality disorder and narcissistic personality disorder, which are known for destructive tendencies in interpersonal relationships and/or difficulties maintaining bonds with others [LZS+04, VA00, SATJ82]. Dichotomous thinking also has been studied as a hallmark bias of traits of antisocial behavior [JOS+18]. Particularly simplistic world views with no tolerance for single missteps (“You are either with us or against us”) have further shown themselves to be harmful on a larger scale, which can be seen e.g.in extreme political partisanship [LJS12] and the legislatory deadlocks and social fractures it can evoke.

We should however note here that the positive results that we have presented here have been observed in a restricted setup. They naturally cannot be automatically generalized to more complex scenarios where mutation rates are higher, or where the leading eight norms have to compete with mutants besides just ALLD and ALLC. For example, higher mutation rates might affect the performance of $L8$, which we have observed to be less stable in its reputation dynamics when both ALLC and ALLD are present in the population. This issue is the subject of ongoing work. Also, one can assume that population structure [LHN05, ON08] can play a significant role in the question of how well quantitative assessment can diminish the effect of errors. For very sparse topologies, it is not trivial to answer this question as too much buffer for negative experiences can become a hindrance when a player cannot observe certain actions at all. These limitations notwithstanding, our present findings still clearly show that the main problem that previous work has pointed out, i.e. the fact that the leading eight strategies overall cannot maintain cooperation against defectors when information is private, noisy and imperfect, can in fact be overcome with more fine-grained assessment.

Overall, our study thus serves to highlight how broader definitions of goodness and nuanced thinking can benefit a community using more complex social norms, and play an important role in maintaining positive social relations.

Indirect reciprocity describes the role of reputation and social norms for the evolution of cooperation. This mechanism assumes that when individuals interact with someone, they are subject to observation and judgement by their peers. Helping someone at own cost increases individuals’ reputation, which benefits helpers in future interactions and therefore can sustain cooperative behavior. Previous literature has identified eight highly successful social norms, the “leading eight”. However, these norms require all relevant information to be reliable and public to be effective. Otherwise, the resulting disagreements can proliferate and prevent cooperation from being sustained. Studies on the performance of the leading eight usually assume that reputations are binary: players can be thought to be either good or bad. In contrast, here we study the case where reputations are more nuanced. With this “quantitative” assessment, every individual privately keeps track of others' integer reputation scores, incrementing and decrementing these scores according to observations and norm-specific rules. To make an overall judgement of a peer, an individual then compares the corresponding score with a threshold. Our work suggests that such quantitative assessment can act as a powerful stabilizer of cooperation for four of the leading eight strategies when they compete with defectors and cooperators. Keeping track of reputations in a more nuanced way has error-correcting properties, thus increasing the resilience against the effects of noise by enabling recovery from disagreements. For cooperation to be reliably sustained by complex social norms, such comparably more sophisticated assessment and a broader definition of goodness thus seems to be necessary in order to overcome the challenge of noisy environments with incomplete information.

CHAPTER 5

A unified framework of direct and indirect reciprocity

Direct and indirect reciprocity are key mechanisms for evolution of cooperation. Direct reciprocity means individuals use their own experience to decide whether to cooperate with another person. Indirect reciprocity means they also consider the experiences of others. Although the two mechanisms are intertwined, they are typically studied in isolation. Here, we introduce a mathematical framework that allows us to explore both kinds of reciprocity simultaneously. We show that the well-known strategy 'Generous Tit-for-Tat' of direct reciprocity has a natural analogue in indirect reciprocity, which we call 'Generous Scoring'. With an equilibrium analysis, we characterize under which conditions either of the two strategies can maintain cooperation. With simulations, we additionally explore which kind of reciprocity evolves when members of a population engage in social learning to adapt to their environment. Our results draw unexpected connections between direct and indirect reciprocity, while highlighting important differences regarding their evolvability.

5.1 Introduction

Reciprocity is a principle that guides many aspects of our social life [Tri71, Sug86, Now06, Sig10]. Whenever people repay a favor, write a positive evaluation of an online seller, or build up trust over multiple interactions, they engage in reciprocal behavior. Previous work distinguishes two kinds of reciprocity. Direct reciprocity [AH81, NS92, HS97, PD12, HNS13, SP14a, SP14c, Aki16, PHRZ15, HRZ15, MH16b, IM18, HCN18, GvV18, RHR+18] means that my behavior towards you depends on what you have done to me. Indirect reciprocity [NS98b, LH01, OI04, SSP18, Sig12, NPSH15] means that my behavior towards you also depends on what you have done to others. Direct reciprocity requires that the same individuals interact repeatedly, which enables them to respond to their interaction partner in future transactions (**Fig. 5.1a**). Indirect reciprocity does not require individuals to have a joint history of previous interactions, nor does it require them to ever meet again. It is solely based on the premise that by helping someone, you can increase your public standing. This reputational gain is valuable in future interactions with others (**Fig. 5.1b**). Experiments suggest that human behavior is shaped by both direct [FGF01, GGLM+14] and indirect reciprocity [WM00, OYS+18, MvdBE13].

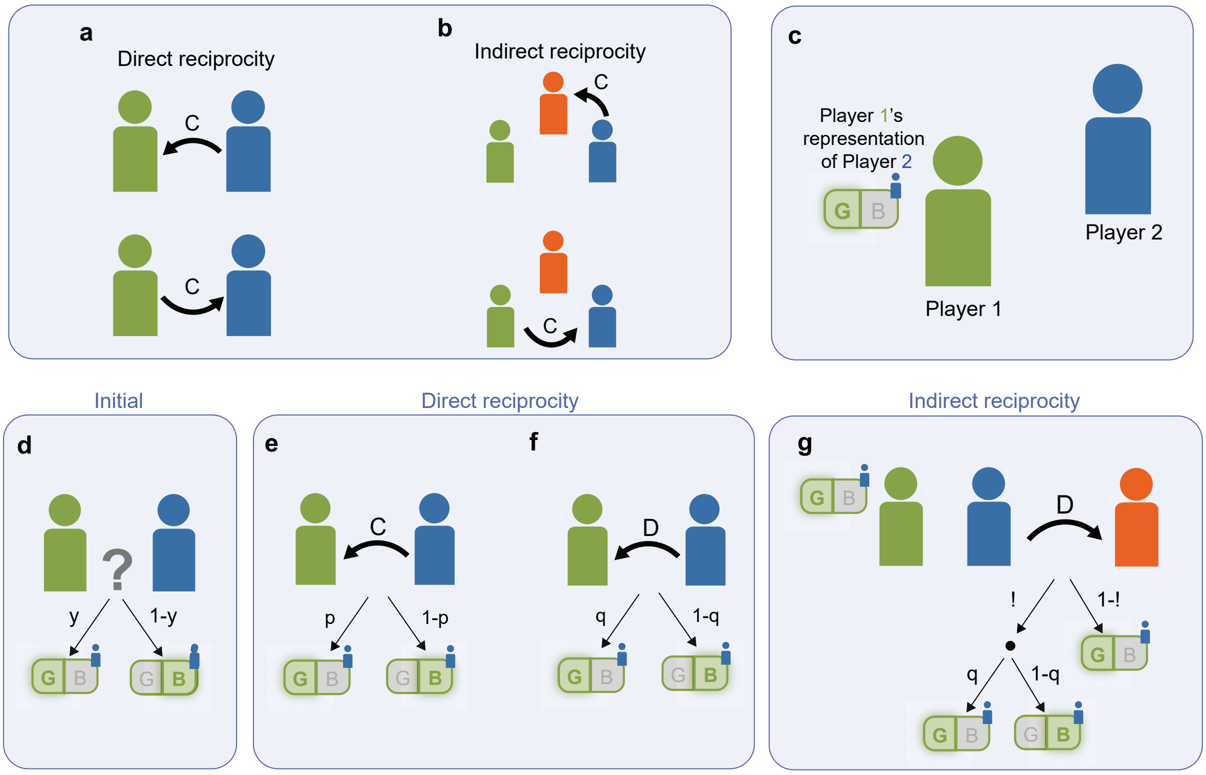

Figure 5.1: **A unifying framework for direct and indirect reciprocity. a,** Under direct reciprocity, an individual's cooperation is returned directly by the beneficiary. **b,** Under indirect reciprocity, cooperation is not returned by the beneficiary, but by some observer. **c,** To model direct and indirect reciprocity we consider individuals who assign one of two possible reputations to their co-player, good (G) or bad (B). The current assignment is highlighted in green. Individuals cooperate (C) with those co-players they consider as good, and they defect (D) against those they deem as bad. **d–g,** Whether an individual considers a co-player as good depends on her strategy (y, p, q, λ). Here, y is the initial probability to assign a good reputation to the co-player, without having any information; p and q are the probabilities to assign a good reputation after the co-player has cooperated or defected in a direct interaction, respectively. The receptivity λ is the probability with which an individual takes third-party interactions of the respective co-player into account. For $\lambda = 0$, we obtain a model of direct reciprocity. For $\lambda = 1$, we obtain a model of indirect reciprocity. While the illustrations depict one-way interactions for simplicity, our model considers two-way interactions. When two players are chosen to interact, they both decide simultaneously whether to cooperate or defect. All other population members observe their choices.

While direct and indirect reciprocity are related, the respective models are strikingly different. Studies of direct reciprocity [AH81, NS92, HS97, PD12, HNS13, SP14a, SP14c, Aki16, PHRZ15, HRZ15, MH16b, IM18, HCN18, GvV18, RHR$^+$18] report substantial cooperation rates even if subjects only remember a minimum of information. Successful strategies like Tit-for-Tat [AH81] (TFT) and Generous Tit-for-Tat [NS92] ($GTFT$) only keep track of the very last interaction. In contrast, studies of indirect reciprocity stress that cooperation can only be maintained when strategies are sufficiently complex [LH01, OI04, SSP18]. To describe how complex strategies need to be, this literature distinguishes different classes of strategies. The most elementary class are the first-order strategies, where a player's reputation only depends on her previous actions. A well-known example is image scoring [NS98b]. Here, reputations are represented by an integer score. A player's score increases when she cooperates, and it drops when she defects. Individuals only cooperate with those who have a sufficiently high score. Classic image scoring, however, is unstable [LH01]. After all, individuals have no incentive to retaliate against defectors, because this would impede their own score. This instability suggests considering second-order strategies. Here, reputations do not only depend on what an individual did, but also to whom. For example, when an individual defects against a co-player with a bad reputation, this defection may be considered as justified. The hierarchy of strategies can be further extended to third order. Here, players additionally take into account the focal individual's reputation.

In a landmark study, Ohtsuki & Iwasa explored which strategies of up to third order sustain cooperation [OI04]. In their study, reputations are required to be binary (good or bad), strategies are deterministic (the same behaviour always yields the same reputation), and all information is public and mutually agreed upon. Within this setup, they show there are no stable first-order strategies that give rise to cooperation. However, there are two second-order strategies and six third-order strategies. These so-called leading eight strategies can sustain cooperation because they allow for more sensible judgments than image scoring. At the same time, they require more information than most well-known strategies of direct reciprocity.

The two kinds of reciprocity also differ in how susceptible they are to misunderstandings and other types of errors. Whereas *GTFT* and similar strategies of direct reciprocity are robust with respect to noise [Sig10, HMVCN17], the leading eight strategies of indirect reciprocity are not[US13, HST$^+$18].

Due to such differences, it has been difficult to analyse the two modes of reciprocity within a single theoretical framework. Previous work has taken two approaches. The first approach is to suggest particular strategies that combine elements of direct and indirect reciprocity, and to analyse their stability [RW90, PD92]. The second approach uses computer simulations to let different strategies compete [Rob07, NK04, SN16]. Two noteworthy studies of the latter kind are by Nakamaru & Kawata [NK04] and Seki & Nakamaru [SN16]. They explore the evolution of reciprocity when players can fake their own reputation, or misrepresent the reputation of others. The two studies observe that when outside information becomes unreliable, players tend to ignore it. Computational studies, however, make it difficult to compare the different kinds of reciprocity directly. They often involve comparisons between strategies of different complexity. Moreover, the relative advantage of each type of reciprocity can only be inferred by comparing simulations for specific parameter choices.

Instead, here we propose a framework that can be analysed explicitly. For our study, we extend the theory of zero-determinant strategies from direct [PD12, HNS13, SP14a, SP14c, Aki16, PHRZ15, HRZ15, MH16b, IM18] to indirect reciprocity. This approach allows us to draw analytic conclusions when comparing the two mechanisms.

5.2 Results

5.2.1 A unified framework of reciprocity

To introduce a model that entails both kinds of reciprocity, we consider a population of n players. Players engage in the following sequence of interactions. In the beginning, two players are randomly drawn from the population to interact in one round of the prisoner's dilemma. In this game, each player independently decides whether to cooperate (C) or defect (D). Cooperation means to pay a cost $c>0$ to provide a benefit $b>c$ for the co-player. After each such interaction, with probability d again two players are randomly drawn from the population to engage in another round of the game. Otherwise, with probability $1-d$, no further interaction occurs. Once there are no more interactions, we calculate the payoffs of each player by averaging over all pairwise games in which the respective player participated.

To make their decisions, players represent each co-player by a separate finite-state automaton. Each automaton has two possible states, labeled as Good (G) and Bad (B), see **Fig. 5.1c**. Players cooperate with those co-players they currently deem as good, and they defect against those they consider as bad. They update the current state of each co-player according to their strategy (y, p, q, λ). The parameter y is the initial probability that a co-player is considered as good, in the absence of any information (**Fig. 5.1d**). The parameters p (and q), determine the probability to assign a good reputation to a co-player who has just cooperated (defected) in a direct interaction (**Fig. 5.1e,f**). The parameter λ is a player's receptivity to indirect information. If a co-player interacts with a third party, then with probability λ the focal player updates that co-player's state accordingly. In that case, again the co-player obtains a good reputation with probability p or q, depending on whether she cooperated or defected (**Fig. 5.1g**, **Fig. 5.2**). For simplicity, we assume in the main text that all population members observe everybody else's interactions. However, they may misinterpret the outcome of games between others with probability ε. When such an observation error occurs, a third-party observer mistakenly interprets a player's cooperation as defection, and vice versa.

If $\lambda=0$ for all individuals, players base their decisions entirely on their own experiences. In that case our framework reduces to the standard model of direct reciprocity with reactive strategies[NS92]. On the other hand, if $\lambda=1$ for all individuals, then players take all interactions of their opponents equally into account, no matter whether they are directly involved. In that case our framework yields a model of indirect reciprocity among players with stochastic first-order strategies[Oht04]. It is important to note that even players with $\lambda=1$ do not ignore any directly obtained information they may have. For example, if the same two individuals are chosen to interact for two consecutive rounds, their second-round behavior will naturally depend on the outcome of the first round. However, in large populations in which such consecutive encounters are unlikely, the role of direct information on players with $\lambda=1$ becomes negligible. In **Section 5.5.6**, we compare this baseline model with an alternative setup, where we consider a 'purified' version of indirect reciprocity. In that alternative setup, players can choose to ignore all direct experiences they have, such that they solely rely on third-party information. The results of that alternative model are similar to the results presented herein.

Our strategy space contains several well-known strategies of direct and indirect reciprocity. Examples include $TFT=(1,1,0,0)$, $GTFT=(1,1,q,0)$, and an elementary image scoring rule[NS98a] referred to as Simple Scoring[Ber11], $SCO=(1,1,0,1)$. However, our model is more general than these previous studies on either direct or indirect reciprocity in two ways. First, it allows for populations in which some players use direct reciprocity ($\lambda=0$), whereas others use indirect reciprocity ($\lambda=1$). Second, it allows players to combine the two modes of reciprocity, by choosing $0<\lambda<1$. In that case, players always take direct experiences into account, but they would occasionally also consider a co-player's

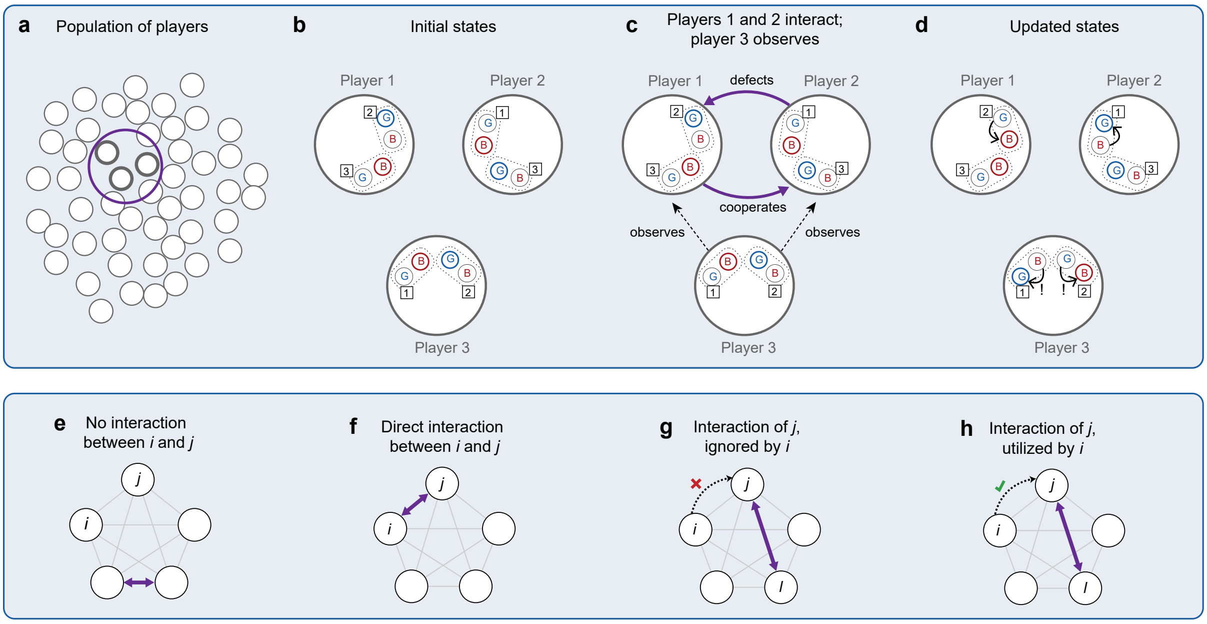

Figure 5.2: **Schematic representation of the model. a,** We consider a population of size n. To illustrate the basic workings of our model, we focus on three arbitrary players that are fully interchangeable in all their abilities. **b,** Each player has a separate finite-state automaton with two possible states G and B for each co-player. The current state is marked in bold. In this example, player 1 considers player 2 as good and player 3 as bad. **c,** In each round, two players are chosen at random to interact in a prisoner's dilemma. Players cooperate if they consider their co-player to be good and they defect otherwise. The other population members do not participate in the game, but they observe its outcome at no cost to themselves. **d,** After the interaction, both active players update their respective automata, depending on their strategy and on the co-player's action. In addition, each observer independently updates her automata with respect to players 1 and 2 with probability λ each. **e–h,** We can mathematically describe how player i's automaton with respect to player j changes over time by distinguishing four possible events. First, player j is not chosen to interact, such that player i's automaton remains unaffected (**e**); second, players i and j interact with each other and update their respective states accordingly (**f**); third, player j interacts with someone else, but player i does not take this interaction into account (**g**); fourth, player j interacts with someone else, and player i updates j's state accordingly (**h**).

interactions with others.

Models of indirect reciprocity often assume 'public information'[OI04, SSP18]. This does not only mean that all individuals learn all relevant information. Instead, respective models also assume that everyone agrees on each co-player's reputation. Such an assumption can be problematic when individuals receive information from independent sources, or when information transmission is noisy[HST+18]. Moreover, even if individuals agree on all past events, they may still disagree on which reputation a co-player should have if they apply different social norms. Such different assessments can easily arise, for example, when some players base their decisions on direct reciprocity, whereas others use indirect reciprocity. Because we are exactly interested in such scenarios, our model is necessarily one of 'private information', as in Nakamaru & Kawata[NK04] and Seki & Nakamaru[SN16]. As a result, different players may hold different views on any given population member.

Throughout the main text, we will use the above baseline framework to explore the dynamics of direct and indirect reciprocity. However, in **Section 5.5.6**, we explore the effect of several model extensions. In particular, we discuss how our results change when we allow for alternative kinds of errors[BS04, Uch10, MVC13] and for incomplete information[NM11]. Moreover, we describe how our framework can be adapted to capture more complex strategies, including finite-state automata with more than two states[TSM13, GvV18] or the leading eight[OI04].

5.2.2 Equilibrium conditions for reciprocal cooperation

Because the strategies of the baseline model only require first-order information, we can compute the players' payoffs explicitly. The respective formula, derived in detail in **Section 5.5.3**, is valid for any population size, arbitrary population compositions, and all parameter values. Based on this explicit representation of payoffs, we first characterised all Nash equilibria among the strategies (y, p, q, λ). In a Nash equilibrium, no player can improve her payoff by unilaterally deviating (not even by using a more complex strategy that uses arbitrary amounts of past information). By extending the theory of zero-determinant strategies[PD12, HNS13, SP14a, SP14c, Aki16], we find that for every $\lambda \in [0,1]$, there can be exactly one generic Nash equilibrium strategy (y, p, q, λ) that yields full cooperation. These strategies are explicitly derived in **Section 5.5.4**. In the following, we summarise the corresponding results.

For direct reciprocity ($\lambda = 0$) the unique strategy that yields stable cooperation is given by the classical *GTFT* strategy (**Fig. 5.3a**), with $y = p = 1$ and

$$q_0 = 1 - \frac{c}{\delta b}. \tag{5.1}$$

Here, δ is the probability that two interacting players interact again some time in the future. This pairwise continuation probability can be derived from the population-wide continuation probability d (**Section 5.5.4**). For indirect reciprocity ($\lambda = 1$), the Nash equilibrium has $y = p = 1$ and

$$q_1 = 1 - \frac{1 + (n-2)\delta}{1 + (n-2)(1-2\varepsilon)} \frac{c}{\delta\, b}. \tag{5.2}$$

In analogy to *GTFT*, we call this strategy Generous Scoring (*GSCO*, **Fig. 5.3b**). Both strategies have in common that they always assign a good reputation to cooperators, and that they occasionally assign a good reputation to defectors. However, they differ in which information they take into account when making these assessments. While *GTFT* only considers direct interactions, *GSCO* takes all interactions of a co-player into account.

The above descriptions of *GTFT* and *GSCO* only give rise to a sensible strategy if their q is non-negative. By requiring $q \geq 0$, equations (5.26) and (5.2) thus characterise when cooperation can be sustained at all. We find that the game's continuation probability needs to be sufficiently large, $\delta \geq \delta_\lambda$. The respective threshold values for direct ($\lambda = 0$) and indirect ($\lambda = 1$) reciprocity are

$$\delta_0 = \frac{c}{b} \quad \text{and} \quad \delta_1 = \frac{c}{b + (n-2)\Big((1-2\varepsilon)b - c\Big)}. \tag{5.3}$$

The threshold δ_0 for direct reciprocity is simply given by the cost-to-benefit ratio of cooperation[Now06]. The threshold δ_1 for indirect reciprocity can be greater or lower,

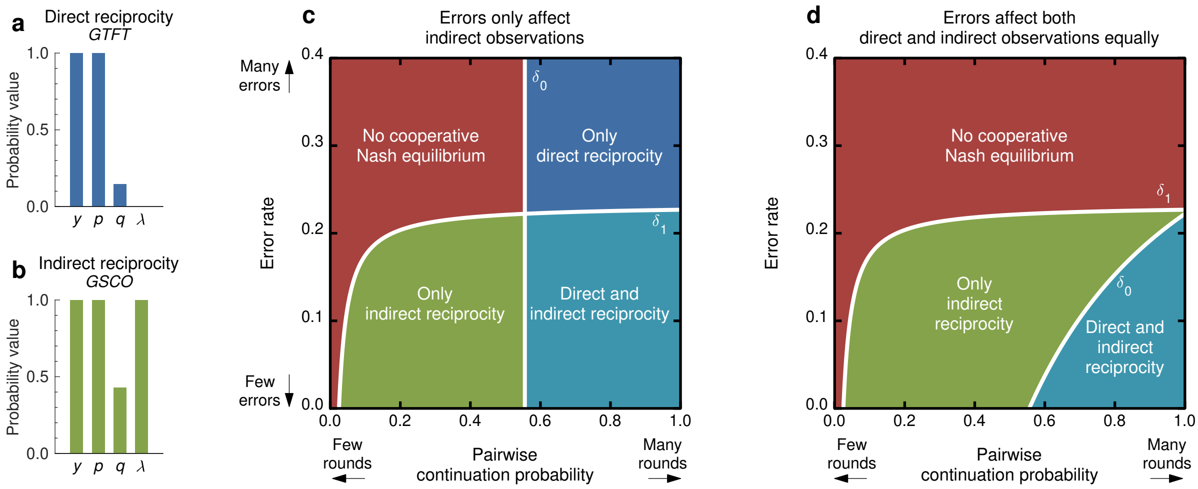

Figure 5.3: **An equilibrium analysis reveals when direct or indirect reciprocity can sustain cooperation. a,b,** Within the reactive strategies, there is one cooperative Nash equilibrium for direct reciprocity ($GTFT$), and one such equilibrium for indirect reciprocity ($GSCO$). Both strategies have in common that they always cooperate in the first round, or if the co-player has cooperated in the last relevant interaction ($y=p=1$). They differ in how they react to a co-player's defection, as described by Eqs. (5.26) and (5.2), and in whether they take into account indirect information. **c,** Depending on the parameters of the game, there are up to four scenarios: (*i*) When there are few rounds and many perception errors, cooperation is infeasible; (*ii*) When there are intermediately many rounds and few perception errors, cooperation can be sustained by indirect but not by direct reciprocity; (*iii*) When there are many rounds and many perception errors, cooperation can be sustained by direct but not by indirect reciprocity; (*iv*) When there are many rounds and few errors, both direct and indirect reciprocity support cooperation. **d,** In case direct and indirect observations are subject to the same error rate, there is no region in which direct reciprocity can sustain cooperation but indirect reciprocity cannot. The figure shows the case of $n=50$, $b=1.8$ and $c=1$. In **c,** the white lines depict the continuation probabilities δ_0 and δ_1 given by Eq. (5.3). In **d,** they are given by $\delta_0=c/\big((1-2\varepsilon)b\big)$ and $\delta_1=c/\big((n-1)(1-2\varepsilon)b-(n-2)c\big)$, where ε is now the joint error probability for both direct and indirect observations.

depending on whether or not outside information is sufficiently reliable (i.e., depending on whether the probability ε of an observation error is greater or lower than $(1-c/b)/2$). The two thresholds in (5.3) give rise to four possible cases (**Fig. 5.3c**). Either (*i*) cooperation is not feasible at all, (*ii*) it is only feasible through indirect reciprocity, (*iii*) it is only feasible through direct reciprocity, or (*iv*) it is feasible through both kinds of reciprocity. We have derived analogous thresholds for δ under the alternative assumption that both direct and indirect observations are subject to the same error rate (**Section 5.5.6**). In that case, the third region vanishes: if cooperation is feasible at all, it is always feasible through indirect reciprocity (**Fig. 5.3d**).

In addition to the extremal cases of direct reciprocity ($\lambda=0$) and indirect reciprocity ($\lambda=1$), we have also explored whether the equilibrium conditions for a cooperative equilibrium can be met more easily if players use intermediate values of λ. Interestingly, the answer is negative. Specifically, we prove that if there is a cooperative Nash equilibrium for some $0<\lambda<1$, then either $GTFT$ or $GSCO$ is already an equilibrium. From an equilibrium perspective, intermediate degrees of receptivity thus do not further extend the possibilities for cooperation. Moreover, in the limit of rare errors we find that the conditions (5.3) are strict even as we allow for arbitrarily complex strategies: if neither *GTFT* nor *GSCO* can sustain cooperation for the given parameters of the game, no other Nash equilibrium can.

5.2.3 Comparing the dynamics of direct and indirect reciprocity

The previous equilibrium results highlight different strategies that can maintain cooperation if adopted by sufficiently many in the population. However, the above results do not imply that such strategies would automatically evolve. After all, $ALLD = (0,0,0,\lambda)$ is also an equilibrium for all parameter values (**Section 5.5.4**). In a next step, we have thus explored under which conditions cooperation can emerge when players engage in social learning.

To this end, we no longer assume that players use equilibrium strategies. Rather they may start out with some arbitrary strategy (y,p,q,λ). Over time players adopt new strategies based on a pairwise comparison process[ST98, TPN07]. This process assumes that in each time step, one individual is randomly drawn from the population. This player then has the opportunity to revise her strategy. She can do so by either adopting a randomly chosen strategy (akin to a mutation in biological models), or by imitating the strategy of another group member (akin to selection). Imitation events are biased such that strategies with a high payoff have a better chance to be imitated (see **Methods**, Section 3.4). This elementary strategy updating step is then iterated over many time periods. We use simulations to record which strategies the players adopt over time and how often they cooperate. To this end, we sometimes assume that mutations are rare. The limit of rare mutations is mathematically well understood [FI06, FNTI06, IN10, WGWT12, McA15] and it has been prominently employed in previous studies of reciprocity [IFN05, GT12b, vSPLS12, SP13, SP15, SSP16, HHCN19] and beyond [HTB$^+$07, SDSTH10, GT12a, HI12, LIDS19]. When mutations are rare, the population consists of at most two strategies, residents and mutants. The mutant strategy goes extinct or fixes before the next mutation arises. The assumption of rare mutations allows simulations to be run more efficiently. This in turn makes it easier to systematically explore the entire strategy space (see **Section 5.5.5** for details). We complement the respective results with simulations with frequent mutations.

We first explore the two limiting cases of reciprocity separately, by fixing either $\lambda=0$ or $\lambda=1$. We consider two different scenarios (**Fig. 5.4**). In the first scenario, individuals interact only for a few rounds. In the other scenario, we consider the limiting case that they interact for infinitely many rounds. This limit has been employed in many previous studies [NS92, HS97, PD12, HNS13, SP14a, SP14c, Aki16] as it naturally reduces the dimension of the strategy space (see **Fig. 5.4a,b**). Similar results can be obtained if the number of rounds is large but finite (**Fig. 5.5**).

In all scenarios we observe that for rare mutations, players either tend to adopt a strategy close to always defect, $ALLD = (0,0,0,\lambda)$, or a conditionally cooperative strategy $(1,1,q,\lambda)$. As expected from our equilibrium analysis, indirect reciprocity is overall more favourable to cooperation when individuals interact only for a few rounds. Interestingly,

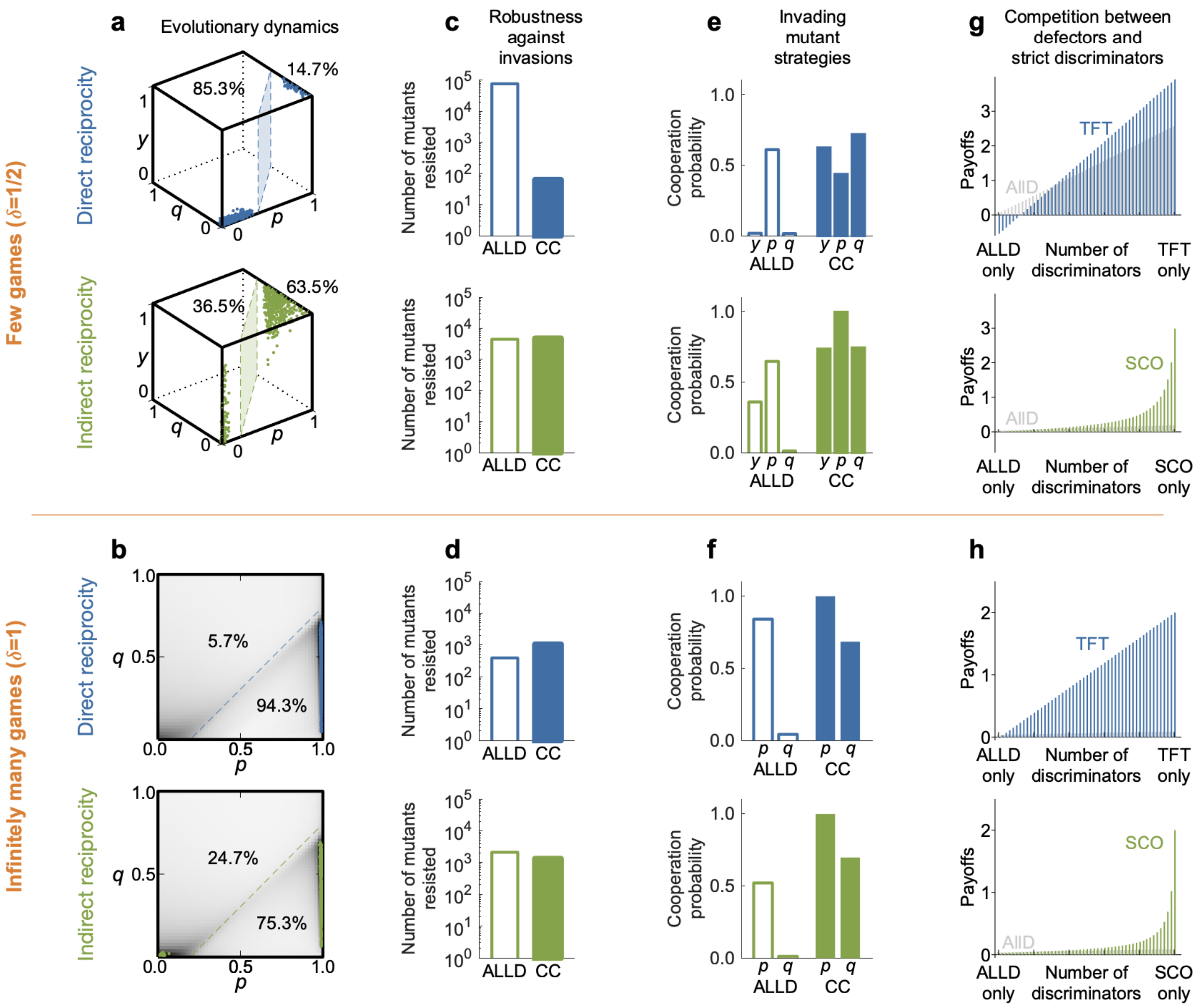

Figure 5.4: **Evolutionary dynamics of direct and indirect reciprocity.** We use individual based simulations to explore the dynamics if either all players engage in direct (blue) or indirect reciprocity (green). We consider two scenarios: individuals either engage in only a few games (top) or in infinitely many games (bottom). **a,b,** We find that over the course of evolution, populations cluster in two regions of the strategy space. Populations are either in the vicinity of *ALLD* (where $y \approx p \approx q \approx 0$), or in the vicinity of conditionally cooperative strategies (where $p \approx 1$). Percentages represent the fraction of time spent in each of these two neighbourhoods. Dots represent the 500 most long-lived resident strategies. As the impact of the first round is negligible for $\delta = 1$, the state space degenerates to a square instead of a cube. **c,d,** We have recorded how many mutants it takes to invade a population of defectors or conditional cooperators. We find that a larger number of rounds undermines the stability of *ALLD*, and enhances the stability of the cooperators. **e,f,** In addition, we have recorded which mutant strategies invade these two resident strategies. On average, defectors are invaded by conditionally cooperative strategies with $p \gg q$. **g,h,** Under direct reciprocity, the payoff of a discriminating mutant (*TFT*) in an *ALLD* population increases linearly in the number of mutants. Under indirect reciprocity, the payoff of a discriminating mutant (*SCO*) is nonlinear. As baseline parameters in our evolutionary simulations, we use $n = 50$ and $b/c = 5$. For the exact setup of these simulations, see **Methods**, Section 3.4.

however, direct reciprocity is more effective in maintaining cooperation when many rounds are played, even in the absence of any observation errors.

To gain some analytical understanding for why direct reciprocity becomes superior, we consider an initial population that either employs *ALLD* or a conditionally cooperative strategy. For both resident strategies, we record how long it takes until a different strategy can invade (**Fig. 5.4c,d**) and which strategies are most likely to do so (**Fig. 5.4e,f**). When many games are being played, conditional cooperators have a similar invasion time for both direct and indirect reciprocity (**Fig. 5.4d**). However, *ALLD* can be invaded more easily when players use direct reciprocity. To explore this differential robustness of defectors, we analysed the competition between *ALLD* and a conditionally cooperative strategy $(1, 1, q, \lambda)$. When only these two strategies are present, the respective payoffs π_D and π_C can be calculated explicitly (see **Section 5.5.5**). In the limit of large populations and rare errors, the payoffs under direct reciprocity ($\lambda = 0$) become

$$\begin{aligned} \pi_C^0 &= (b-c)\cdot z - (1-\delta+\delta q)c\cdot(1-z) \\ \pi_D^0 &= (1-\delta+\delta q)b\cdot z. \end{aligned} \tag{5.4}$$

Here, z is the fraction of conditional cooperators in the population. Eq. (5.4) yields two insights. First, provided that $q < 1-c/(\delta b)$, the dynamics is bistable. If cooperators are common ($z \approx 1$), they have the higher payoff. In contrast, when cooperators are rare ($z \approx 0$), defectors are favoured. Second, the payoff of the two strategies increases linearly in the fraction of cooperators. When we perform the same analysis for indirect reciprocity ($\lambda = 1$), we obtain

$$\begin{aligned} \pi_C^1 &= \frac{q + q(1-q)(1-z)}{1-(1-q)z}\cdot z(b-c) - q(1-z)c \\ \pi_D^1 &= qb\cdot z. \end{aligned} \tag{5.5}$$

Again, for q sufficiently small these payoffs result in a bistable competition. However, while the defectors' payoffs continue to increase linearly in the fraction of cooperators, the cooperators' payoffs are now nonlinear (**Fig. 5.4g,h**).

This analysis highlights two crucial effects that distinguish indirect from direct reciprocity. On the one hand, indirect reciprocity leads to a faster spread of information throughout a population. As a consequence, indirect reciprocity is more effective in restricting the payoff of a defector (i.e., $\pi_D^1 < \pi_D^0$ for all $z > 0$). On the other hand, successful cooperation in indirect reciprocity is based on non-linear synergy effects. Cooperators only obtain high payoffs when they are sufficiently common. Which of the two effects is dominant depends on the population size, the error rate, and on how often players interact on average (**Fig. 5.6**). Once players interact for many rounds, indirect reciprocity ceases to have any advantage (because $\pi_D^1 = \pi_D^0$ for $\delta \to 1$). In that case, defectors are always more readily invaded under direct reciprocity.

Due to their nonlinear returns, cooperative strategies of indirect reciprocity are most effective when they are common. This observation suggests that indirect reciprocity may be less likely to evolve when the evolutionary process itself prevents cooperative strategies to form a large majority. Such a case can occur, for example, when mutations are abundant, such that many different strategies are routinely present in the population. To explore this issue in more detail, we have systematically varied the mutation rate of the evolutionary process (**Fig. 5.5**). Indeed, while mutations only have a minor effect on direct reciprocity,

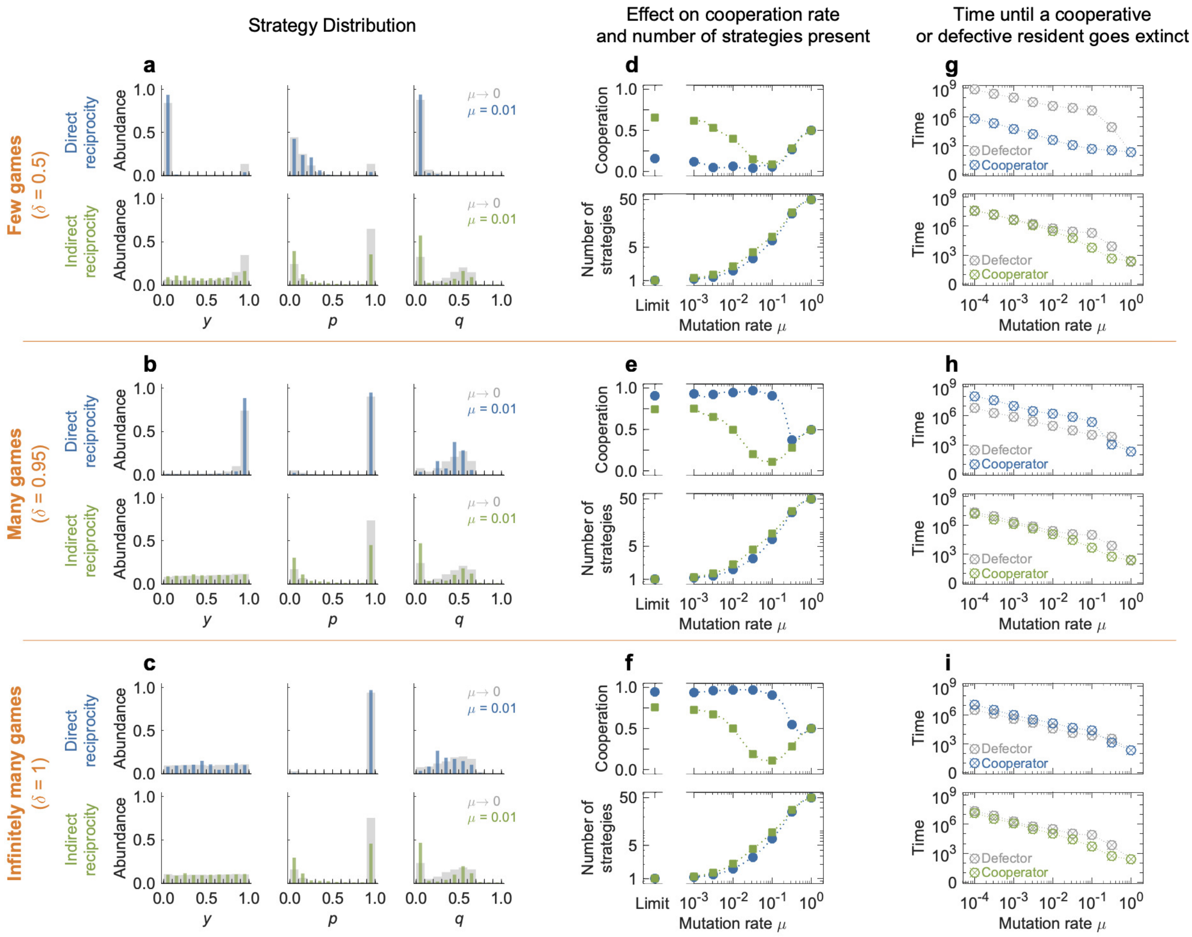

Figure 5.5: **Impact of mutations on either direct or indirect reciprocity.** We have run additional simulations to explore how larger mutation rates affect the results in **Fig. 5.4**. We consider the same two scenarios with few games and infinitely many games, and in addition a scenario where the number of games is large but finite. **a–c,** We have first run simulations for a particular positive mutation rate (coloured). We compare them with the results for the limit of rare mutations (grey). The bar diagrams depict how often we are to observe players to use strategies (y, p, q, λ) either for direct or indirect reciprocity. Similar to **Fig. 5.4**, we find that players are clustered in two regions of the strategy space. Either they tend to defect $(y \approx q \approx 0)$ or they are conditionally cooperative $(p \approx 1,\ q < 1)$. We note that the scenario for infinitely many games yields similar results as the scenario with many games. However, now the initial propensity to cooperate y becomes irrelevant. **d–f,** When we vary the mutation rate systematically, we find that while cooperation is relatively stable for direct reciprocity, cooperation under indirect reciprocity is reduced (upper panels). Interestingly, the number of different strategies that are simultaneously present in the population only differs marginally between direct and indirect reciprocity (lower panels). **g–i,** To explore what would cause this reduction in cooperation for indirect reciprocity, we have checked the stability of defectors and conditional cooperators for various mutation rates, as in **Fig. 5.4c,d**. In the interval $0.01 < \mu < 0.1$ where indirect reciprocity yields the lowest cooperation rate, we find that the stability of defectors is enhanced, whereas the stability of cooperators is reduced. For the exact setup of these simulations, see **Methods**, Section 3.4.

the impact on indirect reciprocity can be substantial as mutation rates become large. In

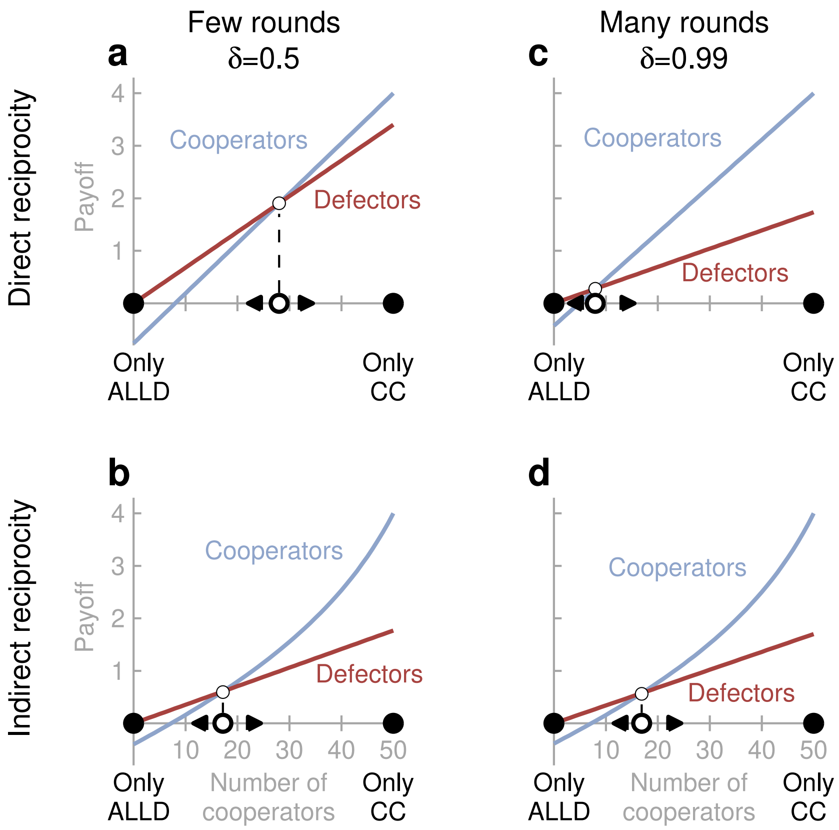

Figure 5.6: **Competition between conditional cooperators and defectors.** We compare the performance of conditional cooperators with strategy $(1,1,1/3,\lambda)$ in a population of defectors, $(0,0,0,\lambda)$. We consider four scenarios, depending on whether players use direct (**a,c**) or indirect (**b,d**) reciprocity and depending on whether pairs interact only a few times (**a,b**) or often (**c,d**). Each panel shows the payoff of cooperators and defectors depending on how many of the 50 population members are cooperators, for $b=5$ and $c=1$. In all four cases we find bistability (as indicated by the arrows on the x-axis). That is, defectors have the higher payoff when there are few cooperators and the lower payoff when there are many cooperators. However, the threshold number of cooperators necessary to make cooperation beneficial differs. Indirect reciprocity has the lower threshold when there are only few rounds, because cooperators are better able to restrict the payoff of defectors (as indicated by the smaller slope of the red line in **b** compared to **a**). Direct reciprocity has the lower threshold when there are many rounds. Here, already a few cooperators suffice to invade the defectors. In contrast, for indirect reciprocity cooperators need to establish a critical mass because their payoffs increase nonlinearly.

particular, around $\mu n=1$ (i.e., once there is more than one mutation per generation), cooperation rates decline quickly. Further simulations suggest that this downfall of cooperation is due to both a reduced stability of conditionally cooperative strategies and an enhanced stability of populations with a majority of defectors (**Fig. 5.5g-i**). These results highlight an important difference between direct and indirect reciprocity. While reactive strategies of direct reciprocity are largely robust to mutations, the corresponding strategies of indirect reciprocity are more sensitive. Strategies like Generous Scoring are most powerful in environments with little noise. To spread, they do not only need outside information to be faithful (small ε) but also the evolutionary process (small μ).

5.2.4 The co-evolution of direct and indirect reciprocity

The above findings raise the question whether the players themselves are able to learn when to use indirect information. To explore this issue, we have first considered a simplified setup in which players can freely choose between all strategies (y, p, q, λ) where either $\lambda = 0$ and $\lambda = 1$. That is, players can choose whether they only take direct interactions into account, or whether they take all of a co-player's interactions equally into account. We study three different scenarios in the limit of rare mutations: one with noisy information and few interactions (**Fig. 5.7a**), one with reliable information and intermediately many interactions (**Fig. 5.7b**), and one with noisy information and many interactions (**Fig. 5.7c**). The results confirm our previous analytical findings. While defectors are predominant in the first scenario, individuals adopt conditionally cooperative strategies in the second and third scenario, showing a bias towards indirect and direct reciprocity, respectively. In a next step, we have systematically varied how often individuals interact with each other, and how noisy third-party information is (**Fig. 5.7d**). Again, we find that indirect reciprocity is most abundant when there are intermediately many rounds, such that cooperation cannot evolve through direct reciprocity alone.

We have repeated all simulations for an evolutionary process with more frequent mutations (**Fig. 5.7e–h**). While the qualitative results are similar, we recover our previous observation that larger mutations rates disfavour indirect reciprocity. Even in those parameter regions in which individuals learn to incorporate third-party information, evolving cooperation rates tend to be lower than in the scenario with rare mutations (**Fig. 5.7f**). The effect of other game parameters on the evolution of cooperation is discussed in **Section 5.5.5**, and visualised in **Fig. 5.8** and **Fig. 5.9**.

Finally, we have also explored which strategies evolve when players can adopt intermediate values of λ (**Fig. 5.10**). To allow for a fair comparison between direct and indirect reciprocity, mutant strategies are drawn such that an average mutant would resort to their direct experience in approximately half of the cases (for details, see **Section 5.5.5**). Overall, we observe a similar trend as before: When information is noisy and there are very few rounds, individuals learn not to cooperate (**Fig. 5.10e**); when there is little noise and intermediately many interactions, individuals learn to cooperate predominantly based on indirect information (**Fig. 5.10f**); and when there is an intermediate amount of noise and many interactions, individuals tend to cooperate based on direct information (**Fig. 5.10g**).

5.3 Discussion

When deciding whether to cooperate, humans often resort to the co-player's reputation arising from third party interactions [WM00, OYS$^+$18], sometimes even if the two players have a joint history of direct interactions [MvdBE13]. Most theoretical studies, however, do not investigate how subjects choose between these two sources of information. They either study direct reciprocity using repeated games [AH81, NS92, HS97, PD12, HNS13, SP14a, SP14c, Aki16, PHRZ15, HRZ15, MH16b, IM18, HCN18, GvV18] or indirect reciprocity using donor-recipient games [NS98b, LH01, OI04, SSP18]. Here, we have proposed a general framework that unifies direct and indirect reciprocity.

To make such a comparison between different kinds of reciprocity most transparent, throughout the main text we have focussed on a comparably simple setup. For example,

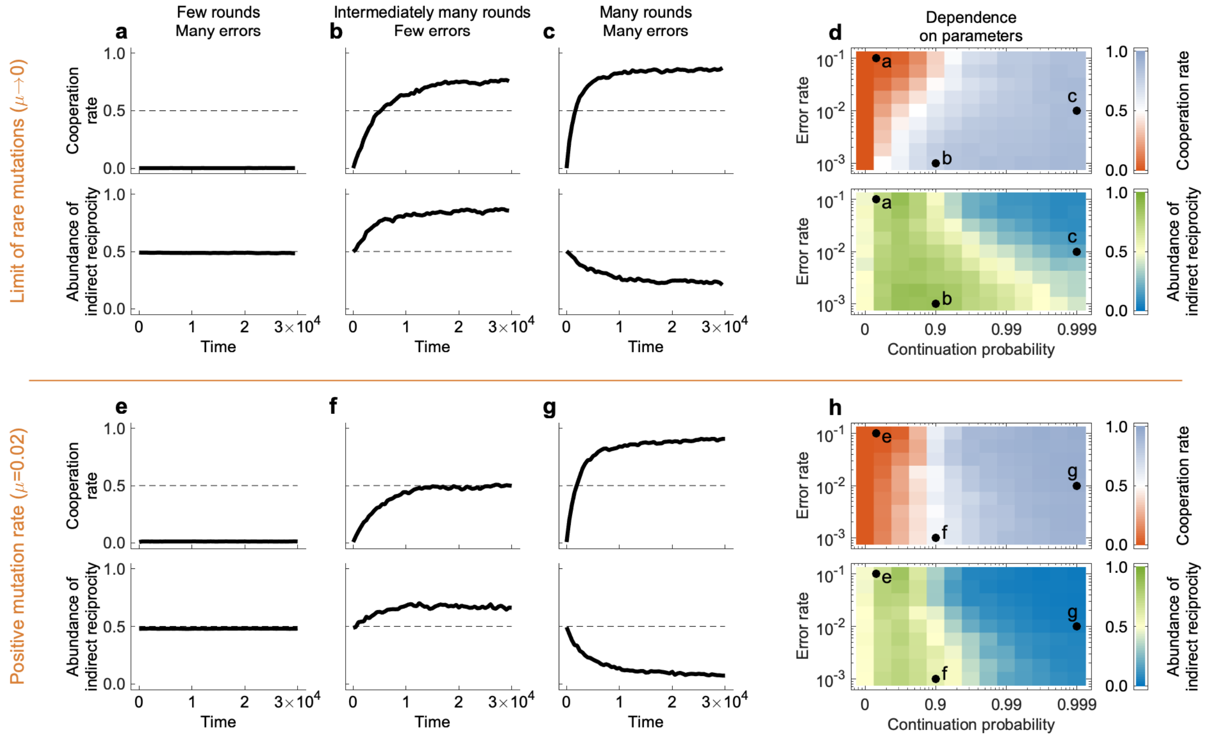

Figure 5.7: **The coevolution of conditional cooperation and information use.** To explore when individuals themselves learn to use indirect information, we ran simulations in which players can either use direct information only ($\lambda = 0$) or all information ($\lambda = 1$). **a–c,** We started with three particular scenarios in the limit of rare mutations. The scenarios differ in how often subjects interact on average and how noisy indirect information is. When there are only a few interactions and considerable noise, cooperation does not evolve at all (**a**). In the other two scenarios, cooperation either evolves due to indirect (**b**) or to direct (**c**) reciprocity. **d,** In a next step, we systematically varied the continuation probability δ and the error rate ε. Again, indirect reciprocity evolves for intermediate continuation probabilities. **e–g,** We obtain qualitatively similar results for positive mutation rates ($\mu = 0.02$). **h,** However, the green region in which individuals take into account indirect information is substantially diminished. For the exact setup of these simulations, see **Methods**, Section 3.4.

we have not modelled explicitly how information from third party interactions spreads throughout a population. We have assumed that individuals observe each others' interactions directly. Instead, one may equally assume that individuals use rumours and gossip to exchange information about their past experiences with other population members. Such communication can add another layer of complexity to the model because players may have an incentive to strategically misrepresent their reputation. For example, defectors are naturally incentivised to prevent others from faithfully learning about their past behaviours. As demonstrated in previous work [NK04, SN16], such miscommunication does not render cooperation impossible. However, it imposes additional bounds on when indirect reciprocity can evolve. While our model does not consider the effects of false gossip explicitly, we may capture some of its workings by assuming that observation errors may be biased. For example, acts of defection may be more likely to be misperceived than acts of cooperation. In **Section 5.5.6**, we show that all our analytical results naturally carry over to this more realistic setting.

Similarly, while we have explored the effect of observation errors in detail, we have neglected other types of errors. As an example, implementation errors have received considerable attention in the previous literature [BS06]. Such errors lead players to mis-execute their intended actions. These players may fail to cooperate although they planned to do so, perhaps because of a 'trembling hand' [Sel75]. The consequences of implementation errors can be rather different from observation errors, because only the former become publicly known. Nevertheless, we can show in **Section 5.5.6** that such implementation errors can be naturally included into our framework (see also **Fig. 5.11**).

Finally, in the main text we have restricted ourselves to the simplest class of strategies, which only depend on a player's previous action. Within this class, we have identified a remarkable strategy of indirect reciprocity. This strategy, called Generous Scoring ($GSCO$), is the analogue of Generous Tit-for-Tat ($GTFT$) [NS92]. It routinely cooperates with other cooperators, but it is also willing to forgive a defector occasionally. Unlike $GTFT$, however, $GSCO$ does not require repeated interactions between two players; it can sustain cooperation even if individuals are likely to never meet again. When previous research on indirect reciprocity identified stable cooperative strategies, the strategies are only shown to be stable within a given strategy class [OI04, SSP18]. This kind of analysis does not rule out that the respective equilibria turn out to be unstable once more complex mutant strategies are permitted. In contrast, Generous Scoring is a Nash equilibrium with respect to all possible mutant strategies, independent of whether mutants themselves use direct or indirect reciprocity, or how much information they are able to process.

This stability of Generous Scoring may be surprising. After all, first-order strategies have been suspected to be incapable of sustaining cooperation [LH01, OI04, SSP18]. For example, Image Scoring is unstable because players have no incentive to retaliate against defectors in the first place [LH01]. By defecting, they would only harm their own reputation, which puts them at risk to receive less cooperation in the future. Generous Scoring circumvents this risk by punishing defectors stochastically, with a well-defined probability. This probability is chosen such that the expected long-term loss in reputation exactly matches the short run gains from saving the cooperation costs. We note that this does not require the players to know all relevant game parameters in advance, or to explicitly calculate the respective probabilities. Instead, our simulations suggest that individuals may well be able to learn such strategies through elementary exploration and imitation processes.

Our results also suggest that direct and indirect reciprocity require different environments to emerge. Generous Tit-for-Tat requires players to interact sufficiently often, whereas Generous Scoring can also sustain cooperation when players only interact occasionally. However, for Generous Scoring to evolve, mutation rates need to be smaller than under direct reciprocity, and outside information needs to be sufficiently reliable (**Fig. 5.4** – **Fig. 5.7**). While our results in the main text focus on simple first-order strategies, our general framework is equally applicable to more elaborate norms of indirect reciprocity. In particular, in **Section 5.5.6** we explore how our framework can be adapted to study strategies represented by finite-state automata with more than two states[GvV18] or the leading eight [OI04], see also **Fig. 5.12**–**Fig. 5.15**.

We have explored how to make decisions when different sources of information are available. When individuals interact regularly, we find that they rely on direct information. They trust their own experiences more than indirect information which may be subject to noise. In contrast, when relationships are short-lived or superficial, cooperation can only be sustained when people act upon public reputations. Previous work suggests that indirect reciprocity requires social norms that are sufficiently complex [LH01, OI04, SSP18]. These norms make use of an unlimited regress; when assigning a new reputation to a person, observers need to take into account the reputation of the person's co-player, which in turn depends on the reputation of the co-player's previous interaction partner. Our model proposes a different view. To sustain cooperation, simple probabilistic rules based on a minimum of information can suffice.

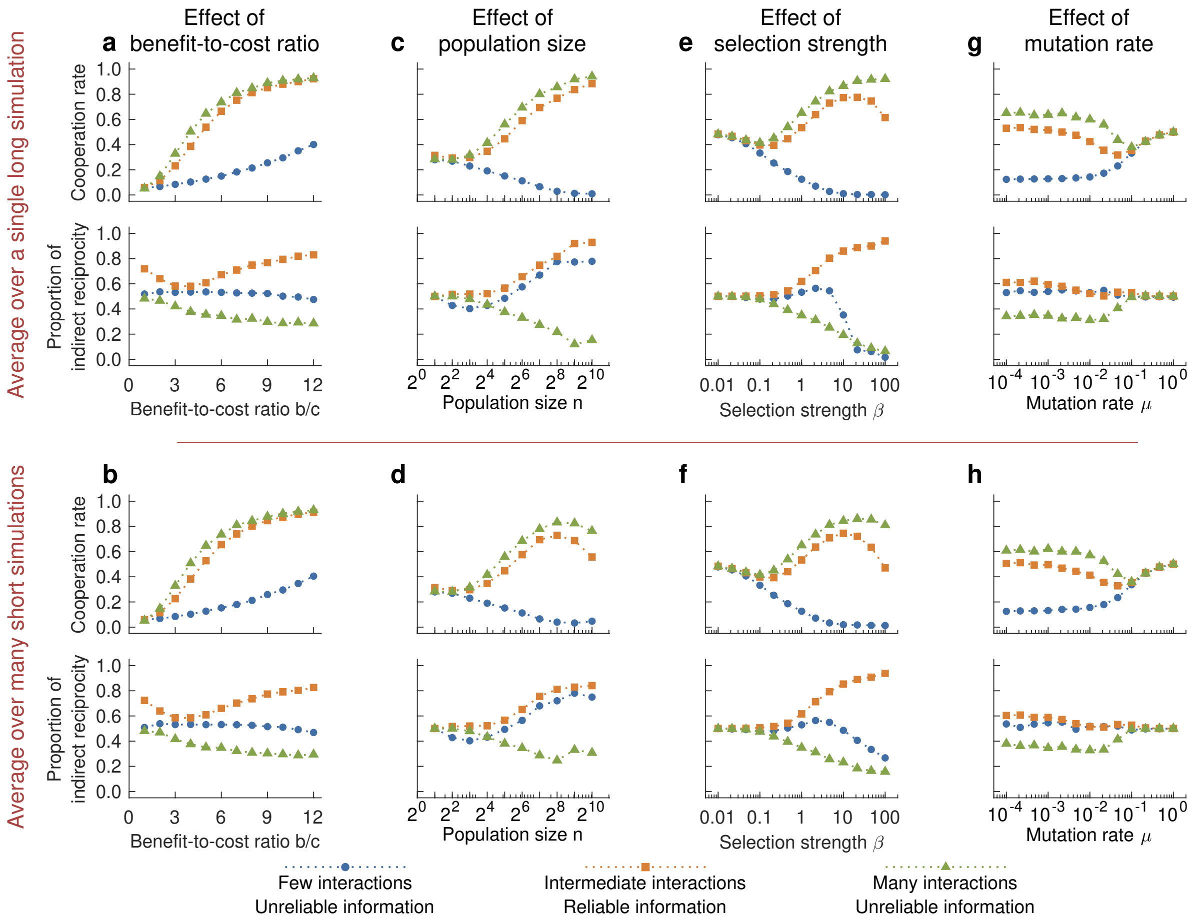

Figure 5.8: **Impact of different model parameters on the co-evolution of direct and indirect reciprocity.** We show how our evolutionary results in **Fig. 5.7** are affected as we change different parameters of our model. In each panel, we vary one parameter and leave all others constant. We consider the same three scenarios as in **Fig. 5.7a–c**: few interactions and unreliable information (blue), intermediate interactions and reliable information (orange), and many interactions and unreliable information (green). We employ two complementary simulation techniques. In the upper half, each data point represents the average of a single simulation. This simulation was run for sufficiently long such that the averages converge and are independent of the initial condition. This typically happens after 10^7 mutant strategies have been introduced into the population. In the lower panels, each data point represents the average of 200 simulations with a random initial population. Here, each simulation only introduces 10^5 mutant strategies. For the parameters, we consider variation in the benefit-to-cost ratio (**a,b**), the population size (**c,d**), the selection strength (**e,f**), and the mutation rate (**g,h**). Our simulations suggest that each of these parameters can have a considerable impact on the evolving cooperation rates and the player's propensity to adopt indirect reciprocity. For example, for the orange curve in panel **e,** we observe that the effect of selection strength on cooperation can be non-monotonic. We further discuss these dependencies in **Fig. 5.9** and **Section 5.5.5**. In general, however, we recover the following regularities from **Fig. 5.7**: (*i*) Substantial cooperation only evolves in the second and third scenario (i.e., for the cooperation rates, the blue curve is systematically below the other curves). (*ii*) If cooperation evolves, players prefer indirect reciprocity when there are intermediately many interactions and outside information is reliable. They prefer direct reciprocity when there are many interactions and when outside information is noisy (i.e., for the proportion of indirect reciprocity, the orange curve is systematically above the green curve).

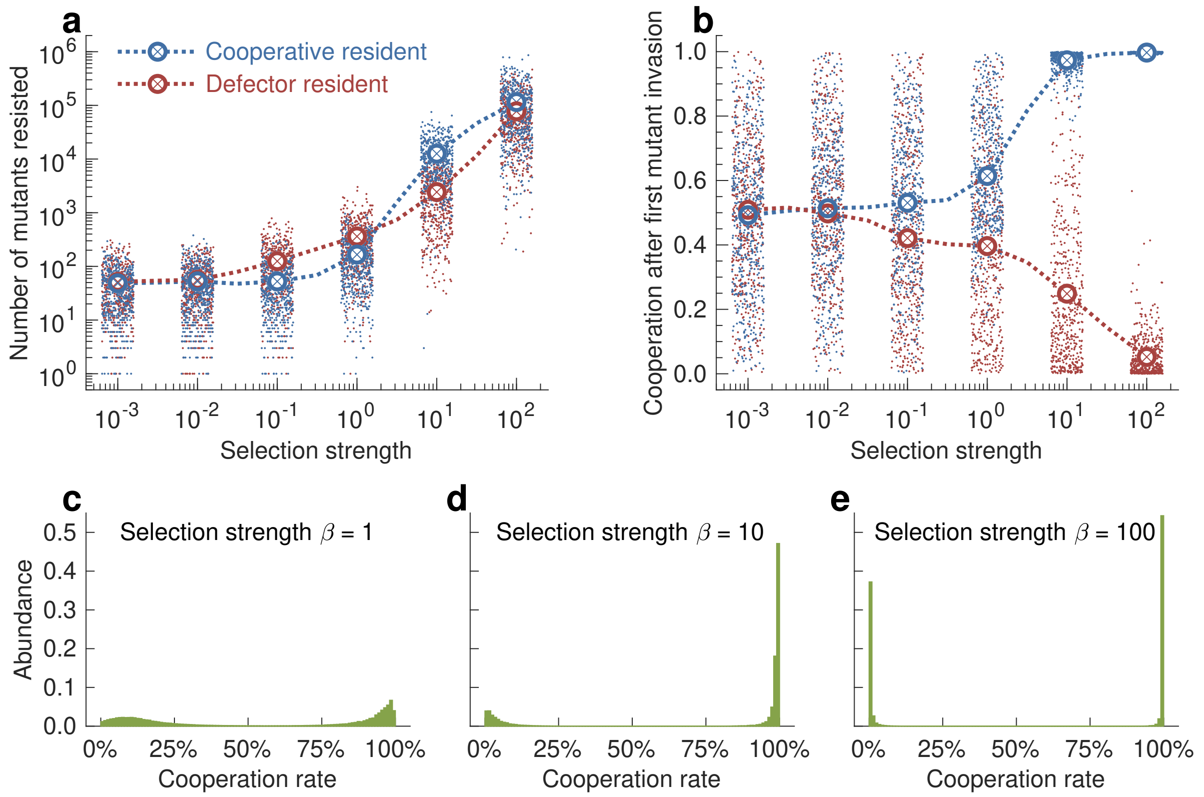

Figure 5.9: **Impact of selection strength on indirect reciprocity.** As shown in the upper panel of **Fig. 5.8e**, selection can sometimes have a non-monotonic effect on cooperation. For intermediate interactions and reliable information ($\delta = 0.9, \varepsilon = 0.001$, depicted by the orange curve in **Fig. 5.8e**), we have observed that the evolving cooperation rate is 53.4% for $\beta = 1$, increases to 77.3% for $\beta = 10$, and reduces to 61.5% for $\beta = 100$. Here we present additional simulations to shed further light on this non-monotonicity. **a,b,** We considered initial resident populations that either adopt a defective strategy or a conditionally cooperative strategy. We recorded how long it takes the evolutionary process until the resident strategy is replaced, and what the cooperation rate of the invading strategy is. Dots show the outcome of individual simulations, and the curves represent averages. The results suggest that the non-monotonicity of cooperation is not due to a reduced stability of cooperative strategies. They remain highly robust even for large selection strengths. Moreover, when selection is strong, they are typically invaded only by other cooperative strategies. **c–e** In a next step, we recorded the distribution of cooperation over time for three different selection strengths for the process considered in **Fig. 5.8e**. We find that this distribution becomes more extreme with increasing selection strength: individuals either become highly cooperative or highly non-cooperative. However, the proportion of non-cooperative populations grows faster than the proportion of cooperative populations.

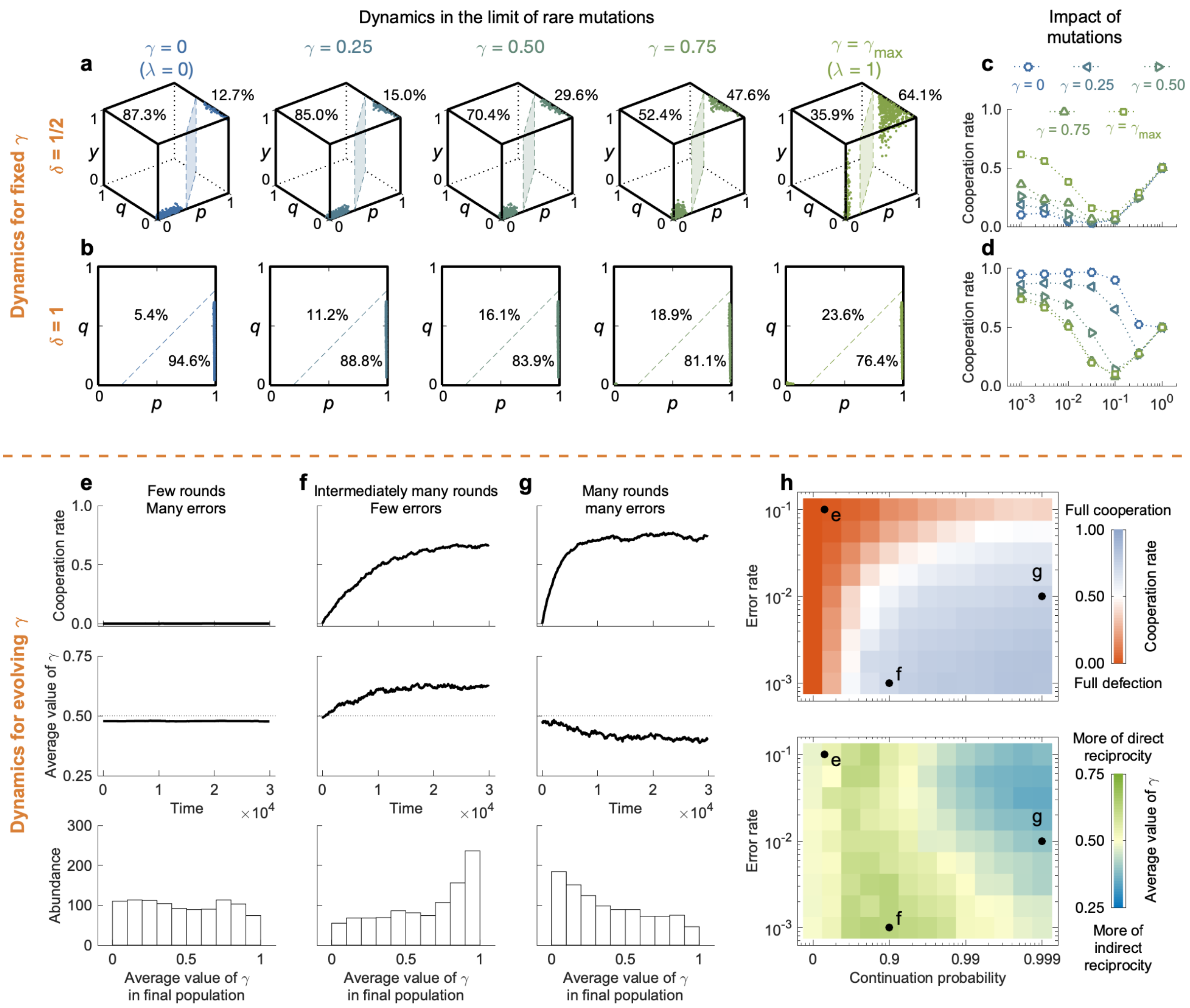

Figure 5.10: **Evolution of cooperation for players with intermediate degrees of receptivity.** In the main text figures **Fig. 5.4**–**Fig. 5.7**, we explore situations in which individuals can choose strategies where they either only take direct information into account ($\lambda = 0$), or where they take all information into account ($\lambda = 1$). Here we repeat these simulations in a setup where intermediate values of λ are permitted. To this end, we define a quantity γ. This quantity is the probability that a player's decision is based on the co-player's behavior towards third parties, see Eq. (5.10) in **Methods**, Section 3.4. For $0 \le \lambda \le 1$ we obtain $0 \le \gamma \le \gamma_{\max} := (n-2)/(n-1)$. **a,b,** We repeat the simulations in **Fig. 5.4a,b** for various values of γ. We observe that cooperation is never most likely to evolve for intermediate values of γ. Either most cooperation evolves for $\gamma = \gamma_{\max}$ (in panel **a**), or for $\gamma = 0$ (in panel **b**). **c,d,** Similarly, we repeat the simulations in **Fig. 5.5d,f** for various values of γ. Again, the average cooperation rates for intermediate γ are strictly in between the results for $\gamma = 0$ and $\gamma = \gamma_{\max}$. **e–h,** Finally, we repeat the simulations shown in **Fig. 5.7a–d**, allowing for mutant strategies (y, p, q, λ) that lead to arbitrary values of γ between 0 and $\gamma_{\max}$. Especially for larger error rates, we observe that the evolving cooperation rates are now smaller. Nevertheless, the general patterns of **Fig. 5.7** remain: (*i*) When there are only few rounds and many observation errors, cooperation does not evolve. (*ii*) When there are intermediately many rounds and few errors, cooperation evolves and players tend to put more weight on indirect information (that is, γ tends to be larger than 1/2). In particular, strategies with $\gamma \approx \gamma_{\max}$ are most abundant. (*iii*) When there are many rounds and intermediately many errors, cooperation evolves and players tend to put more weight on direct information. Here, players are most likely to adopt a strategy with $\gamma \approx 0$. See **Section 5.5.5** for details.

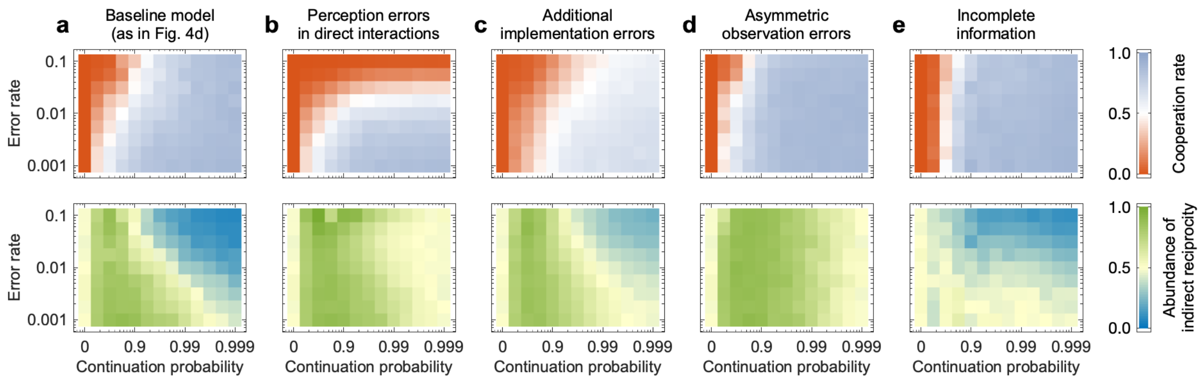

Figure 5.11: **Effect of different types of errors and incomplete information on cooperation.** **a,** To explore how sensitive our results are to different kinds of errors and incomplete information, we have repeated the rare mutation simulations shown in **Fig. 5.7d**, reproduced here. **b,** While the baseline model assumes that only indirect observations are subject to perception errors, here we explore the effects when direct observations are equally prone to errors. We find that cooperation is substantially reduced compared to the baseline scenario. Moreover, direct reciprocity is only favoured for very large continuation probabilities. **c,** We have also explored the effect of additional implementation errors on cooperation. To this end, we assume here that players mis-implement their intended action with fixed probability $e = 0.01$. Compared to the baseline model without such errors, we find that there is less cooperation and less direct reciprocity. **d,** To mimic the dynamics that arises when defectors strategically conceal their bad actions, we have also considered a model in which defective actions are misperceived with probability ε, whereas cooperative actions are always observed faithfully. Because this assumption reduces the total rate at which errors occur compared to the baseline scenario, we observe more cooperation and players are more reliant on indirect reciprocity. **e,** Here we assume that individuals observe third-party interactions only with probability $\nu = 0.01$. Due to the scarcity of information, players who take any third-party information into account are almost indistinguishable from those players who do not. As a result, cooperation is largely independent of observation errors, and the region in which indirect reciprocity is favoured has vanished. Unless noted otherwise, all parameters are the same as in **Fig. 5.7d**.

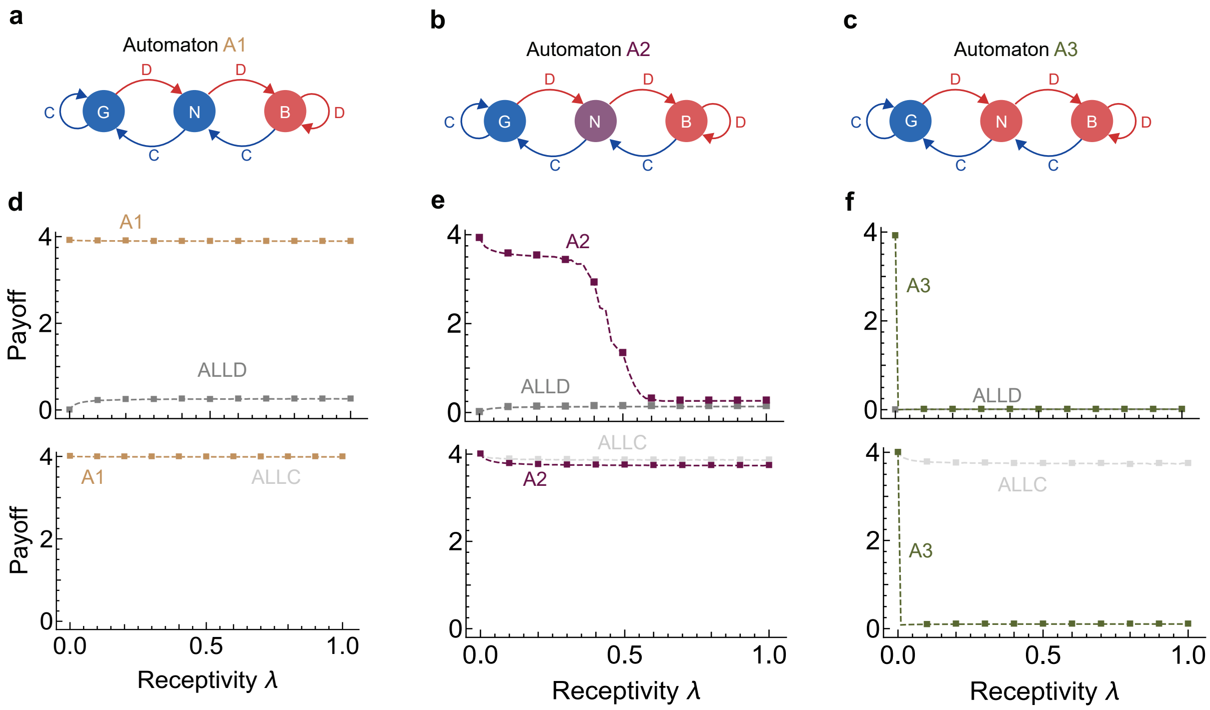

Figure 5.12: **Direct and indirect reciprocity for finite-state automata with three states.** In an extension of our model, we allow players to assign more nuanced reputations to their co-players. We illustrate this approach by considering finite state automata with three states - good (G), neutral (N) and bad (B), with G as the initial state. We assume $n-1$ residents employ the respective finite-state automaton strategy, while the remaining player uses either $ALLC$ or $ALLD$. We simulate the players' payoffs for various values of $\lambda \in [0,1]$. We consider three different automaton strategies employed by the residents. The automata differ in how they deal with co-players that are assigned a neutral reputation. **a,** Players with the first automaton $A1$ are fully cooperative when they encounter a co-player with neutral reputation. This strategy can sustain cooperation among itself. However, a single $ALLC$ player obtains approximately the same payoff as the residents, and hence can invade by (almost) neutral drift (**d**). **b,** According to the second automaton $A2$, players cooperate against neutral opponents with 50% probability. This strategy can be invaded by $ALLC$ for all $\lambda > 0$ (**e**). **c,** According to $A3$, players defect against co-players with a neutral reputation. This strategy is not stable against $ALLC$ for $\lambda > 0$ (**f**), and residents fail to cooperate with each other altogether.

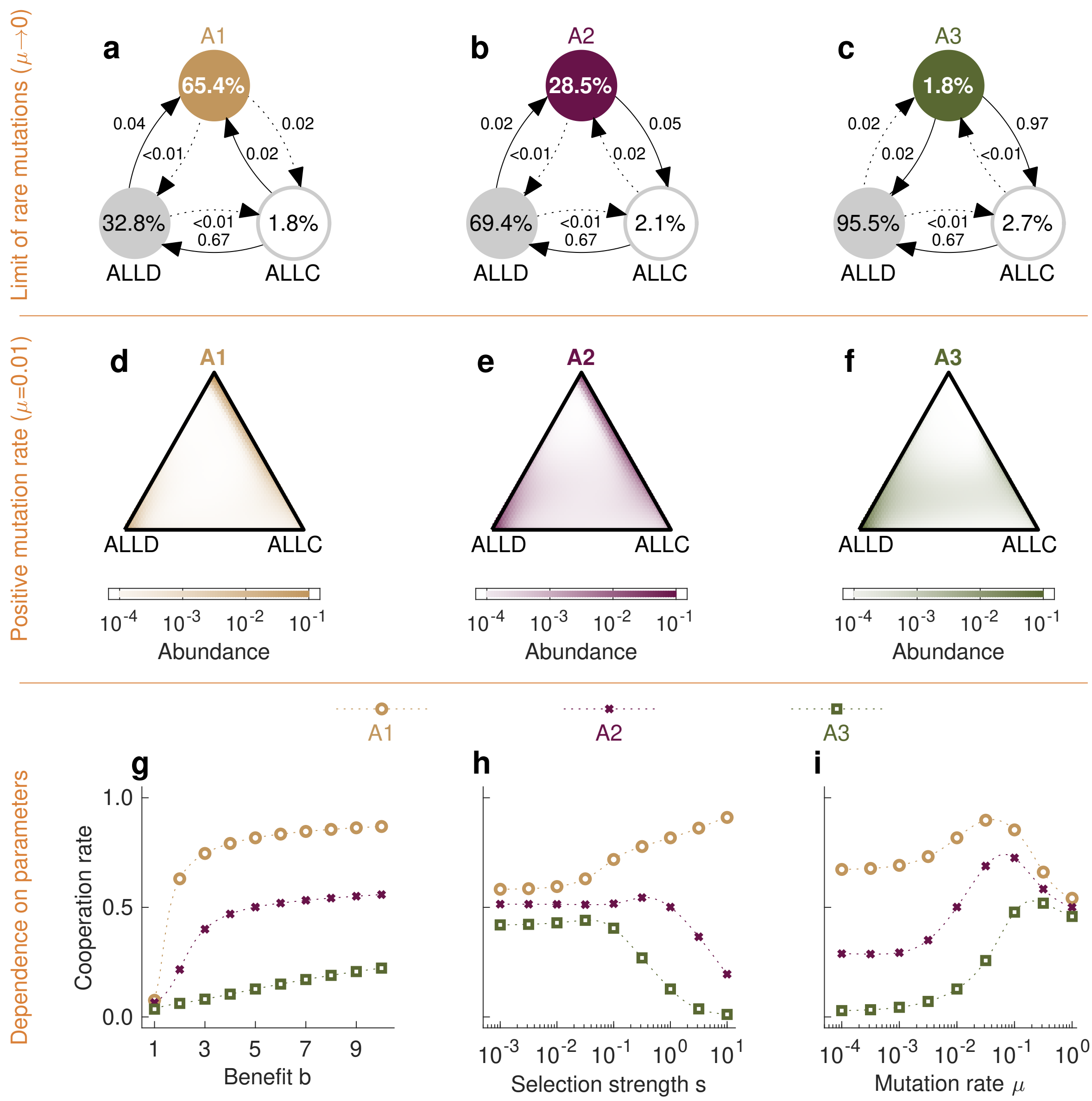

Figure 5.13: **Evolutionary competition between finite state automata, ALLC, and ALLD.** We have explored the evolutionary dynamics when population members can choose between *ALLC*, *ALLD*, and one of the three finite-state automata introduced in **Fig. 5.12**. **a–c,** First, we have explored the limit of rare mutations, using the same game payoffs as in **Fig. 5.12**, and a fixed receptivity $\lambda = 0.1$. The numbers in each circle denote how often the respective strategy is played on average. Arrows illustrate how likely a single mutant fixes in the respective resident population. Solid arrows indicate that the fixation probability is larger than the neutral $1/n$, whereas for dotted arrows this probability is smaller than neutral. We find that only the first automaton A_1 can outperform both *ALLC* and *ALLD*. **d–f,** In a next step, we have explored the same scenario for a positive mutation rate $\mu = 0.01$. The triangles represent the possible population compositions. Each corner corresponds to a homogeneous population, whereas the center corresponds to a perfectly mixed population. The color code reflects how often we observe the respective population composition over the course of evolution. We find that most of the time, populations are either in the neighborhood of *ALLD*, or they represent some mixture between the automaton strategy and *ALLC*. **g–i,** We have re-run the simulations in panels **d–f**, but now varying either the benefit of cooperation, the selection strength, or the mutation rate. In all cases, we observe that the first automaton is most favorable to cooperation. Interestingly, we observe the largest cooperation rate for intermediate mutation rates. This result, however, may be due to the fact that players can only choose from an unbalanced strategy space, as discussed in detail in **Section 5.5.6**.

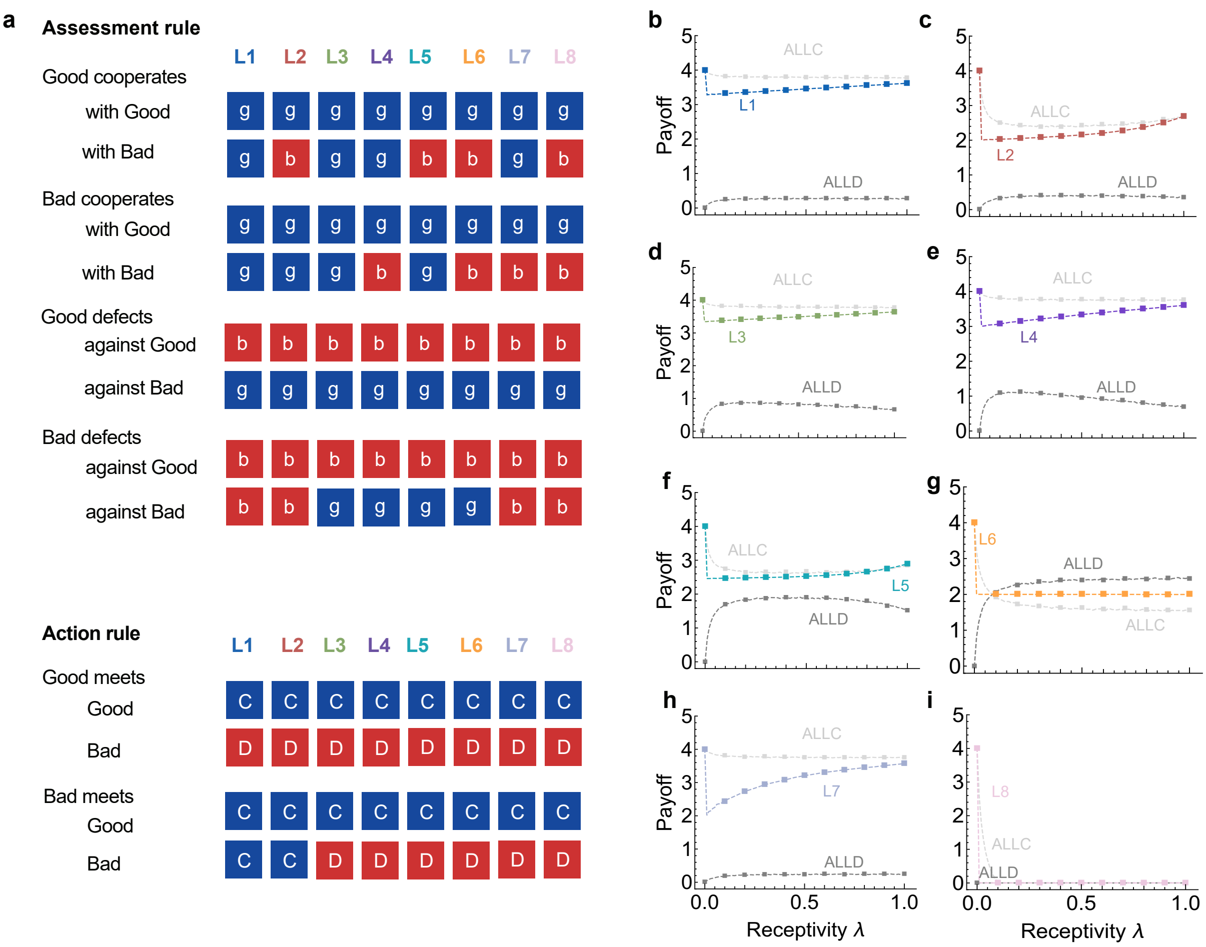

Figure 5.14: **Performance of leading-eight strategies under direct and indirect reciprocity. a,** Previous research has suggested that there are eight stable third-order strategies of indirect reciprocity that can sustain cooperation[OI04], called the leading eight, $L1$–$L8$. They consist of two components, an assessment rule and an action rule. The assessment rule determines how players evaluate each other's actions, depending on the previous reputations of the involved players. The action rule determines how to interact in the game, depending on one's own reputation and on the reputation of the co-player. **b–i,** To explore the stability of these strategies, we consider a population in which $n-1$ players adopt one of the leading-eight strategies. The remaining player either adopts $ALLC$ or $ALLD$. Our results for $\lambda > 0$ reflect previous findings[HST+18]: in the presence of perception errors, all leading-eight strategies are susceptible to invasion by either $ALLC$ or $ALLD$. Only for $\lambda = 0$ (when perception errors are absent), the leading-eight strategies are stable against both mutant strategies.

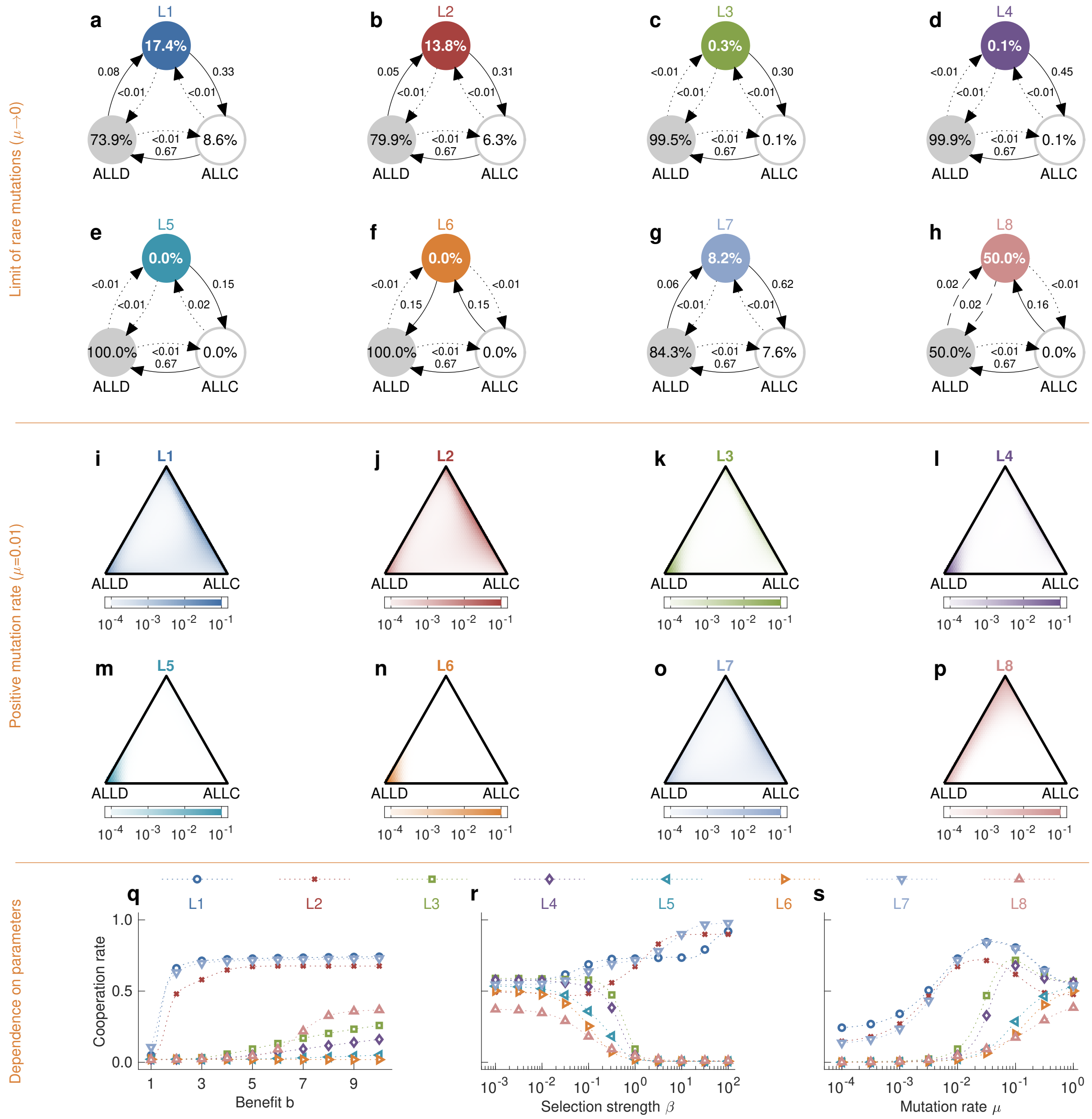

Figure 5.15: **Evolutionary dynamics of the leading-eight.** Similar to **Fig. 5.13** for finite state automata, this figure explores how each of the leading-eight fares in an evolutionary competition against *ALLC* and *ALLD* for a fixed receptivity $\lambda = 0.1$. **a–h,** When mutations are rare, only 'Judging' (L_8) is played in notable proportions. However, in the presence of perception errors, this strategy tends to assign a bad reputation to other players with the same strategy, such that everyone defects eventually[HST+18]. **i–p,** When mutations are more common, some of the leading-eight strategies can stably coexist with *ALLC*. We observe such cooperative coexistences for L_1, L_2, and L_7. **q–s,** These three strategies also yield substantial cooperation rates when we vary the benefit of cooperation, the selection strength, and the mutation rate. With respect to mutation, we again observe that intermediate mutation rates are most favorable to cooperation. However, this finding may not be robust, because the strategy space is again unbalanced. For a more detailed discussion, see **Section 5.5.6**.

5.4 Methods

In the following, we provide a more technical summary of our framework. We explain how it can be used to (*i*) derive the players' payoffs, (*ii*) characterise all Nash equilibria among reactive strategies, and (*iii*) study the co-evolution of direct and indirect reciprocity. For all details and proofs, we refer to **Section 5.5.2** – **Section 5.5.5**.

5.4.1 General framework

For the baseline model considered throughout the main text, we consider a game in a well-mixed population with n individuals. In each round, two individuals are randomly drawn to interact in one round of a prisoner's dilemma. They can either cooperate (C) or defect (D). Cooperation means to pay a cost $c > 0$ to provide a benefit $b > c$ to the co-player. Defection means to pay no cost, and for the co-player to gain no benefit. Both players decide independently. Their actions are observed by all population members. However, we assume indirect information is subject to perception errors: those members who only indirectly witness the interaction may misinterpret each player's action with probability ε. That is, with probability ε, outside observers take a C for a D, or vice versa. After the two interacting individuals have made their decisions, with probability d there is another round. In that case, again two individuals are chosen at random from the population to interact in a prisoner's dilemma. Otherwise, with probability $1-d$, the game is over. The players' payoffs for the population game are defined as their average payoff over all rounds in which they participated in.

Each individual represents every other population member by a separate finite-state automaton. Each automaton can be in one of two possible states, 'good' (G) or 'bad' (B). The current state of the automaton depends on the individual's strategy, on the co-player's past actions, and on whether or not an error has occurred. In the main text, strategies are 4-tuples $(y, p, q, \lambda) \in [0, 1]$. The first entry y is the initial probability for the automaton to be in the good state. The second entry p and the third entry q are the conditional probabilities that the automaton is in the good state, given that the respective co-player just cooperated (defected) in a direct interaction, respectively. Finally, the value of λ is the probability that a player's indirect interactions with third parties are taken into account to update the automaton accordingly. For $\lambda = 0$, third-party interactions are completely ignored, and the automaton's state only depends on direct interactions. For $\lambda = 1$, every interaction of the respective co-player is equally taken into account, no matter whether or not the focal individual is directly involved. Individuals cooperate with those co-players they consider as good, and defect against those co-players they consider as bad.

We refer to the case of $\lambda = 0$ as direct reciprocity, and to $\lambda = 1$ as indirect reciprocity. We note that in exceptional cases, even a player with $\lambda = 1$ may base her decisions on direct experiences. This happens, for example, when the same two players are chosen to interact for two consecutive rounds. In that case, the players' second-round behavior will depend on their direct experience in the first round. In **Section 5.5.6**, we contrast this model with an alternative specification. In that alternative model, players who use indirect reciprocity ignore all direct information entirely. With minor modifications, all results presented herein carry over (see also **Fig. 5.17**).

5.4.2 Derivation of a unified payoff equation

For our baseline framework, the players' payoffs can be calculated explicitly, without having to simulate the game. To derive the respective payoff equation, let each player i adopt some arbitrary but fixed strategy $(y_i, p_i, q_i, \lambda_i)$. Let $\bar{w}=2/n$ denote the probability that a particular player is chosen to interact in the next round of the prisoner's dilemma. Similarly, $w=2/\big(n(n-1)\big)$ is the probability that a particular pair of players is chosen. Finally, we denote by $x_{ij}(t)$ the probability that player i considers player j to be good after t games have been played in the population. Given the value of $x_{ij}(t)$, we can recursively compute $x_{ij}(t+1)$ as

$$\begin{aligned} x_{ij}(t+1) = {} & (1-\bar{w})\, x_{ij}(t) \\ & + w\left(x_{ji}(t)\, p_i + (1-x_{ji}(t))\, q_i\right) \\ & + (\bar{w}-w)\,(1-\lambda_i)\, x_{ij}(t) \\ & + w\lambda_i \sum_{l\neq i,j}\bigg(\Big(1-\varepsilon\Big)x_{jl}(t)+\varepsilon\Big(1-x_{jl}(t)\Big)\bigg)\, p_i + \bigg(\Big(1-\varepsilon\Big)\Big(1-x_{jl}(t)\Big)+\varepsilon x_{jl}(t)\bigg)\, q_i. \end{aligned} \tag{5.6}$$

The summands on the right hand side reflect the following four possible events (illustrated in **Fig. 5.2e–h**):

(*i*) The first line on the right hand side corresponds to the case that player j does not interact at time t. This happens with probability $1-\bar{w}$. In this case i's automaton with respect to j does not change.

(*ii*) The second line corresponds to the case that i directly interacts with j at time t. This happens with probability w. In this case, we assume players always get a chance to update the co-player's reputation. The updated reputation state depends on the values of p_i and q_i, and on j's actions. Player j's action is C with probability $x_{ji}(t)$ and D with probability $1-x_{ji}(t)$.

(*iii*) The third line corresponds to the case that j interacts with some third party, which happens with cumulative probability $(\bar{w}-w)$, but player i decides not to react to this indirect information, with probability $(1-\lambda_i)$. In this case i's automaton with respect to j does not change.

(*iv*) The last line represents the case that j interacts with some third party l, which has probability w each, and i updates her automaton with respect to j accordingly. In this case, player i's updated state depends on whether or not j cooperates, whether or not there is a perception error, and on the values of p_i and q_i. We sum up over all possible interactions of player j with third parties.

Given this recursion, we can calculate the value of $x_{ij}(t)$ for all future times t based on the initial condition $x_{ij}(0) = y_i$. This allows us to compute the weighted average $x_{ij} := (1-d)\sum_{t=0}^{\infty} d^t \cdot x_{ij}(t)$. This average corresponds to the probability to find player i's automaton in the good state in a randomly picked round. Its value can be computed explicitly, by representing Eq. (5.6) in matrix notation (**Section 5.5.3**). Based on the values of x_{ij}, player i's expected payoff becomes

$$\pi_i = \frac{1}{n-1}\sum_{j\neq i}(x_{ji}\, b - x_{ij}\, c). \tag{5.7}$$

This equation allows the explicit calculation of payoffs for arbitrary population compositions. Its results are in agreement with the payoffs that one obtains when simulating the game dynamics explicitly (**Fig. 5.16**).

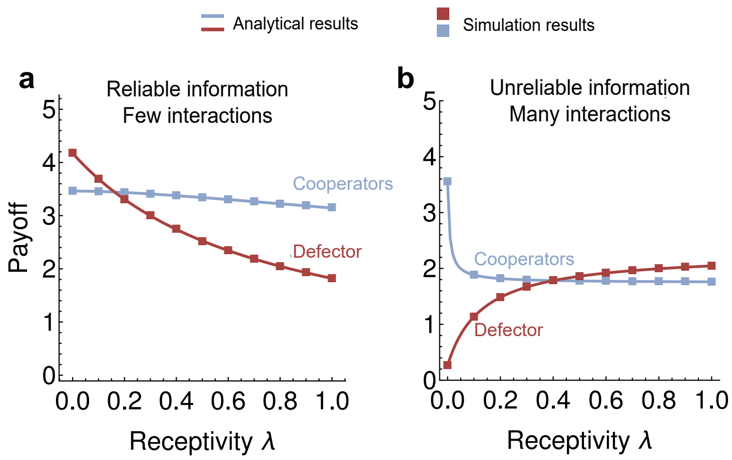

Figure 5.16: **Simulations of the game dynamics confirm the results of the analytical payoff calculations.** Here, we explore whether equation (5.19) gives an accurate prediction of the resulting payoffs when all players adopt some (reactive) strategy (y, p, q, λ). To this end, we consider $n-1$ conditional cooperators with strategy $(1, 1, 0.01, \lambda)$. The remaining player is a defector. We calculate payoffs in two different ways, by using the formula (5.19), and by simulating the game dynamics explicitly. **a,** When errors are rare and the continuation probability is comparably small, cooperators can only outperform defectors when they take indirect information into account. **b,** In contrast, when information is noisy and there are many pairwise interactions, cooperators obtain a better payoff when they ignore indirect information. In both cases, our analytical results agree with the simulations.

5.4.3 Equilibrium analysis

Based on the payoff formula (5.19), we can explicitly characterise the generic Nash equilibria of our model (i.e., those Nash equilibria that are robust with respect to small parameter changes). To this end, it is useful to introduce the variable δ, which is the pairwise continuation probability (the probability that two players interact again, given that they just had an interaction). This probability can be calculated explicitly. It depends on the population-wide continuation probability d and on the population size n, and it is given by $\delta = 2d/\big(2d+(n-1)n(1-d)\big)$. For a derivation, see **Section 5.5.4**.

By extending the theory of zero-determinant strategies[PD12, HNS13, SP14a, SP14c, Aki16, PHRZ15, HRZ15, MH16b, IM18], we prove that a reactive strategy (y, p, q, λ) is a generic Nash equilibrium for $0 < \delta < 1$ if it is either $ALLD = (0, 0, 0, \lambda)$, or if $p - q = r^*_\lambda$ with

$$r^*_\lambda := \frac{1 + (n-2)\delta\lambda}{1 + (n-2)(1-2\varepsilon)\lambda} \cdot \frac{c}{\delta b}. \tag{5.8}$$

In **Fig. 5.4a,b**, the set of all strategies that satisfy $p - q = r^*_\lambda$ is depicted by a coloured face and a coloured dashed line, respectively. If the entire population adopts one of these Nash equilibrium strategies, no single player can gain a higher payoff by deviating.

We call a generic Nash equilibrium *cooperative*, if it has the additional property that all players are fully cooperative in the limit of rare errors. Due to this latter property, the strategy needs to satisfy

$$y = p = 1. \tag{5.9}$$

That is, the strategy always needs to assign a good reputation to unknown players, and to players who have cooperated in the latest relevant interaction. Combining Eqs. (5.8) and (5.9) shows that within the space of reactive strategies, there is exactly one cooperative Nash equilibrium of direct reciprocity ($\lambda = 0$). This strategy is $GTFT$, as defined in Eq. (5.26). Similarly, there is exactly one cooperative Nash equilibrium of indirect reciprocity ($\lambda = 1$), the strategy *GSCO*, defined by Eq. (5.2). In addition to these distinguished boundary cases, we can use Eqs. (5.8) and (5.9) to construct infinitely many cooperative Nash equilibria, one for every value of $\lambda \in [0, 1]$. We refer to the class of all these strategies as *Generous Reciprocators*. For all respective details, see **Section 5.5.4**.

5.4.4 Evolutionary analysis

We model the evolutionary spread of strategies in the population by a pairwise comparison process [ST98, TPN07]. Initially, players adopt an arbitrary strategy (y, p, q, λ). Then one player is randomly chosen from the population to update her strategy. There are two distinct mechanisms how this updating can occur.

(*i*) With probability μ, there is a mutation event. In that case, the focal player abandons her old strategy and instead switches to a new strategy (y', p', q', λ'). The first three entries, y', p', q', are uniformly and independently drawn from the unit interval $[0, 1]$. For simplicity, we assume in most figures that the last entry λ' is either predetermined (for those simulations in which players are restricted to either direct or indirect reciprocity), or that it is randomly taken from the set $\{0, 1\}$.

In addition, in **Fig. 5.10**, we explore how evolution operates when players can also adopt strategies with intermediate λ. For these simulations, we first compute how likely it is for a given strategy (y, p, q, λ) that a player's state with respect to a given co-player is updated between two consecutive games of the two players. The respective probability γ can be calculated as (**Section 5.5.5**)

$$\gamma = \frac{(n-2)\lambda}{1 + (n-2)\lambda}. \tag{5.10}$$

As one may expect, $\lambda = 0$ implies that $\gamma = 0$. That is, a player who ignores all third-party information only updates the co-player's state in a direct encounter and never in between. Similarly, $\lambda = 1$ implies that $\gamma = (n-2)/(n-1)$. That is, a player who takes all information into account has an $(n-2)/(n-1)$ chance to update the co-player's state before the two players interact again (the only exception occurs when the co-player engages in no third-party interaction in between, which happens with a $1/(n-1)$ probability). For the simulations shown in **Fig. 5.10**, we randomly draw mutant strategies (y', p', q', λ') such that the respective γ' according to Eq. (5.10) is evenly distributed in $[0, (n-2)/(n-1)]$. In this way, we ensure that a randomly drawn mutant is approximately equally likely to base her decisions on direct and on third-party information, respectively.

We note that alternatively, one could also consider a mutation scheme where λ itself is uniformly drawn from the unit interval $[0,1]$. We do not employ this alternative mutation scheme here because the resulting mutant strategies would rarely engage in direct reciprocity. Intuitively, players in large populations have many more third-party interactions than they have direct interactions. As a result, even for a comparably small value of λ, the resulting γ according to Eq. (5.10) is typically close to one (especially if the population size n is large). For uniform λ, players would thus rarely act based on their direct experience with the respective co-player. For further details, see **Section 5.5.5**.

(*ii*) With probability $1-\mu$, there is an imitation event. In that case, the focal player randomly chooses another player from the population as a potential role model. If the focal player's payoff according to Eq. (5.19) is given by π_F and the role model's payoff is π_R, the focal player adopts the role model's strategy with probability

$$\rho = \frac{1}{1+\exp\Big[-\beta(\pi_R-\pi_F)\Big]}. \tag{5.11}$$

The parameter $\beta \geq 0$ measures the strength of selection. For small values of β, the imitation probability is roughly 1/2, independent of the strategies of the involved players. As the value of β increases, the more likely it becomes that the focal player only adopts those strategies that yield a higher payoff.

For positive values of μ and finite values of β, the two mechanisms of mutation and imitation give rise to an ergodic stochastic process on the space of all population compositions. To explore the evolutionary dynamics, we have simulated this process for a large number of updating events. We record which strategies the players adopt over time, and how often they cooperate. Because the process is ergodic, the time averages of these quantities converge, and they are independent of the initial population[KT75].

Specific methods employed for the figures. Fig. 5.4a,b depicts simulation results of the evolutionary process when all players are either required to use direct ($\lambda=0$) or indirect reciprocity ($\lambda=1$). We simulated the process for $2\cdot10^7$ mutant strategies. For panels **c–f**, we have looked at simulations where the initial population either employs noisy variant of *ALLD*, $(0.01, 0.01, 0.01, \lambda)$, or a conditionally cooperative strategy, $CC=(0.99, 0.99, 0.5, \lambda)$. We then recorded how long it takes on average until a mutant strategy reaches fixation, and which mutant strategy succeeds. Each bar depicts an average over 10^3 simulations. Panels **g,h** depict the players' payoffs when the population consists of a mixture of defectors and noisy discriminators (*TFT* in the case of direct reciprocity, *SCO* in the case with indirect reciprocity, with $p=0.99$, $q=0.01$). As parameters for this figure, we use $n=50$, $b=5$, $c=1$, $\varepsilon=0$, $\beta=10$, and the limit of rare mutations $\mu\to0$.

Fig. 5.5 explores how different mutation rates affect the results of **Fig. 5.4**. As in **Fig. 5.4**, all players are either restricted to use direct ($\lambda=0$) or indirect reciprocity ($\lambda=1$). For the panels **a–f**, we have then simulated the evolutionary process for different continuation probabilities. Panels **a–c** depict which strategies the players use over time, for either the limit $\mu\to0$ (grey bars) or a mutation rate of $\mu=0.01$ (coloured bars). The upper panels of **d–e** depict how much players cooperate on average, for different values of μ. The lower panels show how many different strategies are simultaneously present in a population on average. This number ranges from 1 in the limit of rare mutations to $n=50$ when mutations are abundant. Panels **g–i** consider populations that are initialised either with the same noisy variant of *ALLD* considered in **Fig. 5.4**, or with the same conditionally cooperative strategy. Again, we record how long it takes on average until the respective strategy has been removed from the population by the evolutionary process. Each data point represents an average of 50 independent simulations. Time is scaled such that units correspond to number of introduced mutant strategies. For $\mu\to 0$ these extinction times converge to the values in **Fig. 5.4c,d**. For $\mu\to 1$, extinction times converge to a value that is the same for both strategies and all continuation probabilities. Unless noted otherwise, parameters are the same as in **Fig. 5.4**.

Fig. 5.7 shows evolutionary results for the case that in addition to the y, p, q values, players are also free to choose $\lambda \in \{0,1\}$. We consider two sets of simulations, either for the case of rare mutations ($\mu\to 0$), or for a positive mutation rate ($\mu=0.02$). The panels **a–c** and **e–g** depict average trajectories for three specific scenarios. The scenarios differ in the game's continuation probability and the error rate. The specific parameters we use are $\delta=0.3$, $\varepsilon=0.1$ (**Fig. 5.7a,e**), $\delta=0.9$, $\varepsilon=0.001$ (**Fig. 5.7b,f**), and $\delta=0.999$, $\varepsilon=0.01$ (**Fig. 5.7c,g**). The time trajectories represent an average over 1,000 simulations. In the initial population, all players employ $ALLD=(0,0,0,\lambda)$ with randomly chosen $\lambda\in\{0,1\}$. Time is scaled such that units correspond to the number of introduced mutant strategies since the beginning of the simulation. In **Fig. 5.7d,h**, we systematically vary the continuation probability and the error rate of the game. Each data point corresponds to the time average of a single simulation with 10^7 time steps. Unless noted otherwise, parameters are the same as in **Fig. 5.4**.

Fig. 5.8 explores how the results for the three scenarios considered in **Fig. 5.7** change as we vary four model parameters. The upper half of the figure considers the same basic setup as in **Fig. 5.7**: the population is initialised such that every player uses *ALLD*. Then we simulate the process for a sufficient time (at least until 10^7 mutant strategies have been introduced). This time is chosen such that the average cooperation rate and the average proportion of indirect reciprocity equilibrates, and that these quantities are independent of the chosen initial population. In the lower half we consider an alternative simulation scheme. Here, we take the average over 200 simulations with randomly chosen initial populations. Each simulation is only run for 10^5 time steps (mutant strategies introduced). The evolutionary parameters that we vary are the benefit-to-cost ratio b/c (between 1 and 12), population size n (between 2 and 1,024), selection strength β (between 0.01 and 100) and mutation rate μ (between 0.0001 and 1). As the baseline parameters, we use the same values as in **Fig. 5.7a–c**.

Fig. 5.9 investigates in more detail the non-monotonicity of evolving cooperation rates in **Fig. 5.8e**. To this end, we again consider the scenario with intermediately many interactions and reliable information ($\delta=0.9$, $\varepsilon=0.001$, orange curve in **Fig. 5.8e**). Panels **a,b,** explore how many mutant strategies it takes on average to invade two different resident strategies. Because the non-monotonicity arises in a scenario that favours the evolution of indirect reciprocity, we consider two residents with $\lambda=1$. The defector resident is given by (0.001, 0.001, 0.001, 1), whereas the cooperative resident adopts the strategy

(0.999, 0.999, 0.650, 1). We have run 1,000 simulations for different values of the selection strength parameter ($10^{-3} \leq \beta \leq 10^2$). Dots represent outcomes of individual simulations, whereas the red and blue curve represent average values. For panels **c–e,** we have run the same evolutionary process as in **Fig. 5.8e** for the scenario with intermediately many interactions and reliable information. For three different intensities of selection, we have recorded the distribution of cooperation rates over a simulation with $2 \cdot 10^7$ time steps. Time steps are measured in number of mutant strategies introduced by the process.

In **Fig. 5.10**, we repeat the simulations in **Fig. 5.4**, **Fig. 5.5**, and **Fig. 5.7**, but now allowing for intermediate values of λ. In **Fig. 5.10a-d**, we consider the case of a fixed λ value. To this end, we use five different values of λ, which according to Eq. (5.10) map to the values of $\gamma \in \{0, 1/4, 1/2, 3/4, \gamma_{\max}\}$, with $\gamma_{\max} = (n-2)/(n-1)$. For the panels **a, b**, we consider the same setup as in **Fig. 5.4a,b**, and use the same parameter values. For the panels **c,d**, we use the same setup and the same parameters as in **Fig. 5.5d,f**. Finally, for the panels **e–h**, we use the same parameters and the same general setup as in **Fig. 5.7a–d**. However, while in **Fig. 5.7a–d**, players are restricted to strategies with either $\lambda = 0$ or $\lambda = 1$, here they can adopt arbitrary strategies with $0 < \lambda < 1$. The bottom panels of **Fig. 5.10e–g** show how often residents adopt different values of γ by the end of each simulation (for 1,000 simulations in total). For details on how the respective mutant strategies are generated, see **Methods**, Section 3.4 and **Section 5.5.5**.

For **Fig. 5.11**, we have re-run the simulations in **Fig. 5.7d** for different noise scenarios. Except for the changes explicitly mentioned (by changing the error scenario, or the information available to the players), the simulations have been performed exactly as for **Fig. 5.7d**.

Fig. 5.12 explores the stability of three different finite-state automata against a single *ALLD* or *ALLC* mutant. Because there is no efficient payoff formula that computes the payoffs of arbitrary automata in the context of indirect reciprocity with noisy observations, we computed the payoffs by simulations. To this end, we assumed all automata are initialized in the good state. Then players engage in $2 \cdot 10^6$ pairwise interactions. To compute the players' average payoffs, we take the mean over all their payoffs in the second half of the simulation, as in previous work[HST+18]. Taking the average over all rounds would not alter our conclusions. As game parameters, we use $n = 50$, $b = 5$, $c = 1$, and $\varepsilon = 0.05$.

For **Fig. 5.13**, we have first simulated the players' payoffs for all possible population compositions (k_A, k_C, k_D). Here, k_A is the number of players who adopt the respective automaton strategy, k_C is the number of unconditional cooperators, and k_D is the number of defectors. For these payoff calculations, we have employed the same process as in **Fig. 5.12**. For pre-computed payoffs, the fixation probability of a given mutant strategy into any other resident strategy can be computed explicitly[NSTF04]. Based on all pairwise fixation probabilities, one can then compute how often each strategy is played on average[FI06]. This yields the panels **a–c.** For positive mutation rates, the abundance of each strategy can still be computed explicitly, by formulating the evolutionary process as a Markov chain. The states of this Markov chain are all possible population compositions (k_A, k_C, k_D). When $n = 50$, there are 1,326 such states. Thus, the dynamics can be described by an $1,326 \times 1,326$ transition matrix. The entries of this transition matrix describe with which probability the population moves from state (k_A, k_C, k_D) to state (k'_A, k'_C, k'_D) after one evolutionary updating event (see for example SI Section 6.2 in Ref. HST+18). The invariant distribution of this Markov chain can be computed directly. It describes how often each state is visited by the evolutionary process, as illustrated in panels **d–f**. Based on this invariant distribution, we can also calculate how often players cooperate on average, as shown in panels **g–i**.

Fig. 5.14 and **Fig. 5.15** use exactly the same method as **Fig. 5.12** and **Fig. 5.13**, respectively. As the only difference, the finite-state automata are replaced by leading-eight strategies.

Fig. 5.16 considers a population consisting of 49 conditional cooperators and a single defector. The cooperators employ the strategy $(1, 1, 0.01, \lambda)$, whereas the remaining defector applies the strategy $(0, 0, 0, \lambda)$. We use two independent approaches to compute the players' payoffs, the payoff equation (5.19), and explicit simulations of the game dynamics (we averaged over 10^5 iterations per parameter combination). The parameters are **a,** $\varepsilon = 0.001$, $\delta = 0.9$ and **b,** $\varepsilon = 0.45$, $\delta = 0.999$. The respective Python scripts used to run the simulations and for solving Eq. (5.19) are provided online, see **Code availability**.

Finally, for **Fig. 5.17**, we repeat the simulations done in **Fig. 5.7a–d**, but now using the alternative strategy set (y, p, q, κ) described in **Section 5.5.6**. Here, players who employ indirect reciprocity ignore all direct information they may have. Game parameters are the same as in **Fig. 5.7a–d**.

5.5 SI

In this work, we introduce a model that combines direct and indirect reciprocity within a unified framework. Players do not only react to the games they are involved in. They also observe how their co-players act in interactions with third parties. We use this model to explore how individuals make use of different sources of information, and to compare how direct and indirect reciprocity facilitate the spread of cooperation throughout a population.

In the following, we describe the framework and the employed methods in detail. In **Section 5.5.1**, we begin by summarizing the previous literature on direct and indirect reciprocity. **Section 5.5.2** introduces our baseline model, for which we assume that players use simple reactive (first-order) strategies. In **Section 5.5.3**, we derive explicit formulas for the players' payoffs for any given population composition. **Section 5.5.4** offers an equilibrium analysis. We show under which conditions reactive strategies suffice to sustain a fully cooperative equilibrium. We prove that the respective conditions are stringent: if reactive strategies cannot sustain cooperation, more complex strategy classes cannot sustain it either. In **Section 5.5.5** we study the co-evolution of direct and indirect reciprocity among reactive players. We show that indirect reciprocity is most likely to evolve when there are only a few interactions, information is reliable, and mutations are not too abundant. In addition, we provide a framework for the evolution of strategies when individuals are able to use a mixture of direct and indirect reciprocity. **Section 5.5.6** discusses several model extensions. We explore how our model can capture different kinds of errors and incomplete information. We also discuss an alternative implementation of indirect reciprocity, according to which players can choose to ignore any direct information they may have. Finally, we incorporate higher-order strategies. The **Appendix** (Section 5.5.7) contains the proofs of our analytical results.

5.5.1 Related literature

Previous literature on direct reciprocity.

There is by now an extensive literature on direct reciprocity, that is, the evolution of cooperation in repeated games. This literature has suggested various strategies that succeed in evolutionary simulations and tournaments [AH81, Mol85, NS92, KK89, IFN07, vSPLS12, FFG+13, PVSP14, RSC15, LHF15, DLTZ15], and it has discussed under which conditions cooperation can be evolutionarily stable [BL87, FW89, Boy89, FM90, BS95, vVGRN12, GvV16]. Moreover, it has explored how the evolution of cooperation depends on model parameters, such as the players' memory [HS97, BJHN16, SP16, HMVCN17] or which strategies players have access to [GT12b, CZN12, SP15]. A general summary of this literature can be found in recent reviews [GvV18, HCN18].

In the context of our paper, the literature most relevant is the recent work on zero-determinant (ZD) strategies for repeated games [PD12, SP12, HNS13, SP13, HNT13, SP14c, SP14b, HWTN14, HRZ15, LLXH15, Aki16, MH16b, MH16a, XRW+17, IM18]. For the repeated prisoner's dilemma, Press and Dyson [PD12] have shown that a player can use such strategies to unilaterally enforce a linear relationship between the players' payoffs. A special case of these zero-determinant strategies are *equalizers* [BNS97], with which a player can enforce that any co-player will get some fixed payoff, independent of the co-player's strategy. In our work, we use such equalizer strategies to construct Nash equilibria for both direct and indirect reciprocity.

Our work is also related to a previous study on crosstalk in repeated games [RHR+18]. Under crosstalk, players pay forward somebody's cooperation: if Alice helps Bob, this may increase the chance that Bob helps some third unrelated player Charlie, even though Bob has no prior positive experiences with that player. This form of generalized reciprocity is maladaptive: crosstalk undermines cooperation, and is disfavored to evolve in the first place. Herein, we adapt some of the mathematical techniques used in that paper [RHR+18] to derive a general payoff formula for direct and indirect reciprocity.

Our study adds to the previous literature on direct reciprocity in the following way:

(*i*) We fundamentally generalize the theory of ZD strategies, by extending it to settings in which players are unlikely to ever meet again, and in which direct reciprocity fails to maintain cooperation.

(*ii*) We explore in which scenarios players engage in direct reciprocity in the first place. To this end, we analyze for which parameters players learn to ignore any indirect information they might have about their present co-player.

Previous literature on indirect reciprocity.

Indirect reciprocity is an alternative mechanism for cooperation [Now06]. It can sustain cooperation even when players only interact once with each other, such that direct reciprocity is infeasible [Ale87, Kan92, OFP95]. Much of the work has focused on the question how complex strategies need to be in order to establish stable cooperation [NS98b, LH01, PB03, BS04, OI04, OI06, PSC06, US09, MVC13, UYOS18, SSP18]. This work has been summarized recently [NS05, Sig12, Oka20].

Since the influential work of Ohtsuki and Iwasa [OI04, OI06], it is a widely shared belief that successful strategies of indirect reciprocity need to be sufficiently complex. The complexity of strategies is typically evaluated in terms of how much information is needed to assess the reputation of a co-player. When players use *first-order strategies*, they assess a player purely based on the players' actions. For example, simple scoring assigns a good reputation to players who cooperated in their last interaction, and it assigns a bad reputation to players who defected [NS98a]. *Second-order strategies* additionally depend on the reputation of the recipient. For example, according to the 'Stern' strategy [PSC06], a player who defects against a good co-player deserves a bad reputation, whereas a player who defects against a bad co-player deserves a good reputation. In addition, *third-order strategies* take the actors' original reputation into account. For example, when players adopt the Staying strategy [SON17], a good player should keep his good reputation no matter how he treats a co-player who is deemed bad. By systematically exploring all deterministic third-order strategies, Ohtsuki and Iwasa have shown that there are only eight strategies that yield stable cooperation, called the 'leading eight' [OI04, OI06]. None of these eight strategies is first-order.

Most relevant to our study, Ohtsuki [Oht04] has explored the adaptive dynamics of stochastic first-order strategies. He shows that the dynamics admits a fixed point in which everyone cooperates. However, since $ALLD$ is the only locally stable fixed point, he concludes that no first-order strategy can sustain cooperation. Some subsequent studies have suggested that first-order strategies may suffice when they record more than a player's last action [Ber11, BG16, CFW19]. For example, in the recent paper by Clark *et al* [CFW19] players count how often each other population member has defected so far. To sustain at least partial cooperation in the population, the paper suggests an innovative strategy called *GrimK*. Players with that strategy cooperate provided that both players' defection record is below K. For games in which there is a coordination-motive to cooperation, they show that one can always find a K such that *GrimK* is a strict Nash equilibrium.

Our study adds to this literature on indirect reciprocity in the following way:

(*i*) We prove that cooperation can be sustained in a Nash equilibrium when players use Generous Scoring. *GSCO* has an intuitive interpretation, it only depends on the co-player's very last action, and it is robust with respect to various kinds of errors (see **Section 5.5.6**).

(*ii*) Our results continue to hold when players differ in the information they have about each co-player, which has been a major obstacle for cooperation in some previous models of indirect reciprocity[HST+18].

(*iii*) We show through simulations that the strategy dynamics in finite populations exhibits cycles. *ALLD* is typically invaded through conditionally cooperative strategies, which in turn are invaded by even more cooperative strategies.

Literature that combines elements of direct and indirect reciprocity.

There is only a handful of studies that explore how direct and indirect reciprocity interact. In an early study on the subject, Raub and Weesie explore the effectiveness of reciprocity when players are placed on a lattice [RW90]. They explore the stability of cooperation for three scenarios, differing in whether players do or do not receive information about their co-players' interactions with third parties. To this end, they analyze whether Grim/Trigger is an equilibrium, assuming there are no implementation or perception errors. In addition they explore the case that third-party information is received with some time lag. The study finds that immediate information about third-party interactions is most favorable to cooperation.

Pollock and Dugatkin propose a strategy for the repeated prisoner's dilemma called 'Observer Tit For Tat' (*OTFT*) [PD92]. Against a co-player with a joint previous history of play, *OTFT* behaves the same way as a *TFT* player. However, against an unknown co-player, *OTFT* takes into account third-party information. The paper explores the static competition between three strategies, *OTFT*, *TFT*, and *ALLD*. It is shown that *OTFT* can be stable against these three strategies even if the continuation probability approaches zero. However, for larger continuation probabilities, *TFT* is shown to be superior.

Roberts presents simulations for a meta-population setup when players can choose among a finite set of strategies [Rob07]. The strategy set represents a selection from the direct and indirect reciprocity literature, and it includes first-order strategies (scoring) as well as second-order strategies (standing). When there are only few interactions between each pair of players, Roberts observes that most players adopt strategies of indirect reciprocity. This trend towards indirect reciprocity is even stronger when players have access to standing strategies. Importantly, however, the study assumes that there is public information about each player's reputation (i.e., it requires that all players agree on a given co-player's reputation at any point in time). For noisy and private information, most previously considered higher-order strategies of indirect reciprocity fail to maintain cooperation [HST+18].

Finally, our work is related to two previous studies in which players are able to misrepresent their own reputation[NK04, SN16].

The first study is by Nakamaru and Kawata [NK04]. They consider a setup where players engage in two kinds of interactions. First, players interact in a series of prisoner's dilemma games. As in our study, players can decide whether to cooperate or defect, depending on the reputation of the opponent. Second, players are matched in pairs to communicate which reputation they assign to every population member. In particular, the model considers the case of private information – different individuals may assign different reputations to the same co-player. Compared to our study, strategies are more complex. Players do not only need to determine what they do in the prisoner's dilemma. They also need to specify to which extent they participate in rumour exchange, and whether or not they initiate wrong rumours about themselves. To this end, the study considers typical archetypes of strategies. For example, 'liars' defect in the prisoner's dilemma, and they misrepresent themselves as cooperative. In contrast, 'Advisors' are conditionally cooperative in the prisoner's dilemma. In the rumour exchanges, they spread true rumour about those co-players who defected against them. Finally, TFT-like players ignore rumours and just implement the traditional tit-for-tat strategy. The strategy dynamics is explored through computer simulations with up to 39 different strategies. Defectors win if players interact on average for one round or less. When there are slightly more interactions, rumour-based strategies like Advisor can succeed. Finally, with many pairwise interactions, TFT-like strategies persist.

The second study by Seki and Nakamaru [SN16] considers a similar setup as the first. However, this time strategies are represented differently. Each player's strategy is encoded as a list of numbers. One number represents how the player acts in the prisoner's dilemma. Two further numbers represent under which circumstances the player would spread positive or negative rumour, respectively. Another number represents under which condition the player would take third-party rumour into account into his own assessment of a player. This decision may depend on the reputation of the person who communicates the respective rumour. With four further numbers, the player represents to which extent direct information (positive or negative) or indirect information (positive or negative) affect the respective co-player's reputation. This reputation is measured by a number between -5 and 5. In addition to the liars whose aim is to self-advertise themselves, this study now also involves players who aim to destabilise the reputation system altogether. For example, such players may spread good rumours about everyone, which adds further noise to the system. The relative weight that players attribute to indirect information does not evolve. However, through extensive computer simulations, the paper shows that the more noise defectors introduce, the more difficult it becomes for cooperation to evolve through indirect reciprocity. Strategies that put more weight on direct reciprocity are more effective under these conditions.

Our study adds to this literature in the following way:

(*i*) We provide a general framework to explore the interaction of direct and indirect reciprocity. This framework allows us to systematically explore the evolution of reciprocity when players can choose among all possible first-order strategies, including stochastic strategies. It can easily be extended to include more complex strategies (see **Section 5.5.6**).

(*ii*) By focusing on comparably simple strategies in the main text, we allow for a more transparent comparison between direct and indirect reciprocity. In particular, we are able to derive analytical conditions when each kind of reciprocity is stable.

(*iii*) Simulation results necessarily depend on which strategies have competed. Adding further strategies can sometimes change the respective conclusions. In contrast, our analytical results are robust. When we find a cooperative strategy to be an equilibrium, it is stable against any possible deviation, even if more complex strategies become available.

5.5.2 Model description

We consider a well-mixed population of n players. Interactions take place in discrete time. In each time step, two players are chosen from the population at random to engage in one round of the prisoner's dilemma. The two chosen players independently decide whether to cooperate (C) or to defect (D). A cooperator pays a cost $c>0$ which yields a benefit $b>c$ to the co-player. Hence, the possible payoffs are $b-c$ if both players cooperate, $-c$ if only the focal player cooperates, b if only the co-player cooperates, and 0 if both defect. After each game, the process iterates with probability d. That is, with probability d, again two players are chosen at random to interact in a prisoner's dilemma. Otherwise, with probability $1-d$, the game terminates. Upon termination, the players' payoffs are calculated by averaging over all interactions in which they participated in.

To model how a player forms and updates her opinions about other group members, we consider players who use separate finite state automata to represent each other group member's reputation. In the baseline model, we assume each automaton only has two states, which we denote by G (the player considers the respective group member to be 'good') and B (the respective group member is considered 'bad'). Players cooperate with those group members they consider as good, and they defect against the bad ones. Such a binary representation of the co-player's current social standing has become standard in the literature on indirect reciprocity [OI04, NS05, Sig12]. However, in **Section 5.5.6** we illustrate how our framework can be extended to allow for more nuanced representations of the co-player. With this extended framework, we can capture previous models in which players have a third 'neutral' [TSM13] or 'unknown' [NM11] state, or models in which the co-player's score is measured in integer values [NS98b, LH01]. The extended framework also allows us to capture situations in which a player's assessment of a co-player depends on more than the co-player's last observed action [HS97, Ber11].

While a player can be in different states with respect to different group members, each player uses a uniform rule to update all his automata. We interpret this rule as the player's strategy. In the baseline model, players use simple reactive strategies. The strategy of player i is represented by a vector $(y_i, p_i, q_i, \lambda_i) \in [0,1]^4$. The first entry y_i is the probability that player i initially assigns a good reputation to a given co-player (without having observed any interaction of that co-player before, see **Fig. 5.1d**). We assume that all of the player's $n-1$ automata are initialized independently. In particular, for $0<y_i<1$, the player may assign different initial states to different co-players. However, all our qualitative results remain unchanged if a player's initial assignments are fully correlated, such that the player assigns the same initial reputation to all other group members. The values of p_i and q_i determine how the player updates a co-player's state after a direct interaction (**Fig. 5.1e,f**). A cooperative co-player is assigned a good reputation with probability p_i, whereas a defecting co-player is assigned a good reputation with probability q_i. As an example, the strategy *ALLD* sets $y = p = q = 0$, whereas *ALLC* uses $y = p = q = 1$.

Finally, the last parameter λ encodes to which extent players take indirect information into account when assigning reputations to others (**Fig. 5.1g**). To model the impact of indirect information, we assume that all players can observe the interactions of all other population members (for the case of incomplete information, see **Section 5.5.6**). After player i witnesses an interaction between player j and some third party, i updates player j's reputation with probability λ_i. If player j's reputation is updated, the new reputation is good with probability p_i if player j has cooperated with the third party, and it is good with probability q_i if player j has defected. We refer to λ as the player's *receptivity*, as it controls to which extent the player is receptive to indirect information.

In the special case that all players set $\lambda = 0$, they completely ignore third party interactions. We refer to this case as 'direct reciprocity'. In the other limit $\lambda = 1$, a player equally takes into account all actions of the other group members, no matter whether the player is directly involved. We refer to this case as 'indirect reciprocity'. We note that even a player with $\lambda = 1$ may occasionally cooperate based on her direct experience with the given co-player. Such an instance occurs for example if the same two players are randomly chosen to interact twice in a row. In that case, a player's behavior in the second round will depend on what happened in the first. This seems natural: in most applications, even a player who routinely takes into account third-party information would not ignore any piece of information merely because it stems from a direct encounter. Nevertheless, it can be useful from a conceptual perspective to consider an alternative model of indirect reciprocity, where all decisions are solely based on third-party information. We consider such a model in **Section 5.5.6**.

The archetypal strategy of direct reciprocity, Tit-for-Tat, corresponds to the vector $TFT = (1, 1, 0, 0)$. The analogous strategy in the indirect reciprocity literature, simple scoring[NS98a], is given by $SCO = (1, 1, 0, 1)$. In **Fig. 5.2a–d**, we provide a graphical illustration of our model and the resulting reputation dynamics. **Table S1** gives a summary of the parameters of our model.

Previous work has shown that the evolution of reciprocity is sensitive to the presence of noise [BS06, Uch10, HST+18]. We thus assume that observations may be subject to perception errors. In this way, we allow indirect information to be less reliable than direct information. Specifically, we assume that when observing an indirect interaction, player i misinterprets somebody's cooperation as defection with probability $\varepsilon < 1/2$. Similarly, a player's defection may be taken as cooperation with the same probability. For simplicity direct interactions are not subject to perception errors in the baseline model. However, in **Section 5.5.6**, we analyze a model extension that includes perception and implementation errors for both modes of reciprocity.

We note that the strategies in this baseline model only make use of first-order information [BS04, OI06]. A player's reputation only depends on the player's action. It does not depend on the standing of the recipient, or on the player's previous reputation. The assumption of first-order strategies greatly facilitates the calculation of the players' payoffs in **Section 5.5.3**. However, in **Section 5.5.6** we revisit higher-order strategies, and we discuss how they can be captured by our framework.

	Parameter	Interpretation
Fixed Parameters	n	Population size
	b, c	Benefit and cost of cooperation
	ε	Error rate for indirect information
	d	Probability of another interaction in the entire population
	δ	Probability of another interaction between a given pair of players, introduced in **Section 5.5.4**
	μ	Mutation rate, introduced in **Section 5.5.5**
	β	Selection strength, introduced in **Section 5.5.5**
Evolving traits	y	Probability to assign a good reputation to unknown players
	p	Probability to assign a good reputation to co-players who cooperate
	q	Probability to assign a good reputation to co-players who defect
	λ	Probability to use indirect information

Table 5.1: **Parameters of the model.** Our model involves a number of fixed parameters that are the same for all players and kept constant over time. In addition, our model considers four evolving traits. The values of the evolving traits may differ between individuals. They are kept constant over the course of a game, but they may change over an evolutionary timescale (see **Section 5.5.5**).

Finally, we note that starting with the influential work of Ohtsuki and Iwasa [OI04, OI06] much of the literature on indirect reciprocity considers the case of public information [PSC06, SSP16, SON17, UYOS18]. In such models, players do not observe each others' interactions with third parties directly. Rather there is a central observer who monitors all interactions in a population, assigns new reputations, and disseminates the updated reputations to all population members. Models of public information have the useful mathematical property that all players agree on the reputation of any other population member (because everyone receives the same information from the same source). In contrast, herein we are interested in the formation of reputations when some players base their decisions on direct interactions, whereas others may also take indirect information into account. In such a situation, players may no longer agree on the reputation they assign to some given co-player. Our model is thus – necessarily – a model of private information[Uch10, HST+18, NK04, SN16].

5.5.3 A unified payoff equation for direct and indirect reciprocity

In this section we derive an explicit expression of the payoffs when all players use reactive strategies (y, p, q, λ). To this end, let $x_{ij}(t)$, be the probability that player i assigns a good reputation to co-player j at time t. As shown in the **Methods**, Section 3.4 section of the main text, this quantity satisfies the recursion,

$$\begin{aligned} x_{ij}(t{+}1) &= (1{-}\bar{w})\, x_{ij}(t) \\ &\quad + w\left(x_{ji}(t)\, p_i + (1{-}x_{ji}(t))\, q_i\right) \\ &\quad + (\bar{w}{-}w)\,(1{-}\lambda_i)\, x_{ij}(t) \\ &\quad + w\lambda_i \sum_{l\neq i,j}\Big(\big(1{-}\varepsilon\big)x_{jl}(t){+}\varepsilon\big(1{-}x_{jl}(t)\big)\Big)\, p_i + \Big(\big(1{-}\varepsilon\big)\big(1{-}x_{jl}(t)\big) + \varepsilon x_{jl}(t)\Big)\, q_i, \end{aligned} \tag{5.12}$$

with

$$x_{ij}(0){=}y_i. \tag{5.13}$$

The parameter $\bar{w}{=}2/n$ is the probability that a particular player is chosen to interact in the next round, and $w{=}2/\big(n(n{-}1)\big)$ is the probability that a particular pair of players is chosen. To obtain an expression for the payoffs, we take Eq. (5.12), collect all terms with $x_{kl}(t)$, and define $r_i{:=}p_i{-}q_i$, which yields

$$\begin{aligned} x_{ij}(t{+}1) &= \Big(1{-}w{-}\lambda_i(\bar{w}{-}w)\Big)x_{ij}(t) \;+\; w\, r_i\, x_{ji}(t) \;+\; w\lambda_i(1{-}2\varepsilon)r_i \sum_{l\neq i,j} x_{jl}(t) \\ &\quad + \Big(w\, q_i + \lambda_i(\bar{w}{-}w)(\varepsilon r_i + q_i)\Big). \end{aligned} \tag{5.14}$$

Eq. (5.14) indicates that the value of $x_{ij}(t+1)$ is a linear function of the respective probabilities $x_{kl}(t)$ in the previous round.

For further manipulation, it is useful to rewrite Eq. (5.14) using matrix notation. To this end, we collect the players' probabilities to assign a good reputation to their co-players in an $n(n{-}1)$-dimensional column vector,

$$\mathbf{x}(t) := \Big(x_{12}(t),\ldots,x_{1n}(t);\; x_{21}(t),\ldots,x_{2n}(t);\; \ldots\; ;\; x_{n1}(t),\ldots x_{n(n-1)}(t)\Big)^{\mathsf{T}}. \tag{5.15}$$

Similarly, we collect the factors in the first line of Eq. (5.14) in an $n(n{-}1){\times}n(n{-}1)$ matrix $\mathbf{M}{=}(m_{ij,kl})$, with entries

$$m_{ij,kl} = \begin{cases} 1{-}w{-}\lambda_i(\bar{w}{-}w) & \text{if } i{=}k \;\text{ and } j{=}l \\ w\, r_i & \text{if } i{=}l \;\text{ and } j{=}k \\ w\lambda_i(1{-}2\varepsilon)r_i & \text{if } i{\neq}l,\; j{\neq}l, \;\text{ and } j{=}k \\ 0 & \text{otherwise.} \end{cases} \tag{5.16}$$

Finally, we collect the constant term in the second line of Eq. (5.14) in an $n(n-1)$-dimensional column vector $\mathbf{v} = (v_{ij})$ with entries

$$v_{ij} = w\, q_i + \lambda_i(\bar{w}{-}w)(\varepsilon r_i + q_i) \quad \text{for all } j. \tag{5.17}$$

Using this notation, we can write Eq. (5.14) as $\mathbf{x}(t{+}1){=}\mathbf{M}\mathbf{x}(t){+}\mathbf{v}$. Based on this equation, we calculate the weighted average

$$\mathbf{x} := (1{-}d)\sum_{t=0}^{\infty} d^t \cdot \mathbf{x}(t) = (\mathbf{1}{-}d{\cdot}\mathbf{M})^{-1}\Big((1{-}d)\mathbf{x_0} + d\mathbf{v}\Big). \tag{5.18}$$

Here, $\mathbf{1}$ denotes the identity matrix, and $\mathbf{x_0}$ is the shorthand notation for $\mathbf{x}(0)$ with entries as defined by Eq. (5.13). The $n(n{-}1)$ entries x_{ij} of this vector $\mathbf{x}$ can be interpreted as the probability to find player i's automaton with respect to j in the good state in a randomly picked round. We use Eq. (5.18) to compute the expected payoff π_i of player i as

$$\pi_i = \frac{1}{n-1} \sum_{j \neq i} (x_{ji}\, b - x_{ij}\, c). \tag{5.19}$$

In **Fig. 5.16**, we compare this analytically derived payoff with the payoff obtained from simulations of the game dynamics. For these simulations, we consider a population where $n-1$ players use the conditionally cooperative strategy $\sigma_C = (1, 1, q, \lambda)$, whereas the remaining player is a defector, $\sigma_D = (0, 0, 0, \lambda)$. We compute and simulate the payoffs for different values of λ for two different scenarios (with different continuation probabilities and error rates). In all cases, we observe that the analytically derived payoffs fully match the simulation results.

When using the above equations to calculate payoffs, the computationally most expensive step is to find the inverse of the $n(n-1) \times n(n-1)$ matrix $(\mathbf{1} - d\mathbf{M})$ in Eq. (5.18). This computation can be made more efficient when several players in a population adopt the same strategy (which will often be the case in evolutionary simulations). In that case, one can exploit the symmetries of a well-mixed population: all players who adopt the same strategy are expected to receive the same payoff. For the computation of payoffs it is thus not necessary to distinguish between all n players. It suffices to distinguish between all strategies that are present in the population – a usually far smaller number. In the **Section 5.5.7**, we show how payoffs can be computed more efficiently when these symmetries are taken into account.

5.5.4 Equilibrium analysis

Characterization of all reactive Nash equilibria

In the following, we wish to characterize which reactive strategies $\sigma_i = (y_i, p_i, q_i, \lambda_i)$ are Nash equilibria (in the next subsection we will then focus on those Nash equilibria that yield full cooperation). A strategy is a Nash equilibrium if no player has an incentive to unilaterally deviate from it. To simplify the notation, and to make our results comparable to the previous literature on direct and indirect reciprocity, it is useful to introduce an additional parameter δ, which is the continuation probability for a pair of players.

Lemma 1. *Consider a population of size n, and let d be the probability that another random pair is drawn from the population after the current round. Let δ be the probability that a given pair of players interacts again after it has just participated in an interaction. Then*

$$\delta = \frac{2d}{2d + (n-1)n(1-d)}. \tag{5.20}$$

All proofs are provided in the **Section 5.5.7**. While d is the global continuation probability (the probability that another game will be played in the entire population), the parameter δ is the continuation probability for each pair of players. We note that the above formula implies that $\delta = 1$ if and only if $d = 1$, and that $\delta = 0$ if and only if $d = 0$, as one may expect. Using the pairwise continuation probability δ, we can formulate the following result that will help us to characterize the set of all Nash equilibria.

Lemma 2. *Consider a homogeneous population with strategy $\sigma=(y,p,q,\lambda)$, and let $r{:=}p{-}q$. Then on average, players assign a good reputation to each other with probability*

$$x = \frac{\Big(1{-}\delta\Big)y + \delta\Big(q + \lambda(n{-}2)(q{+}r\varepsilon)\Big)}{\Big(1{-}\delta\Big) + \delta\Big(1{-}r + \lambda(n{-}2)(1{-}r{+}2r\varepsilon)\Big)} \tag{5.21}$$

In particular, for a generic game with $n{>}2$, $\varepsilon{>}0$, and $0{<}\delta{<}1$ we have:

1. *The population is fully cooperative (all players' automata are in the G state for the entire game) if and only if $y{=}p{=}1$ and either $\lambda{=}0$ or $q{=}1$.*
2. *The population is fully defecting (all players' automata are in the B state for the entire game) if and only if $y{=}q{=}0$ and either $\lambda{=}0$ or $p{=}0$.*

Finally, the following lemma describes which payoff a single player can get from deviating from the *resident* strategy (the strategy everyone else in the population applies). For brevity, we will sometimes refer to the deviating player as the *mutant*.

Lemma 3. *Consider a population where all but one player apply the resident strategy $\sigma{=}(y,p,q,\lambda)$. Let $r:=p{-}q$. Then the mutant's payoff π' takes the form*

$$\pi' = A_1 + A_2(r - r^*_\lambda)x'. \tag{5.22}$$

Here, x' is the mutant's average cooperation rate against the residents, $A_1, A_2 > 0$ are constants that only depend on the resident strategy, and

$$r^*_\lambda = \frac{1 + (n{-}2)\delta\lambda}{1 + (n{-}2)(1{-}2\varepsilon)\lambda} \cdot \frac{c}{\delta\, b}. \tag{5.23}$$

According to Lemma 3, the payoff of the mutant is a linear function of the mutant's cooperation rate. Due to the properties of linear functions, it follows that the mutant's payoff is either maximized by choosing $x'{=}1$ (if $r{\geq}r^*_\lambda$), or by choosing $x'{=}0$ (if $r{\leq}r^*_\lambda$). That is, for a resident strategy σ with $r{>}r^*_\lambda$, *ALLC* is a best response. In analogy to the case of reactive strategies for the repeated prisoner's dilemma[HS98], we call the set of all such strategies $\sigma{\in}[0,1]^4$ the *cooperation rewarding zone*. Conversely, for any strategy σ with $r{<}r^*_\lambda$, *ALLD* is a best response, yielding the *defection rewarding zone*. In between these two zones, for $r = r^*_\lambda$, any mutant strategy obtains the same payoff $\pi'{=}A_1$. Strategies for which $r{=}r^*_\lambda$ are called *equalizers*. Equalizer strategies have been previously described in models of direct reciprocity[BNS97, PD12, HTS15]. Lemma 3 guarantees that analogous strategies also exist when players take arbitrary amounts of indirect information into account.

It is important to note that Lemma 3 makes no restrictions on the mutant strategy. For the lemma to hold, we do not require the mutant to choose a strategy of the form $\sigma' = (y',p',q',\lambda')$. Instead the mutant may take arbitrarily many past actions of the co-player into account, and she may combine direct and indirect information in non-trivial ways. According to Lemma 3, the mutant's eventual payoff is solely determined by her resultant average cooperation rate.

Based on the above lemmas, we can now characterize all Nash equilibria among the reactive strategies.

Theorem 1 (Characterization of Nash equilibria).
Consider $0<\delta<1$, *and a strategy* $\sigma=(y,p,q,\lambda)$.

1. In a game with $n>2$ *and* $\varepsilon>0$, *a strategy* σ *with* $\lambda>0$ *is a Nash equilibrium if and only if it is either ALLD or an equalizer strategy,*

$$y=p=q=0, \quad \text{or} \quad p-q=r^*_\lambda. \tag{5.24}$$

We refer to strategies of the form (5.24) as generic Nash equilibria.

2. If $\lambda=0$, $n=2$ *or* $\varepsilon=0$, *the strategy* σ *is a Nash equilibrium if and only if it is either generic, or if one of the following two cases applies,*

$$y=q=0, \;\; p<r^*_\lambda \quad \text{or} \quad y=p=1, \;\; q<1-r^*_\lambda. \tag{5.25}$$

Remark 1. *We emphasize that while Theorem 1 characterizes the Nash equilibria among the reactive strategies, it does not restrict the strategies deviating players may employ. The Nash equilibria described in Theorem 1 are robust against any possible mutant strategy, including mutant strategies that take higher order information into account, or mutant strategies that depend on the whole history of previous play.*

Cooperative Nash equilibria

In the following, we are interested in those strategies that can sustain high levels of cooperation in a population. To this end, we call a strategy σ a *cooperative Nash equilibrium* if

(*i*) it constitutes a generic Nash equilibrium, and
(*ii*) the cooperation rate in a homogeneous σ-population approaches one as the error rate ε goes to zero.

From Eqs. (5.21) and (5.24), it follows that cooperative Nash equilibria need to be equalizers of the form $\sigma=(1,1,q^*_\lambda,\lambda)$ with $q^*_\lambda:=1-r^*_\lambda$ and r^*_λ defined by Eq. (5.23). In the case of direct reciprocity ($\lambda=0$), the corresponding cooperative Nash equilibrium thus takes the form

$$y=1, \quad p=1, \quad q^*_0=1-\frac{c}{\delta\, b}. \tag{5.26}$$

This strategy has been described earlier and is known as *Generous Tit-for-Tat* (*GTFT*)[Mol85, NS92]. Surprisingly the analogous strategy for indirect reciprocity ($\lambda=1$) has not been described before, given by

$$y=1, \quad p=1, \quad q^*_1=1-\frac{1+(n-2)\delta}{1+(n-2)(1-2\varepsilon)}\,\frac{c}{\delta\, b}. \tag{5.27}$$

In analogy to the previous case, we call this strategy *Generous Scoring* (*GSCO*). These two Nash equilibria do not need to exist for all parameter values because the respective value of q^* may become negative. Theorem 2 summarizes the necessary and sufficient conditions for cooperative Nash equilibria to exist.

Theorem 2 (Existence of cooperative Nash equilibria).

1. There is a cooperative Nash equilibrium $\sigma=(1,1,q_0^*,0)$ *in which players exclusively use direct information if and only if* $\delta \geq \delta_0$ *with*

$$\delta_0 = \frac{c}{b}. \tag{5.28}$$

2. There is a cooperative Nash equilibrium $\sigma=(1,1,q_1^*,1)$ *in which players use indirect information if and only if* $2\varepsilon < 1-c/b+1/(n-2)$*, and* $\delta \geq \delta_1$ *with*

$$\delta_1 = \frac{c}{b+(n-2)\Big((1-2\varepsilon)b-c\Big)}. \tag{5.29}$$

3. For $0<\lambda<1$*, there is a cooperative Nash equilibrium* $\sigma=(1,1,q_\lambda^*,\lambda)$ *if and only if there is a cooperative Nash equilibrium for* $\lambda=0$ *or* $\lambda=1$*.*

Theorem 2 gives three major insights:

First, for sustaining cooperation in a Nash equilibrium there is no advantage of using both direct and indirect information simultaneously (i.e., to choose $0<\lambda<1$). From the third part of Theorem 2 it follows that if some cooperative Nash equilibrium in reactive strategies exists at all, there is always a cooperative Nash equilibrium for either $\lambda=0$ or $\lambda=1$.

Second, the first two parts of Theorem 2 suggest that for $\delta_0, \delta_1 \in [0,1]$ we can partition the parameter space of the game into four distinct regions:

(*i*) When $\delta < \min\{\delta_0, \delta_1\}$, there is no cooperative Nash equilibrium;

(*ii*) When $\delta_0 < \delta < \delta_1$, full cooperation can be sustained with direct but not with indirect reciprocity;

(*iii*) When $\delta_1 < \delta < \delta_0$, cooperation can be sustained with indirect but not with direct reciprocity;

(*iv*) When $\delta > \max\{\delta_0, \delta_1\}$ both direct and indirect reciprocity allow for stable cooperation.

We provide a graphical illustration of these parameter regions as **Fig. 5.3c** in the main text.

Third, parameter changes affect the feasibility of cooperation as follows.

Corollary 1. *If a cooperative Nash equilibrium exists for given values* n*,* δ*, and* ε*, then there also exists a cooperative Nash equilibrium for any* $n' \geq n$*,* $\delta' \geq \delta$*, and* $\varepsilon' \leq \varepsilon$*.*

In a nutshell, the above corollary states that the stability of cooperation is monotonic in the population size, the pairwise continuation probability, and the error rate. All other parameters kept equal, the conditions for a cooperative Nash equilibrium are easier to meet if the population size is larger, if there are more rounds to be played in expectation, or if there is less noise on indirect information.

Several remarks are in order.

1. While the above results for $\lambda = 0$ fully recover previous results on direct reciprocity among reactive players[Mol85, NS92], the results on indirect reciprocity ($\lambda = 1$) may come as a surprise. Previous work on indirect reciprocity suggests that reactive ('first-order') strategies are unable to sustain cooperation[LH01, PB03, OI04]. The intuition for this pessimistic result is that in a resident population of perfect discriminators (with $y = p = 1$, $q = 0$), a resident may have no incentive to discriminate against a rare $ALLD$ mutant. By punishing a defector, discriminators harm their own reputation, which puts them at risk to receive less cooperation in future. Our model circumvents this problem by allowing for stochastic strategies. By increasing their q-value from zero, discriminators are able to reduce the effective cost of punishment. Once they reach $q = q^*_\lambda$, the expected long-term loss in reduced reputation after defecting against an $ALLD$ mutant exactly matches the short-term gain in saved cooperation costs. In the Nash equilibrium, discriminators no longer bear an effective cost of punishing defectors.

2. The strategies Generous Tit-for-Tat and Generous Scoring are Nash equilibria, but they are not evolutionarily stable. That is, if the entire population adopts one of these strategies, no mutant is favored to invade, but mutants are not opposed to invade either. Evolutionary stability is generally difficult to achieve in repeated interactions. For the repeated prisoner's dilemma, it has been shown that no strategy is evolutionary stable in the absence of errors[BL87, FW89]. Any resident strategy can be invaded by 'stepping-stone' mutations, which in turn may facilitate the invasion of further mutant strategies[GvV16]. Simulations suggest that cooperation comes and goes in cycles; the frequencies of cooperative and defecting strategies oscillate over time[GvV18]. However, how often cooperative strategies are played over an evolutionary timescale critically depends on whether or not cooperative Nash equilibria exist[vVGRN12]. We will revisit this issue in the next section.

   For indirect reciprocity it has been suggested that evolutionary stability is feasible if more complex strategies are permitted[OI04, OI06]. The respective 'leading-eight' strategies require that the reputation of a player does not only depend on a player's action, but additionally on the reputation of the opponent (and sometimes also on the player's own previous reputation). While these results of Ohtsuki and Iwasa have been tremendously important for understanding which properties stable norms ought to have, their framework is based on some rather restrictive assumptions. Most importantly, information transmission is assumed to be public. In particular, all population members agree on the reputation they assign to a given co-player. In contrast, we allow some players to use a co-player's public record to assign reputations, whereas others may make their judgments solely based on their direct experiences. As a consequence, different players may assign a different reputation to the same co-player. For such a private information scenario, it has been recently shown that not even the leading-eight strategies are able to sustain high cooperation rates in the presence of perception errors[HST$^+$18].

3. Previous results on the stability of cooperation with indirect reciprocity only show stability *within* the given strategy set considered [OI04, MVC13, BG16, SON17]. This leaves it open whether more complex strategies could invade, either due to neutral drift or because of a selective advantage. In contrast, the equilibrium results presented in this section do not restrict which strategies mutants are permitted to

take. The Nash equilibria that we describe are equilibria with respect to *all possible* mutant strategies (see Remark 1).

4. In the limit of rare errors, the conditions that we obtain for the feasibility of cooperation with direct and indirect reciprocity are strict, as the following result shows.

Theorem 3 (Optimality of equilibrium conditions).
Consider a population of size n interacting in a population game as introduced in Section 5.5.2, and suppose $\varepsilon \to 0$. Then there exists a Nash equilibrium (not necessarily in reactive strategies) that yields full cooperation if and only if there is a cooperative Nash equilibrium in reactive strategies.

That is, our focus on simple reactive strategies does not restrict the player's ability to sustain cooperation. Theorem 3 says that if full cooperation is feasible at all in the limit of rare errors, it is already feasible with reactive strategies.

With the above equilibrium analysis we explore which strategies can maintain cooperation if the respective strategy is already adopted by a large majority of the population. Importantly, however, *ALLD* is also an equilibrium for all considered parameter values (Theorem 1). Therefore, even if cooperative equilibria exist, our results do not imply that evolving populations would settle at these equilibria. Before we discuss the corresponding evolutionary questions in the next section, we highlight a few reasons why we deem equilibrium analyses, like the one provided above, valuable: (*i*) The existence of cooperative equilibria can be considered a minimum requirement for cooperation to evolve. (*ii*) Our analytical equilibrium conditions in Theorem 2 allow predictions on how certain parameter changes affect the feasibility of cooperation (see Corollary 1). (*iii*) Even if we are to find that cooperative Nash equilibria have problems to evolve, this may have important policy implications. For example, such a finding could suggest that temporarily, additional incentives for cooperation are necessary.

5.5.5 Evolutionary analysis

Description of the evolutionary process

In the previous section, we have considered a static setup. We have explored which reactive strategies can sustain an equilibrium with full cooperation. In the following, we no longer assume that players settle at an equilibrium automatically. Instead players may choose arbitrary reactive strategies. Over time, they adapt their behaviors through social learning.

To model this adaptation dynamics, we consider a pairwise comparison process[TNP06] in a well-mixed population of constant size n. Initially, all players are assumed to use some arbitrary strategy $(y_0, p_0, q_0, \lambda_0)$. Then in every time step of the evolutionary process, two possible events can happen.

Exploration event. With probability μ (the mutation rate), one of the players is chosen at random. All players have the same probability to be chosen. This player then adopts a randomly chosen new strategy (y', p', q', λ'). The new values for y', p', q' are drawn from a uniform distribution on the unit interval. For λ' we first focus on the case that players either use direct or indirect reciprocity, such that $\lambda' \in \{0,1\}$. We discuss the case of intermediate values of λ' in **Section 5.5.5**.

Imitation event. With probability $1-\mu$, two players are randomly drawn from the population, a learner and a role model. Given the composition of the population, their expected payoffs π_L and π_R are calculated using Eq. (5.19). We assume the learner adopts the role model's strategy with probability

$$\xi = \frac{1}{1+e^{-\beta(\pi_R-\pi_L)}} \tag{5.30}$$

The parameter $\beta \geq 0$ reflects the strength of selection. It determines to which extent the learner's decision depends on the role model's relative success. In the special case that $\beta = 0$, the imitation probability simplifies to $\xi = 1/2$, such that imitation occurs purely at random. As β becomes larger, the learner is increasingly biased to only imitate role models with a higher payoff.

The above two elementary events, exploration and imitation, give rise to an ergodic process on the space of all population compositions. We explore this process with computer simulations. For each simulation, we use a fixed set of parameter values. Simulations are run for sufficiently long, such that the population's average cooperation rate has converged and is independent of the initial population. In addition to this cooperation rate, we record which strategies the players use over time.

Such simulations are computationally expensive. Even when using optimized algorithms (see Section 5.5.7), they need to keep track of all s strategies in the population and to invert $s^2 \times s^2$ matrices at each selection event. To make these computations more efficient, we have employed the method of Imhof and Nowak[IN10] for some of our results. This method assumes that mutations are rare, $\mu \to 0$. In that case, once a mutant enters the population, this mutant can be expected to reach fixation or to go extinct before the next mutation arises [FI06, WGWT12]. As a result, at any given time at most two different strategies are present in the population. The mutant's fixation probability in a resident population with strategy σ_R can be calculated explicitly, using the formula [TH09]

$$\rho(\sigma_M, \sigma_R) = \frac{1}{1+\sum_{i=1}^{n-1}\prod_{k=1}^{i} e^{-\beta\,(\pi_M(k)-\pi_R(k))}}. \tag{5.31}$$

Here, $\pi_M(k)$ and $\pi_R(k)$ are the respective payoffs when k individuals in the population employ the mutant strategy. Once the mutant has gone extinct, or taken over the population, the next mutation occurs. This rare-mutation limit allows for more efficient computation of the resulting dynamics. The saved computation time can then be used, for example, to run the process for a larger number of generations. In this way we can ensure that sufficiently many mutant strategies are drawn over the course of a simulation to obtain a representative sample of the entire strategy space.

Overall, the process generates a sequence of strategies $(\sigma_0, \sigma_1, \ldots)$ that the residents use after each introduced mutant strategy. Based on this sequence, we can again calculate the average cooperation rate and the average strategy traits that evolve for different parameter values.

Impact of parameters on the evolutionary results

As summarized in **Table S1**, our model includes a number of different parameters. Throughout most of the main text, we have focussed on two of them, (*i*) the game's continuation probability δ, and (*ii*) the probability ε that indirect observations are subject to perception errors. In addition, our model depends on the benefit-to-cost ratio b/c of cooperation; the population size n; the strength of selection β; and the mutation rate μ. In the following, we describe the effect of each of these parameters on the co-evolution of cooperation and reciprocity. In particular, we focus on two aspects. (1) How do changes in the respective parameter affect the evolution of cooperation? (2) How do changes in the respective parameter affect the proportion of players who use indirect reciprocity? In each case, we first use our static equilibrium results and the previous literature to form expectations on the impact of the respective parameter. After that, we compare these expectations with the actual behavior observed in simulations.

(i) Continuation probability δ.

Because the expected number of rounds between two given players is given by $1/(1-\delta)$, the parameter δ determines how often two players interact on average. Based on our static equilibrium analysis (Corollary 1 in Section 5.5.4), larger values of δ should facilitate the evolution of cooperation. With respect to the abundance of indirect reciprocity, our equilibrium analysis suggests that indirect reciprocity should be particularly favoured for intermediate values of δ, where direct reciprocity has difficulties to evolve. For sufficiently large continuation probabilities, our static equilibrium analysis does not allow us to make a clear prediction, because both *GTFT* and *GSCO* are equilibria for $\delta \to 1$.

We have performed two sets of simulations in which we have varied the continuation probability systematically, one for the limit of rare mutations (**Fig. 5.7d**) and one for a positive mutation rate (**Fig. 5.7h**). In both cases, our simulation results are in line with the above predictions. For any error rate ε, cooperation becomes more likely as we increase δ. Moreover, indirect reciprocity is most abundant when δ is intermediate. For $\delta \to 1$, we find that direct reciprocity is more likely to emerge. We provide some analytical insight into this last effect in Section 5.5.5.

(ii) Error rate ε on indirect observations.

Based on our equilibrium analysis, we would expect that an increase in ε tends to reduce cooperation (Corollary 1 in Section 5.5.4). This reduction may be less severe for large continuation probabilities, where players may switch to direct reciprocity instead. With respect to the abundance of indirect reciprocity, we expect errors to either have no effect (e.g., when no cooperation evolves and information is irrelevant), or to have a negative effect.

Again, there are two sets of evolutionary simulations in which we vary the error rate ε systematically, see **Fig. 5.7d,h**. In general, these simulations are in agreement with the above expectations. In case the error rate has an effect on cooperation at all (i.e., for intermediate δ), this effect is negative. Similarly, in those cases in which the error rate affects the player's propensity to use indirect reciprocity (i.e., for intermediate to large values of δ), this effect is again negative.

The next four parameters are not specific to models of reciprocity. They feature an important role in any evolutionary model of cooperation. Some of these parameters are known to have non-trivial effects on simulation outcomes. For example, it has been shown that strategy abundances may change non-monotonically in population size and selection strength [NSTF04, WGHT13, WD19]. Similarly, an increase in the system's mutation rate can change the evolutionary dynamics altogether [THDS+09]. Analytical predictions for these effects are usually only possible in simple systems with a few strategies. In the following, we therefore only provide some basic intuition.

(iii) Benefit-to-cost ratio b/c.

According to Theorem 2, a larger benefit-to-cost ratio is favorable to both, a *GTFT* equilibrium and a *GSCO* equilibrium. We would thus expect that an increase in b/c enhances cooperation. With respect to the abundance of indirect reciprocity, our equilibrium analysis does not make any predictions.

We explore the effect of the benefit-to-cost ratio on evolution in **Fig. 5.8a,b**. There, we distinguish between three scenarios, depending on how many interactions occur on average and on how noisy third-party information is. In all scenarios, we find that cooperation is an increasing function in b/c. The effect of b/c on the abundance of indirect reciprocity is more irregular. However, for all benefit-to-cost ratios, we observe that an environment with intermediately many interactions and reliable information is more favorable to indirect reciprocity than an environment with many interactions and unreliable information.

(iv) Population size n.

The effect of n is more difficult to predict because population size affects the dynamics on multiple levels. First, according to Eq. (5.29), it affects the stability of conditional cooperation under indirect reciprocity. The intuition here is that the larger the population, the more likely players already have some third-party information when they encounter a given co-player for the first time. Increases in population size would therefore increase cooperation, because players can more easily avoid helping an unconditional defector. However, this effect should become weaker once the population is already sufficiently large. Second, the size of a population affects the evolutionary process. On the one hand, small populations tend to select for spiteful behaviors [NSTF04, Hil11]. The intuition here is that the smaller a population, the easier it becomes for players to show an above-average performance by simply diminishing the payoff of all their co-players. On the other hand, the dynamics in small populations is more stochastic. Even if a strategy is slightly disfavored, it still has a reasonable chance to reach fixation due to stochastic effects if the population is sufficiently small.

Based on these mechanisms, we can expect the following behavior. For the smallest meaningful population size, $n=2$, cooperation is disfavored by the evolutionary process (i.e., the cooperation rate is smaller than 1/2), because of the effects of spite. Moreover, since direct and indirect reciprocity are indistinguishable for $n=2$, both should have equal frequency. While these predictions for $n=2$ hold for all scenarios, the predictions for $n>2$ will depend on the other model parameters. However, here again one may predict that scenarios with sufficiently many interactions tend to favor cooperation. Scenarios with intermediately many interactions and reliable information tend to favor indirect reciprocity.

Our simulations for varying population sizes indeed follow this general pattern (**Fig. 5.8c,d**). We note that in comparably large populations ($n > 500$), cooperation rates are non-monotonic when simulations are run for a short time-span (green and orange curves in **Fig. 5.8d**). In large populations, it takes on average more time until any given population is invaded. Therefore, the outcome of short simulations is more dependent on the initial population. Because the defection rewarding zone of our strategy space is larger than the cooperation rewarding zone (Section 5.5.4), non-cooperative populations are initially favored. This may explain the decline of cooperation for larger population sizes when simulations are run for a short time span.

(v) Selection strength β .

Selection strength only affects the evolutionary process. Therefore we cannot directly infer its effects based on our static equilibrium predictions. Still we can form some intuition in the following way. In the most simple case of $\beta = 0$, payoffs are irrelevant for the spread of a strategy. For such a neutral process, we expect the average cooperation rate and the average abundance of indirect reciprocity to approach 50%. For small but positive selection strengths, we would expect cooperation to be slightly disfavored in all scenarios. The intuition for this is as follows: Because selection is weak, all strategies are played with almost equal abundance. Because the best response to a majority of strategies is *ALLD* (Section 5.5.4), this effect is expected to give a slight advantage to non-cooperative strategies. For intermediate to strong selection, we expect the outcome to depend again on the scenario. In scenarios with sufficiently many interactions, cooperation can evolve. In addition, provided that cooperation evolves, we expect more indirect reciprocity when there are intermediately many interactions and information is reliable.

We systematically vary selection strength in the simulations shown in **Fig. 5.8e,f**. Again, the results seem to generally agree with the intuition provided above. As with population size, we find that the effect of selection strength on cooperation can be non-monotonic. However, this time, we observe this non-monotonicity even in the case that simulations are run for a long time (orange curve in **Fig. 5.8e**; for the respective simulations, we have consecutively introduced $5 \cdot 10^7$ mutant strategies into the population. We have checked with additional simulations that 10^8 mutant strategies yield the same output).

In **Fig. 5.9**, we explore this non-monotonicity in more detail. Contrary to what one may expect, we do not observe that this non-monotonicity is due to a reduced robustness of cooperators (**Fig. 5.9a**). Moreover, mutants that succeed in invading a conditionally cooperative resident are highly cooperative themselves (**Fig. 5.9d**). As a result, highly cooperative populations actually become more abundant as we increase selection strength. At the same time, however, we find that highly non-cooperative strategies also increase, and at a faster rate than the highly-cooperative strategies (**Fig. 5.9c–e**). Combined, these effects result in overall smaller cooperation rates under strong selection.

(vi) Mutation rate μ .

In several of our figures we consider the limit of rare mutations [FI06, WGWT12, IN10] (e.g, **Fig. 5.4**,**Fig. 5.7a–d**). That is, we assume the mutation rate is vanishingly small, such that at any given point in time, the population employs at most two different strategies. This assumption comes with several mathematical and computational advantages[FI06] and it has been employed in many studies of reciprocity [SSP16, SP13, GT12b, HST+18] and beyond [HTB+07, SDSTH10, HT12]. At the same time, such an assumption entails the risk that important mixed equilibria are overlooked, because they cannot be reached by the evolutionary process. To explore the robustness of our rare mutation results, we have thus performed additional simulations for positive mutation rates (e.g., **Fig. 5.5**, **Fig. 5.7e–h**). There we have observed that our rare-mutation results do not require mutation rates to be exponentially small (as assumed by the respective theory [WGWT12]). Instead, we observe a reasonable agreement between the rare-mutation limit and our simulations for positive mutation rates once μ is of the order $1/n$ or smaller.

In **Fig. 5.8g,h**, we further explore the effect of mutation rates. To gain some intuition for the effect of mutation rates, consider first the limit $\mu \to 1$. In this limit, strategy abundances are again independent of the strategies' payoffs. Every strategy is therefore played with equal likelihood. Both, the average cooperation rate as well as the average abundance of indirect reciprocity simplify to 1/2. If μ is sufficiently close to one but smaller than one, we would again expect that cooperation is slightly disfavored (because *ALLD* is a best response to a uniformly random population). While these predictions apply for all scenarios, the behavior of the system for small mutation rates depends on the exact game parameters. As mutations become sufficiently rare, we expect to recover our main text results based on the method by Imhof and Nowak[IN10] for the limit $\mu \to 0$ (as depicted in **Fig. 5.7a–d**).

Our simulations in **Fig. 5.8g,h** largely follow this intuition. Interestingly, however, we again observe that cooperation based on direct reciprocity is more robust than cooperation based on indirect reciprocity when mutations occur at an intermediate rate (**Fig. 5.5**, **Fig. 5.7**). As noted in the main text, this can occur because in indirect reciprocity, the payoffs of conditional cooperators increase nonlinearly in the number of other conditional cooperators. We explore this non-linearity in more detail in the next section.

An analysis of the competition between defectors and cooperators

According to the simulations, we observe a certain bias in favor of direct reciprocity when interactions are common (i.e., when $\delta \to 1$). Surprisingly, this bias exists even if indirect information is completely reliable (i.e., when $\varepsilon \to 0$). A static equilibrium perspective cannot explain this effect. After all, Theorem 2 suggests that whenever cooperation can be sustained with direct reciprocity, it can also be sustained with indirect reciprocity if the error rate is sufficiently small. In the following, we aim to provide an evolutionary argument for this phenomenon. This argument also sheds some light on why cooperative strategies of indirect reciprocity may have problems to evolve for larger mutation rates.

To this end, let us assume that players can only choose between two strategies, a cooperative strategy $(1,1,q,\lambda)$ and $ALLD = (0,0,0,\lambda)$. We explore the competition between these two strategies for both $\lambda = 0$ and $\lambda = 1$ in the limit of no errors, $\varepsilon \to 0$.

1. **Direct reciprocity ($\lambda = 0$).**
   In a population with k cooperators and $n-k$ defectors, we can use Eq. (5.19) to

explicitly calculate the players' payoffs as

$$\begin{aligned}\pi_C^0 &= \frac{k-1}{n-1}\cdot(b-c) - \frac{n-k}{n-1}\cdot\Big((1-\delta)+\delta q\Big)\cdot c\\ \pi_D^0 &= \frac{k}{n-1}\cdot\Big((1-\delta)+\delta q\Big)\cdot b.\end{aligned} \tag{5.32}$$

Importantly, both players' payoffs are linearly increasing in the number of cooperators k (see **Fig. 5.6a,c** for an illustration of π_C^0 and π_D^0 for two different values of δ). We can simplify these payoff expressions further if we define $z := k/n$ to be the fraction of conditional cooperators in the population, and let $n \to \infty$. In that case, payoffs become

$$\begin{aligned}\pi_C^0 &= (b-c)\cdot z - (1-\delta+\delta q)c\cdot(1-z)\\ \pi_D^0 &= (1-\delta+\delta q)b\cdot z.\end{aligned} \tag{5.33}$$

For $z \to 0$, cooperators always obtain a lower payoff than defectors. In the other limit, $z \to 1$, cooperators outperform defectors if $q < 1-c/(\delta b)$. Under that condition, the dynamics is bistable: each strategy is favored if it is adopted by sufficiently many players. The minimum fraction of cooperators to make cooperation beneficial can be calculated as

$$z^0 = \frac{1-\delta+\delta q}{(1-q)\delta}\cdot\frac{c}{b-c}. \tag{5.34}$$

This critical threshold becomes arbitrarily small as the number of repetitions increases ($\delta \to 1$), and as the cooperative strategy approaches *TFT* ($q \to 0$). That is, if only there are sufficiently many repetitions, already a small minority of *TFT* players can easily invade into an *ALLD* population.

2. **Indirect reciprocity ($\lambda = 1$).**
   Again, we consider k conditional cooperators with strategy $(1, 1, q, 1)$ and $n-k$ defectors with strategy $(0, 0, 0, \lambda)$. According to Eq. (5.68), their respective payoffs are now

$$\begin{aligned}\pi_C^1 &= \frac{k-1}{n-1}\;\frac{1+\Big((n-1)q-1\Big)\delta}{1+(n-2)\delta}\;\frac{1+\Big(n-2+(n-k)(1-q)\Big)\delta}{1+\Big(n-2-(k-1)(1-q)\Big)\delta}\cdot(b-c) - \frac{n-k}{n-1}\;\frac{1+\Big((n-1)q-1\Big)\delta}{1+(n-2)\delta}\cdot c.\\ \pi_D^1 &= \frac{k}{n-1}\cdot\frac{1+\Big((n-1)q-1\Big)\delta}{1+(n-2)\delta}\cdot b\end{aligned} \tag{5.35}$$

   The defectors' payoffs still depend linearly on the cooperators' generosity parameter q and on the number of cooperators k. However, the cooperators' payoffs are nonlinear (see **Fig. 5.6b,d**). If we again define $z := k/n$, and let $n \to \infty$, the above payoffs simplify to

$$\begin{aligned}\pi_C^1 &= \frac{q+q(1-q)(1-z)}{1-(1-q)z}\cdot z(b-c) - q(1-z)c\\ \pi_D^1 &= q\cdot zb\end{aligned} \tag{5.36}$$

   We note that $\partial\pi_C^1/\partial z > 0$ and $\partial^2\pi_C^1/\partial z^2 > 0$ for $q > 0$. That is, the payoff of cooperators increases disproportionally in the number of other cooperators. When $z \to 0$, cooperators are again outperformed by the defectors. Conversely, when $z \to 1$, cooperators succeed if $q<1-c/b$. In that case, there is again a bistable competition between cooperators and defectors. The critical threshold for cooperators to be favored is now

$$z^1 = \frac{c}{b(1-q)}. \tag{5.37}$$

In contrast to the case of direct reciprocity, we note that this critical threshold is independent of the continuation probability. In particular, even when the game is infinitely repeated, indirect reciprocity always requires a critical mass of conditional cooperators present in the population for cooperation to succeed. The critical mass required is at least c/b.

The above results highlight that direct and indirect reciprocity differ in their relative strengths.

On the one hand, indirect reciprocity leads to a faster spread of information. As a consequence, conditional cooperators are better able to restrict the payoff of defectors. In fact, we find $\pi_D^1 \leq \pi_D^0$ for all parameter values. This faster spread of information is particularly important if players only interact for a few number of rounds. For $\delta \to 1$, this comparative advantage of indirect reciprocity disappears, and $\pi_D^1 = \pi_D^0$.

On the other hand, successful cooperation in indirect reciprocity is based on synergy effects. Cooperators only obtain a high payoff if they are present in sufficient numbers. That is, $\pi_C^1 < \pi_C^0$ for sufficiently large δ. Overall, a comparison of the thresholds for direct and indirect reciprocity yields

$$z^0 < z^1 \quad \Leftrightarrow \quad \delta \geq \frac{b}{b-c+(1-q)b}. \tag{5.38}$$

That is, once the continuation probability is sufficiently high, direct reciprocity requires fewer cooperative mutants to undermine an *ALLD* population.

Evolutionary dynamics for probabilistic information usage

In our previous analysis of the evolutionary dynamics, players incorporate any third-party information according to a deterministic rule. They either never take such information into account ($\lambda = 0$) or they always do so ($\lambda = 1$). In the following, we consider an evolutionary process where individuals can interpolate between these two extremes.

To this end, we need to adapt the evolutionary process as described in Section 5.5.5 to allow for mutant strategies with intermediate values of λ. One rather immediate way to incorporate values of λ is to assume that new mutant strategies (y', p', q', λ') are uniformly drawn from the four-dimensional unit cube $[0,1]^4$. Such a mutation scheme, however, would predominantly result in mutant strategies that rarely engage in direct reciprocity. To see why, we note that in a population of size n, two players have on average $n-2$ third-party interactions each before they have a direct encounter again. As a result, even players with comparably small values of λ have quite a high probability to update a co-player's reputation state between two direct encounters.

In the following, we formalize the above intuition. To this end, consider a player with strategy (y, p, q, λ) who just had a direct interaction with some given co-player. Let $\bar{\gamma}$

denote the probability that the player does not update the co-player's reputation until the next direct interaction of the two players. This probability can be computed as follows:

$$\bar{\gamma} = \sum_{k=0}^{\infty} \left(\frac{n-2}{n-1}\right)^k \frac{1}{n-1}(1-\lambda)^k = \frac{1}{1+(n-2)\lambda}. \tag{5.39}$$

In sum on the right hand side, $\big((n-2)/(n-1)\big)^k/(n-1)$ is the probability that the co-player engages in k third party interactions until the next direct interaction happens. The remaining factor $(1-\lambda)^k$ is the probability that the focal player decides not to update the co-player's reputation after each of these third-party interactions. The resulting quantity $\bar{\gamma}$ reflects how likely the player's decision in the next game with the co-player is based on their last game with each other. We thus refer to $\bar{\gamma}$ as the effective probability that a player engages in direct reciprocity. Similarly, we refer to

$$\gamma := 1-\bar{\gamma} = \frac{(n-2)\lambda}{1+(n-2)\lambda} \tag{5.40}$$

as the player's effective probability to engage in indirect reciprocity. In particular, for a player with $\lambda=0$ who ignores any third-party information, we obtain $\gamma=0$ as one may expect. On the other hand, if $\lambda=1$, then $\gamma_{\max}:=(n-2)/(n-1)$; in this case, the only time the player's reaction is based on direct experience is when the co-player happens not to have any third-party interactions between two direct encounters (we revisit this case in **Section 5.5.6**).

Importantly, the effective probability to engage in indirect reciprocity according to Eq. (5.40) is nonlinear in λ. For example, even if a player is rather unlikely to take into account any particular piece of third-party information (e.g., $\lambda=10\%$) may have a considerable effective probability to engage in indirect reciprocity ($\gamma\approx 82.8\%$ for a moderate population of size $n=50$). In particular, for simulations in which λ is chosen uniformly between [0,1], most mutant strategies would predominantly engage in indirect reciprocity. For our simulations in **Fig. 5.10**, we thus do not employ such a mutation scheme. Instead, we generate mutant strategies (y,p,q,λ) such that the resulting γ according to Eq. (5.40) is distributed uniformly in $[0,\gamma_{\max}]$. In this way, we ensure that a randomly generated mutant is approximately equally likely to engage in direct and in indirect reciprocity.

For the simulations, we first consider the case that the player's value of γ is fixed. We then let γ vary between $\gamma=0$ and $\gamma=\gamma_{\max}$. For **Fig. 5.10a**, we have re-run the simulations shown in **Fig. 5.4a** for the case that players are comparably unlikely to interact again, $\delta=1/2$. Here we observe that players are most likely to adopt cooperative strategies when γ is large (i.e., when players have a large effective probability to engage in indirect reciprocity). Conversely, for **Fig. 5.10b**, we consider the limit of infinitely many pairwise interactions, $\delta\to 1$. As already suggested by **Fig. 5.4b**, our results here indicate that direct reciprocity is more effective in generating high cooperation rates. While these two results are obtained in the limit of rare mutations, we obtain similar qualitative results for more abundant mutations (**Fig. 5.10c,d**). When $\delta=1/2$, strategies with higher γ generate more cooperation, irrespective of the considered mutation rate. On the other hand, when $\delta=1$, strategies with smaller γ are more conducive to the evolution of cooperation (which reflects our earlier results in **Fig. 5.5**). We note that these simulations make a similar prediction as the theoretical results in our Theorem 2. The results suggest that in general, intermediate values of λ are not more favorable to cooperation. Instead, the highest cooperation rates are achieved either for $\lambda = 0$ (i.e., $\gamma = 0$) or for $\lambda = 1$ (i.e., $\gamma = \gamma_{\max}$).

In a second step, we then allow the value of γ to co-evolve along with the other strategy parameters y, p, and q. Repeating the simulations done for **Fig. 5.7a–c**, we consider three different scenarios. (*i*) If errors are abundant and players only interact for a few rounds, we observe that no cooperation evolves and that selection among the player's γ values is neutral (**Fig. 5.10e**). (*ii*) If errors are rare and players interact for an intermediate number of rounds, cooperation can evolve (**Fig. 5.10f**). In that case, we also observe that players tend to have a higher effective probability to engage in indirect reciprocity. In particular, among all possible values of γ, players most abundantly choose $\gamma \approx \gamma_{\max}$. (*iii*) If errors occur at an intermediate rate and players interact often, players show a bias towards sustaining cooperation based on direct reciprocity (**Fig. 5.10g**). Now the most abundant strategies adopt $\gamma \approx 0$.

In a last step, we have systematically varied the considered error rate and the continuation probability (**Fig. 5.10h**). Compared to the baseline case in **Fig. 5.7d**, we observe some notable differences. On the one hand, when errors are abundant, players seem to have more difficulties to establish cooperation. On the other hand, also the average value of γ seems to be closer to 1/2 (note that the color legend in **Fig. 5.10h** is rescaled to allow for a better contrast between the different regions). Nevertheless, the general patterns outlined in **Fig. 5.7d** are conserved. Cooperation is most likely to evolve when the continuation probability is large and when errors are rare. At the same time, players show a bias towards indirect reciprocity if they only interact for intermediately many rounds, whereas they prefer direct reciprocity when there are many rounds and many errors.

We note that we can gain insight regarding the range of viable λ that can evolve to sustain cooperation independent of the mutation scheme, simply by taking another look at Eqs. (5.21-5.24). To this end, let $0 \leq \lambda \leq 1$ be arbitrary, and consider general form of the equalizer strategy $\sigma = (1, 1, q^*_\lambda, \lambda)$ with $q^*_\lambda := 1 - r^*_\lambda$ and r^*_λ, as defined by Eq. (5.23). By requiring that $q^* > 0$, we get the following condition on δ for the respective cooperative Nash equilibrium to exist:

$$\delta_\lambda = \frac{c}{b + \lambda(n-2)\Big((1-2\varepsilon)b - c\Big)}. \tag{5.41}$$

We can however also express this as two different conditions on λ. For one, when $\varepsilon < (1-c/b)/2$ and $\delta > \delta_1$, we get

$$\lambda > \frac{c/\delta - b}{b + \lambda(n-2)\Big((1-2\varepsilon)b - c\Big)}. \tag{5.42}$$

For $\varepsilon > (1-c/b)/2$ and $\delta > \delta_0$, this is

$$\lambda < \frac{c/\delta - b}{b + \lambda(n-2)\Big((1-2\varepsilon)b - c\Big)}. \tag{5.43}$$

Equation 5.42 tells us that if cooperation is to evolve in a scenario with few errors, the evolving values of λ need to be above a certain threshold (i.e. there is a lower bound on how likely players are to engage in indirect reciprocity). Conversely, if errors are abundant, the evolving values of λ need to be sufficiently small. (i.e. there is a lower bound for how likely the players' actions are based on direct reciprocity). This gives additional intuition for the behavior of our model beyond the restriction to the extremal values $\lambda = 0$ and $\lambda = 1$.

5.5.6 Model extensions

The impact of errors and incomplete information

Our baseline model introduced in Section 5.5.2 is based on a number of simplifying assumptions. In particular, we assumed that only third-party interactions are subject to perception errors, that player always implement the correct action, that both actions (C and D) are equally likely to be misperceived, and that all players can observe all interactions in a population. In the following, we discuss the impact of each of these assumptions separately.

Direct reciprocity with perception errors. We first discuss how our results need to be adapted when players misperceive the actions of a direct interaction partner with probability $\varepsilon_0 > 0$. For better clarity, we now refer the probability that players misperceive third-party interactions as ε_1. The baseline model corresponds to the case $\varepsilon_0 = 0$ and $\varepsilon_1 = \varepsilon$. In general, it is reasonable to assume that indirect observations are more prone to misinterpretation than direct observations, in which case $\varepsilon_0 \leq \varepsilon_1$. When direct reciprocity is subject to perception errors, the main equation (5.12) becomes

$$\begin{aligned} x_{ij}(t+1) = {} & (1-\bar{w})\, x_{ij}(t) \\ & + w\Big(\big((1-\varepsilon_0)x_{ji}(t)+\varepsilon_0(1-x_{ji}(t))\big)\, p_i + \big((1-\varepsilon_0)(1-x_{ji}(t)) + \varepsilon_0 x_{ji}(t)\big)\, q_i\Big) \\ & + (\bar{w}-w)\,(1-\lambda_i)\, x_{ij}(t) \\ & + \lambda_i\, w \sum_{l\neq i,j} \big[\big((1-\varepsilon_1)x_{jl}(t) + \varepsilon_1(1-x_{jl}(t))\big)\, p_i + \big((1-\varepsilon_1)(1-x_{jl}(t)) + \varepsilon_1 x_{jl}(t)\big)\, q_i\big]. \end{aligned} \tag{5.44}$$

We note that the new parameter ε_0 only affects the second line of this recursion, which captures the direct interaction between i and j (**Fig. 5.2f**). As in Section 5.5.2, we can use this equation to calculate the players' average cooperation rates and payoffs. Again, we are interested in which strategies are able to sustain cooperation. To this end, we need to slightly modify our definition of cooperative Nash equilibria. We say a strategy $\sigma = (y, p, q, \lambda)$ is a cooperative Nash equilibrium if

(i) σ is a Nash equilibrium, and
(ii) the cooperation rate in a homogeneous σ-population approaches one as both $\varepsilon_0 \to 0$ and $\varepsilon_1 \to 0$.

Based on recursion (5.44), the equilibrium results of Section 5.5.4 can be adapted as follows.

Theorem 4 (Cooperative Nash equilibria when direct interactions are subject to perception errors)**.**

1. *A strategy* $\sigma=(y,p,q,\lambda)$ *is a cooperative Nash equilibrium if and only if*

$$y=1, \quad p=1, \quad q=1-r^*_\lambda, \quad \text{with} \quad r^*_\lambda=\frac{1+(n-2)\delta\lambda}{(1-2\varepsilon_0)+(n-2)(1-2\varepsilon_1)\lambda}\cdot\frac{c}{\delta b}. \qquad (5.45)$$

2. *In particular, cooperation can be sustained through direct reciprocity* ($\lambda=0$) *if and only if* $\delta\geq\delta_0$ *with*

$$\delta_0=\frac{1}{1-2\varepsilon_0}\cdot\frac{c}{b}. \qquad (5.46)$$

It can be sustained through indirect reciprocity ($\lambda=1$) *if and only if* $\delta\geq\delta_1\in[0,1]$ *where*

$$\delta_1=\frac{c}{(1-2\varepsilon_0)\,b+(n-2)\Big((1-2\varepsilon_1)\,b-c\Big)}. \qquad (5.47)$$

In **Fig. 5.3d**, we illustrate the equilibrium conditions (5.46) and (5.47) for the special case $\varepsilon_0=\varepsilon_1$. As one may expect, it is more difficult to sustain cooperation when also direct interactions are subject to perception errors. This increased difficulty is reflected in the additional $(1-2\varepsilon_0)$-terms in the denominator of equations (5.46) and (5.47). However, the impact of ε_0 on the threshold value δ_1 for indirect reciprocity vanishes as the population size n becomes large. Again, this is intuitive: if a players' reputation is based on his entire behavior, errors that only affect direct encounters become negligible in large populations.

To explore how perception errors in direct interactions affect the evolutionary dynamics of strategies, we have re-run the simulations in **Fig. 5.7d**, but now assuming that $\varepsilon_0=\varepsilon_1=\varepsilon$. As shown in **Fig. 5.11b**, there are two effects. First, players are now less likely to cooperate, especially for high error rates. Second, players become more likely to take indirect information account (as it is now just as reliable as direct information).

Implementation errors. In addition to perception errors, models of direct and indirect reciprocity often consider an alternative source of noise in the form of implementation errors or trembling hand errors[NSES95, Pos99]. To account for these kinds of errors, we assume that when a player intends to cooperate (defect), he mistakenly defects (cooperates) with probability e. Under this assumption, Eq. (5.12) for player i's probability to be in the good state with respect to co-player j becomes

$$\begin{aligned}
x_{ij}(t+1) &= (1-\bar{w})\,x_{ij}(t)\\
&\quad + w\Big(\big((1-e)x_{ji}(t)+e(1-x_{ji}(t))\big)\,p_i + \big((1-e)(1-x_{ji}(t))+ex_{ji}(t)\big)\,q_i\Big)\\
&\quad + (\bar{w}-w)\,(1-\lambda_i)\,x_{ij}(t)\\
&\quad + \lambda_i\,w\sum_{l\neq i,j}\Big[\Big((1-e)(1-\varepsilon)x_{jl}(t)+(1-e)\varepsilon(1-x_{jl}(t))+e(1-\varepsilon)(1-x_{jl}(t))+e\varepsilon x_{jl}(t)\Big)\,p_i\\
&\qquad + \Big(e(1-\varepsilon)x_{jl}(t)+e\varepsilon(1-x_{jl}(t))+(1-e)(1-\varepsilon)(1-x_{jl}(t))+(1-e)\varepsilon x_{jl}(t)\Big)\,q_i\Big].
\end{aligned} \qquad (5.48)$$

Here, the second and fourth line have changed in comparison to Eq. (5.12). For example, the second line reflects that in direct interactions, we need to distinguish four cases, depending on whether or not the co-player intended to cooperate, and whether or not there was an implementation error. Also the payoff formula (5.19) needs to be adapted accordingly,

$$\pi_i = \frac{1}{n-1}\sum_{j\neq i}\Big((1-e)x_{ji}+e(1-x_{ji})\Big)\,b-\Big((1-e)x_{ij}+e(1-x_{ij})\Big)\,c. \qquad (5.49)$$

Again, we are interested in those Nash equilibria in which everyone is fully cooperative as errors become rare (that is, when $\varepsilon\to 0$ and $e\to 0$).

Theorem 5 (Cooperative Nash equilibria under implementation errors)**.**

1. *A strategy $\sigma=(y,p,q,\lambda)$ is a cooperative Nash equilibrium if and only if*

$$y=1,\quad p=1,\quad q=1-r^*_\lambda,\quad \textit{with}\quad r^*_\lambda=\frac{1+(n-2)\delta\lambda}{(1-2e)\Big(1+(n-2)(1-2\varepsilon)\lambda\Big)}\cdot\frac{c}{\delta b}. \qquad (5.50)$$

2. *In particular, cooperation can be sustained through direct reciprocity if and only if $\delta\geq\delta_0$ with*

$$\delta_0=\frac{1}{1-2e}\cdot\frac{c}{b}. \qquad (5.51)$$

It can be sustained through indirect reciprocity if and only if $\delta\geq\delta_1\in[0,1]$ where

$$\delta_1=\frac{c}{(1-2e)\,b+(n-2)\Big((1-2e)(1-2\varepsilon)\,b-c\Big)}. \qquad (5.52)$$

Interestingly, when we compare Theorem 5 with the respective conditions in Section 5.5.4, we note that the effect of implementation errors is equivalent to a rescaling of the benefit parameter from b to $(1-2e)b$. That is, with respect to the static equilibrium predictions, increasing the rate of implementation errors by 10% has the same effect as reducing the benefit of the game by 20%. To explore the effect of implementation errors on the evolution of cooperation, we have re-run the simulations in **Fig. 5.7d**, but now for $e=0.01$ instead of $e=0$ (**Fig. 5.11c**). As one may expect, the addition of implementation errors renders the evolution of cooperation more difficult. In the new scenario, players need to interact for a larger number of rounds to establish substantial cooperation rates.

Asymmetric errors. One way how misperceptions can occur is when the acting player herself has an incentive to misrepresent her action. This is particularly relevant for defectors, who may wish to conceal the true nature of their actions. Excellent models of strategic miscommunication in indirect reciprocity can be found in the papers by Nakamaru and Kawata and by Seki and Nakamaru [NK04, SN16]. In the context of our model we can approximate the workings of strategic miscommunication by assuming that the two possible actions C and D have different probabilities to be misperceived, ε_C and ε_D. A scenario where defectors are more likely to misrepresent their actions can then be represented by assuming $\varepsilon_D>\varepsilon_C$. For asymmetric errors, we obtain the following recursion for the pairwise probability to assign a good reputation to a co-player,

$$\begin{aligned}x_{ij}(t+1) &= (1-\bar{w})\,x_{ij}(t)\\ &\quad + w\,(x_{ji}(t)\,p_i+(1-x_{ji}(t))\,q_i)\\ &\quad + (\bar{w}-w)\,(1-\lambda_i)\,x_{ij}(t)\\ &\quad + \lambda_i\sum_{l\neq i,j} w\,\big[((1-\varepsilon_C)x_{jl}(t)+\varepsilon_D(1-x_{jl}(t)))\,p_i+((1-\varepsilon_D)(1-x_{jl}(t))+\varepsilon_C x_{jl}(t))\,q_i\big].\end{aligned} \qquad (5.53)$$

The players' payoffs are again defined by Eq. (5.19). In the case of asymmetric errors, we speak of a strategy as a cooperative Nash equilibrium if it is stable and if the cooperation rate against itself approaches one if both $\varepsilon_C \to 0$ and $\varepsilon_D \to 0$. We obtain the following characterization:

Theorem 6 (Cooperative Nash equilibria under asymmetric errors)**.**

1. *A strategy $\sigma=(y,p,q,\lambda)$ is a cooperative Nash equilibrium if and only if*

$$y=1, \quad p=1, \quad q=1-r^*_\lambda, \quad \textit{with} \quad r^* = \frac{(1+(n-2)\delta\lambda)}{1+(n-2)\,(1-\varepsilon_D-\varepsilon_C)\lambda}\frac{c}{\delta b}. \tag{5.54}$$

2. *In particular, cooperation can be sustained through direct reciprocity if and only if $\delta \geq \delta_0$ with*

$$\delta_0 = \frac{c}{b}. \tag{5.55}$$

It can be sustained through indirect reciprocity if and only if $\delta \geq \delta_1 \in [0,1]$ where

$$\delta_1 = \frac{c}{b+(n-2)\Big((1-\varepsilon_D-\varepsilon_C)b-c\Big)} \tag{5.56}$$

In case $\varepsilon_C = \varepsilon_D$, these conditions recover the respective results of the baseline model, which is reassuring. In **Fig. 5.11d**, we show simulation results for the asymmetric case where $\varepsilon_C=0$ and $\varepsilon_D=\varepsilon>0$. For the parameter range considered previously, $0\leq\delta\leq 1$ and $0.001\leq\varepsilon\leq 0.1$, we observe more cooperation compared to our baseline model. This occurs because the effective rate at which errors happen is now reduced, because only some of the players' actions are subject to noise.

However, we note that in contrast to the previous scenarios, it is not unreasonable to assume that error probabilities may exceed 10% when they are the result of strategic miscommunication. Larger values of ε_D can impede cooperation considerably. In fact, even if $\varepsilon_C=0$, it follows from Eq. (5.56) that indirect reciprocity is no longer able to sustain cooperation if $\varepsilon_D > (1-c/b)(n-1)/(n-2)$. That is, for cooperation to evolve through indirect reciprocity, there need to be limits on how well defectors can deceive others.

Incomplete information. In our baseline model we assume that each player is informed about everyone else's interactions. This assumption of complete information though standard is unrealistic, especially if the population is large. In the following, we thus introduce an additional parameter $\nu \in [0,1]$. This parameter reflects which fraction of third-party interactions players observe on average. With incomplete information, the main equation (5.12) now becomes

$$\begin{aligned} x_{ij}(t{+}1) = {} & (1-\bar{w})\,x_{ij}(t) \\ & + w\,\Big(x_{ji}(t)\,p_i + (1-x_{ji}(t))\,q_i\Big) \\ & + (\bar{w}-w)\,(1-\nu\lambda_i)\,x_{ij}(t) \\ & + w\nu\lambda_i \sum_{l\neq i,j}\bigg(\Big(1-\varepsilon\Big)x_{jl}(t)+\varepsilon\Big(1-x_{jl}(t)\Big)\bigg)\,p_i + \bigg(\Big(1-\varepsilon\Big)\Big(1-x_{jl}(t)\Big)+\varepsilon x_{jl}(t)\bigg)\,q_i. \end{aligned} \tag{5.57}$$

Here, the third and fourth line have changed in comparison to Eq. (5.12). That is, a player's probability λ_i to take into account indirect information is now scaled by a factor of ν, the probability that this information is learned in the first place. A player with $\lambda_i=1$ no longer takes every action of the other group members into account when assigning reputations, but every action he is aware of. The condition for cooperation to be feasible is now given as follows.

Theorem 7 (Cooperative Nash equilibria under incomplete information)**.**

1. *A strategy $\sigma=(y,p,q,\lambda)$ is a cooperative Nash equilibrium if and only if*

$$y=1, \quad p=1, \quad q=1-r^*_\lambda, \quad \textit{with} \quad r^*_\lambda=\frac{1+(n-2)\delta\nu\lambda}{1+(n-2)(1-2\varepsilon)\nu\lambda}\cdot\frac{c}{\delta b}. \tag{5.58}$$

2. *The condition for cooperation under direct reciprocity is unchanged, $\delta\geq\delta_0=c/b$. For indirect reciprocity there exists a cooperative Nash equilibrium if and only if $\delta\geq\delta_1\in[0,1]$ where*

$$\delta_1=\frac{c}{b+(n-2)\Big((1-2\varepsilon)\,b-c\Big)\nu}. \tag{5.59}$$

In the limiting case that players obtain no third-party information at all, such that $\nu=0$, the two threshold values coincide, $\delta_0=\delta_1$, as expected. In **Fig. 5.11d**, we present evolutionary simulation results when there is almost no third-party information ($\nu=0.01$, such that only every 100th interaction is observed on average). In that case, players who take indirect information into account are almost indistinguishable from those players who do not. Indeed, we find that cooperation only evolves once it can be sustained through direct reciprocity, and that indirect reciprocity is generally disfavored.

A model of pure indirect reciprocity

Our baseline model makes two assumptions on how individuals assign reputations to each other: (*i*) Individuals always update a co-player's reputation after a direct interaction, and (*ii*) if the respective co-player interacts with a third party, the co-player's reputation is updated with probability λ. These assumptions imply that a player with $\lambda=0$ acts based on direct experiences only, whereas a player with $\lambda=1$ equally takes into account all of a co-player's interactions, no matter whether the focal player was directly involved. Since interactions occur randomly, however, it may happen that the same two players are chosen to interact for two consecutive rounds. In such a case, the players' decisions in the second round are necessarily based on their respective direct experience made in the first round – even if players generally base their decisions on indirect reciprocity (that is, even if $\lambda=1$). On a conceptual level, this effect may make it more difficult to compare the two modes of reciprocity. It can thus be useful to study a model where individuals use pure versions of each behavior (such that individuals who opt for indirect reciprocity neglect any additional direct information they may have). In the following we present such a model, and we show that it leads to qualitatively similar conclusions.

For this model, we assume the players' strategies are given by a 4-tuple (y,p,q,κ). The interpretation of the first three entries is the same as in the baseline model. For the last entry, we assume that if a focal player directly directly interacts with a given co-player, the co-player's reputation is updated with probability $1-\kappa$. In contrast, if the co-player interacts with a third-party, the co-player's reputation is updated with probability κ. This model implies that for $\kappa=0$ only direct interactions lead to a reputation update, whereas for $\kappa=1$ reputations are only updated after third-party interactions. Similar to Eq. (5.12) for the baseline model, we can derive a recursion for the probability $x_{ij}(t)$ that player i considers j to be good at time t. This recursion now takes the form

$$\begin{aligned} x_{ij}(t+1) = {} & (1-\bar{w})\, x_{ij}(t) \\ & + w(1-\kappa_i)\Big(x_{ji}(t)\, p_i + (1-x_{ji}(t))\, q_i\Big) \\ & + w\kappa_i \cdot x_{ij}(t) \\ & + (\bar{w}-w)\,(1-\kappa_i)\, x_{ij}(t) \\ & + w\kappa_i \sum_{l\neq i,j}\bigg(\Big(1-\varepsilon\Big)x_{jl}(t)+\varepsilon\Big(1-x_{jl}(t)\Big)\bigg)\, p_i + \bigg(\Big(1-\varepsilon\Big)\Big(1-x_{jl}(t)\Big) + \varepsilon x_{jl}(t)\bigg)\, q_i. \end{aligned} \tag{5.60}$$

This recursion only differs from Eq. (5.12) in the second and third line. These two lines both refer to direct interactions of players i and j, which happen with probability w. For the model considered here, direct interactions lead to a reputation update with probability $1-\kappa_i$. In that case, the new reputation depends on co-player j's action, as encoded by $x_{ji}(t)$, and on the values of p_i and q_i. Otherwise, with probability κ_i, player j's reputation remains unaffected.

Based on this recursion (5.60) can now derive the corresponding versions of all results of the baseline model. We summarize the respective findings in the following.

Theorem 8 (Cooperative Nash equilibria for the model with pure indirect reciprocity)**.** *Consider a group of size* $n\geq 3$.

1. *A strategy* $\sigma=(y,p,q,\kappa)$ *is a cooperative Nash equilibrium if and only if*

$$y=1, \quad p=1, \quad q=1-r^*_\kappa, \quad \textit{with} \quad r^*_\kappa=\frac{1+(n-3)\delta\kappa}{1-\kappa+(n-2)(1-2\varepsilon)\kappa}\frac{c}{\delta b}. \tag{5.61}$$

2. *In particular, if players only use direct reciprocity* ($\kappa=0$)*, we obtain the same condition for the existence of a cooperative Nash equilibrium as in the baseline model,* $\delta\geq\delta_0=c/b$*. For indirect reciprocity* ($\kappa=1$) *there is a cooperative Nash equilibrium if and only if* $\delta\geq\delta_1\in[0,1]$ *where*

$$\delta_1=\frac{c}{b(n-2)(1-2\varepsilon)-c(n-3)}. \tag{5.62}$$

For any $\kappa>0$, we note that the equilibrium value of q according to Eq. (5.61) approaches $q=1-c/(1–2\varepsilon)/b$ in the limit of large populations. This limit coincides with the respective limit of the baseline model according to Eq. (5.27). This is unsurprising: when populations are sufficiently large, the probability that two particular players interact with each other twice in a row becomes negligible. In that case, already our baseline model yields a model of pure indirect reciprocity. But even for finite population sizes, the two models typically yield similar predictions. For example, for the parameters in **Fig. 5.4** ($b/c=5$, $n=50$, $\varepsilon=0$) the baseline model suggests that full cooperation is feasible for $\delta_0=0.2$ (when $\lambda=0$) and $\delta_1=0.0051$ (when $\lambda=1$). For the alternative model considered here, the respective values become $\delta_0=0.2$ (when $\kappa=0$) and $\delta_1=0.0052$ (when $\kappa=1$), respectively. As a result, we also observe a similar dynamics in evolutionary simulations (see **Fig. 5.17**).

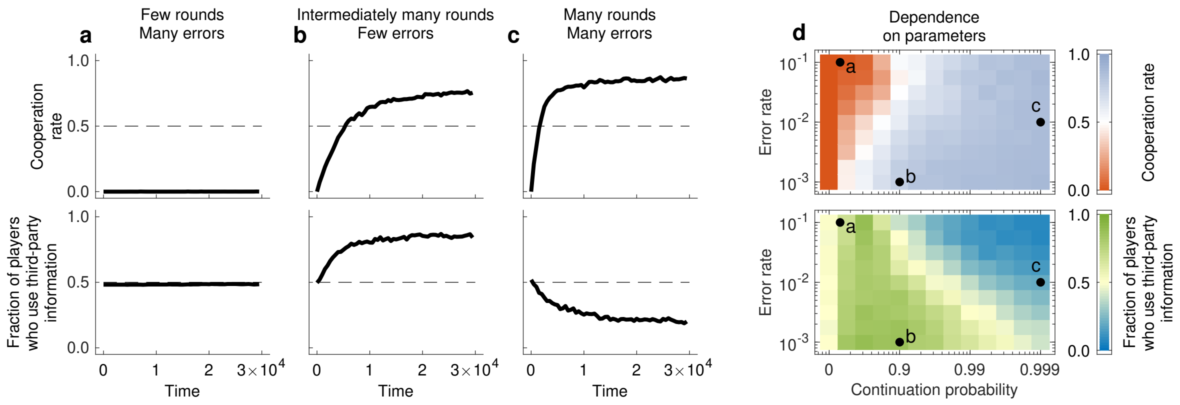

Figure 5.17: **The coevolution of conditional cooperation and information use in a model of "purified" indirect reciprocity.** For the model discussed in the main text, we assume that after a direct interaction, players always update their co-player's reputation. After a indirect interaction, they update their co-player's reputation with probability λ. These assumptions imply that even if $\lambda = 1$, players may occasionally react based on their direct experiences. This happens, for example, if two players have two consecutive interactions. In that case, the players' behavior in the second round will depend on the outcome of the first. While such an assumption seems realistic, it can be useful to compare purified versions of both kinds of reciprocity. We explore such a model in 5.5.6. There, the players' strategies take the form (y, p, q, κ). A value of $\kappa = 0$ means players ignore all of a co-player's third party interactions (similar to the case $\lambda = 0$ discussed before). A value of $\kappa = 1$, however, now means that players ignore all direct information they may have. In the limit of large populations, the two strategy spaces (y, p, q, λ) and (y, p, q, κ) yield identical results. Here we repeat the simulations performed in **Fig. 5.7a–d** for a finite population of $n = 50$, using the alternative strategy space (y, p, q, κ). The results in **a–d** correspond to and are almost indistinguishable from the results shown in **Fig. 5.7a–d**.

Finite-state automata with multiple states

So far we have studied the evolution of direct and indirect reciprocity in the most simple strategy space possible. Players can only assign two possible reputations to their co-players, and a player's updated reputation only depends on the last action taken into account. In the following two subsections, we sketch how our framework can be extended to scenarios that do not meet these two assumptions.

Strategies represented by finite-state automata with multiple states. We begin by allowing players to assign more nuanced reputations to their co-players. To this end, we consider players who use finite state automata with more than two states. In addition to the dichotomous characterization of a player being either 'good' or 'bad', this allows us to capture models in which players can have a third, 'unknown' or 'neutral' reputation. Such third states have been important in models in which players can forget [NM11]. Similarly, these automata also allow us to capture scenarios in which it takes multiple defections in memory to yield a bad reputation [Ber11, BG16].

For this model extension, we assume that each player's strategy takes the form $(\Omega, \gamma, \tau_0, \tau, \lambda)$. Here, the first component $\Omega = \{\omega_1, \ldots, \omega_m\}$ is the number of (reputational) states the player does distinguish. The second component $\gamma : \Omega \to \{C, D\}$ determines for each assigned reputational state whether or not to cooperate with a co-player who has the respective reputation. The third component $\tau_0 \in \Delta^\Omega$ determines the initial reputational state of the co-players. Here, $\Delta^\Omega = \{x \in [0,1]^m \mid x_1 + \ldots + x_n = 1\}$ is the set of probability distributions over the set of states. The fourth component $\tau : \Omega \times \{C, D\} \to \Delta^\Omega$ is the transition function. It determines how likely the updated reputation of a given co-player is $\omega' \in \Omega$, depending on the previously assigned reputation $\omega \in \Omega$ and on the co-player's action $a \in \{C, D\}$. As before, we assume that the co-player's reputation is always updated after a direct interaction. Whether or not the co-player's reputation is also updated if the co-player interacts with a third party depends on the player's receptivity λ. Third party interactions are ignored when $\lambda = 0$, whereas they are fully included when $\lambda = 1$. In the first case, we again obtain a model with direct reciprocity only. In that case, our framework captures previous studies on repeated games among players with finite state automata[vVGRN12, GvV16, GvV18]. If $\lambda = 1$, we obtain a model of indirect reciprocity. In that case, we can capture previous models in which players can have more than two possible reputations[TSM13, NM11, NS98b].

The reactive strategies considered in the previous sections represent a special case of these finite state automaton strategies. For a reactive strategy $\sigma = (y, p, q, \lambda)$ we have

$$\Omega = \{G, B\}, \qquad \gamma(\omega) = \begin{cases} C & \text{if } \omega = G \\ D & \text{if } \omega = B \end{cases}, \qquad \tau_0 = (y, 1-y), \qquad \tau(\omega, a) = \begin{cases} (p, 1-p) & \text{if } a = C \\ (q, 1-q) & \text{if } a = D. \end{cases} \tag{5.63}$$

From this representation, the two important simplifications of reactive strategies become immediately apparent: (*i*) reputations are binary, and (*ii*) when a co-player's reputation is updated, the new reputation only depends on the co-player's action but not on her previous reputation. This latter assumption has been crucial for the explicit calculation of the players' payoffs in Section 5.5.3. Due to this assumption, the pairwise reputation variable $x_{ij}(t+1)$ can be written as a linear function of the respective quantities $x_{kl}(t)$ in the previous round. This no longer needs to hold for more general finite state automata, where transitions depend on both the current state and the respective co-player's action. Instead of calculating the payoffs explicitly, we instead use numerical simulations of the game dynamics to obtain the players' payoffs in the following.

Three examples of finite-state automata with multiple states. We illustrate this approach with three examples (**Fig. 5.12**). In each case, we assume the respective finite state automaton has three states, 'good' (G), 'neutral' (N), and 'bad' (B). Initially, players assign a good reputation to everyone. Each time the co-player defects, the co-player's reputation deteriorates (from good to neutral, or from neutral to bad, respectively). Similarly, each time the co-player cooperates, his reputation improves. Players are assumed to cooperate against good opponents and to defect against bad opponents. The three automata differ in how they deal with opponents with a neutral reputation: we consider the case that players may either cooperate ($A1$, **Fig. 5.12a**), cooperate with 50% probability ($A2$, **Fig. 5.12b**), or defect ($A3$, **Fig. 5.12c**). The automata $A1$ and $A3$ have been referred to as 'generous discriminator' and 'rigorous discriminator', respectively[TSM13].

Stability of the three-state automata against ALLD and ALLC. In a next step we considered a scenario in which n–1 residents employ the respective automaton strategy. The remaining player either employs *ALLD* or *ALLC*. Keeping the population composition fixed, we simulate the players' payoffs for various values of $\lambda \in [0,1]$ (**Fig. 5.12d–f**). We used the fixed game parameters $b=5$, $c=1$, $n=50$, and $\varepsilon=0.05$. To compute the payoffs, we use the same simulation scheme as in a previous study[HST+18]. That is, initially, all players are assumed to assign a good reputation to each other. Then we simulate a game with $2\cdot10^6$ rounds. To compute the players' payoffs , we average over the second half of these rounds; in this way, we avoid any transient effect during the early rounds arising from our assumption on the initial reputation assignments. We checked that averaging over all rounds would not alter our conclusions.

For direct reciprocity only ($\lambda=0$), we observe that all of the three automaton strategies are stable with respect to invasion by *ALLD*, and neutrally stable with respect to invasion by *ALLC*. However, once indirect information is considered ($\lambda>0$), $A2$ and $A3$ can both be invaded by *ALLC*. Moreover, if residents adopt $A3$, they fail to cooperate with each other altogether. Only the first automaton $A1$ is stable against *ALLC* and *ALLD* for all considered λ values. Automaton $A1$ can be considered as a threshold strategy in the sense of Berger and Grüne[BG16]: players are certain to defect only if the co-player defected twice in a row. For appropriate parameter values, such threshold strategies can sustain cooperation although they only rely on first-order information[Ber11, BG16]. Our results in **Fig. 5.12d** suggest that the same is true when players blend direct and indirect reciprocity (i.e., for $0<\lambda<1$).

Evolutionary competition between the three-state automata, ALLD, and ALLC. In a final step, we sketch how our framework can be used to explore the evolutionary performance of finite-state automata which depend on both, direct and indirect information. In the following, we assume players use a fixed receptivity $\lambda=0.1$ (but it should become clear how our methods extend to more general settings). Players can choose between three different strategies: *ALLD*, *ALLC*, and one of the three automata discussed above.

In contrast to reactive strategies, there is no known efficient payoff formula for arbitrary finite-state automata in the context of indirect reciprocity with private information. We therefore determined the players' payoffs by simulation. To this end, we considered all possible population compositions (k_A, k_C, k_D). Here, k_A is the number of players who use the respective finite-state automaton strategy, k_C is the number of unconditional cooperators, and k_D is the number of defectors. The possible population compositions are those for which $k_A+k_C+k_D=n$. Based on these payoffs, we again employ the process described in Section 5.5.5.

First, we considered the dynamics for $\beta=1$ in the limit of rare mutations, $\mu\to0$, for each of the three finite-state automata. For the given parameter values, we find that only the first automaton A_1 is played with substantial frequency (**Fig. 5.13a–c**). We observe that rare defectors are disfavored to invade into an A_1 population. However, cooperators can invade through (almost) neutral drift, because they get a payoff only slightly lower than the residents'. Because unconditional cooperators are in turn easily invaded by *ALLD*, there is a cyclic dynamics. Most of the time, the population either adopts A_1 (65.4%) or *ALLD* (32.8%). As a result, the average cooperation rate over time is below 70%. The other two finite-state automata generate even lower cooperation rates; A_2 is quickly invaded by *ALLC*, whereas A_3 can be invaded by both, *ALLC* and *ALLD*.

Next we have explored the dynamics for a mutation rate of $\mu=0.01$, such that different strategies may coexist in a population. For all three automata we find that most of the time, the population either is in the vicinity of *ALLD*, or in the vicinity of the edge between the the finite-state automaton and *ALLC* (**Fig. 5.13d–f**). How often these two neighborhoods are visited depends on the considered automaton strategy. For A_1 and A_2, we observe substantial cooperation, whereas for A_3 players adopt $ALLD$ most of the time.

In a final step we have re-run these calculations for different parameter values. We have varied the benefit of cooperation, the selection strength, and the mutation rate. As baseline parameters we took the values used in (**Fig. 5.13d–f**). For all parameter values, we find that A_1 is most likely to induce cooperation (**Fig. 5.13g–i**). However, also the second automaton can yield substantial cooperation rates.

Most strikingly, while the mutation rate had a negative impact on cooperation for reactive strategies (**Fig. 5.5**), here we find that an intermediate mutation rate is most favorable to cooperation. However, in our view the two sets of simulations need to be compared with caution. In **Fig. 5.5**, we have considered the entire space of all strategies of a given complexity. In particular, the space is balanced: for every conditional strategy that becomes more cooperative if the co-player cooperates, there is an analogous strategy that reduces its cooperation rate. Only under this assumption, an increase of mutation rates does not per se change the resulting cooperation rate. In contrast, the simulations in this section are based on only a small sample of possible strategies, and the strategy space is not balanced. This shortcoming is not specific to our **Fig. 5.13**, but it is rather common in the indirect reciprocity literature[NS98b, LH01, PB03, BS04, PSC06, US09, MVC13, UYOS18, HST$^+$18]. One reason for this prevalence of unbalanced models is the difficulty of deriving the players' payoffs. In the absence of a general payoff equation that applies to all strategies of a given complexity, researchers have to focus on strategies that seem most relevant. Such an approach does not naturally induce balanced strategy sets. In this sense, we consider the simulations performed for our baseline model as more transparent. They allow for all strategies of some given complexity, without any preselection on the part of the researcher.

Higher-order strategies

Another class of strategies that has received considerable attention in the literature on indirect reciprocity (but less so in direct reciprocity) is the class of higher-order social norms [OI04, NS05, Sig12]. The first-order social norms correspond to the reactive strategies considered in the baseline model. When two players interact, then the updated reputation of each player only depends on whether or not that player has cooperated. Second-order norms additionally take the reputation of the respective opponent into account. An example of such a second-order norm is stern judging [PSC06, SSP16, SSP18]. This norm suggests that to maintain a good reputation, one should cooperate with good opponents and defect against bad opponents. Finally, in third-order norms the updated reputation of some player additionally depends on the player's previous reputation. In that way, higher-order social norms take an increasing number of information into account when assigning a reputation to another group member.

Using an exhaustive method, Ohtsuki and Iwasa have shown that under public information, there are exactly eight deterministic third-order social norms that can maintain cooperation [OI04, OI06]. These norms are called the leading-eight. They consist of two components, an assessment-rule and an action rule (see **Fig. 5.14a**). The assessment rule determines how to update the reputation of other group members, depending on their actions, their previous reputation, and the reputation of the opponent. The action rule determines whether to cooperate; this decision may depend on one own's reputation, as well as on the reputation of the opponent. Crucially, Ohtsuki and Iwasa assume public information. In the following, we explore the performance of the leading-eight rules when reputations are based on both direct interactions and indirect observations, and when information is private.

Representation of third-order strategies. To this end, we now assume the players' strategies have the form $\sigma_i = (\alpha_i, \beta_i, \lambda_i)$. The first component is player i's assessment rule,

$$\alpha = (\alpha_{GCG},\ \alpha_{GCB},\ \alpha_{BCG},\ \alpha_{BCB},\ \alpha_{GDG},\ \alpha_{GDB},\ \alpha_{BDG},\ \alpha_{BDB}) \in \{G, B\}^8. \tag{5.64}$$

For example, an entry of $\alpha_{GCB} = G$ means that i deems it as good if a good population member j cooperates with a bad population member k. Importantly, whether or not j and k had a good and bad reputation to start with, depends on the individual perspective of player i (that is, it depends on $x_{ij}(t)$ and $x_{ik}(t)$ at time t). The second component of player i's strategy corresponds to his action rule,

$$\beta = (\beta_{GG},\ \beta_{GB},\ \beta_{BG},\ \beta_{BB}) \in \{C, D\}^4. \tag{5.65}$$

For example, if player i considers himself to be in a good standing but his co-player to be bad, then an action rule with $\beta_{GB} = D$ would prescribe to defect. Again, these assessments need to be taken from i's individual perspective. To this end, we assume that each player i now also has one additional automaton that captures i's own reputation. For this automaton, we set $x_{ii}(t) = 0$ if player i considers himself to be bad at time t (which depends on i's assessment rule and his previous interactions). Otherwise, if $x_{ii}(t) = 1$, player i deems himself as good. We assume the automaton that captures player i's self-perception is updated every time player i interacts with a co-player. With regard to the reputation of others, we assume that i updates a co-player j's reputation after a direct interaction with probability 1, and after an indirect observation with probability λ_i, as before.

Stability of the leading-eight strategies against ALLD and ALLC. To explore the stability of third-order social norms in the presence of both direct and indirect reciprocity, we proceed along the same lines as in the previous section on finite-state automata. Again, we first consider a population in which $n-1$ players adopt some leading-eight strategy L_i. The remaining player either adopts *ALLC* or *ALLD*. To compute the player's payoffs for various values of λ, again we simulate the game dynamics for fixed game parameters $b = 5$, $c = 1$, $n = 50$, and $\varepsilon = 0.05$. Our results for $\lambda > 0$ resemble previous findings on indirect reciprocity [HST+18]: in the presence of perception errors, all leading-eight strategies are susceptible to invasion. Either a single *ALLC* player or a single *ALLD* player obtains a higher payoff than the residents (**Fig. 5.14b–i**). Only for $\lambda = 0$ (when perception errors are absent), the leading-eight strategies are stable against both mutant strategies.

Evolutionary competition between the leading-eight, ALLD, and ALLC. In a next step, we have again explored how the leading-eight perform in an evolutionary context, when competing with the two unconditional strategies. We use exactly the same setup as in our previous analysis on finite-state automata. That is, we fixed a receptivity value $\lambda = 0.1$ for all population members. Then we pre-computed the payoffs with simulations for all possible population compositions (k_L, k_C, k_D) where $k_L + k_C + k_D = n$ (using the above payoff parameters). Based on these payoffs, we first considered the rare-mutation dynamics. As in a previous analysis for indirect reciprocity only[HST+18], we find that all of the leading-eight strategies have problems to persist in the population (**Fig. 5.15a–h**). The only strategy that achieves notable frequencies in the population is L_8. However, L_8 is sensitive to noise, and thus tends to defect in the presence of errors.

In a next step, we have considered the evolutionary process for a positive mutation rate, $\mu=0.01$. As we observed for the finite-state automaton strategies, populations tend to be clustered in one of two regions of the space of all population compositions (**Fig. 5.15i–p**). Either most players adopt *ALLD*, or the population consists of a mixture of leading-eight players and unconditional cooperators. Cooperative strategies are most abundant for three of the eight cases, for L_1, L_2, and L_7, as observed previously in the case of indirect reciprocity only [HST+18].

Finally, we have again explored how often players cooperate on average as we change three key parameters, the benefit of cooperation, the strength of selection, and the mutation rate. Across all parameters considered, we find that only the three previously identified strategies L_1, L_2, and L_7 can result in a population that predominantly cooperates (**Fig. 5.15q–s**). For these strategies we find more cooperation when the mutation rate is intermediate, for $0.01 \leq \mu \leq 0.1$. This again disagrees with the findings of our baseline model, according to which such mutation rates are detrimental to indirect reciprocity. However, as explained before, the two scenarios should not be compared directly. The simulations for our baseline model consider a complete and balanced strategy space. In contrast, the simulations **Fig. 5.15** only allow for three particular strategies, none of which being anti-reciprocal (i.e., none of which reducing its cooperation rate in response to a cooperative co-player). When a strategy space is not balanced, larger mutation rates may yield more cooperation even in the absence of selection.

To address such problems, it would be desirable to have simulations in which all third-order strategies (or strategies with different values of λ) compete. However, because there is no general payoff formula for third-order strategies under private information, such an approach is computationally out of reach. For our baseline model, we have thus analyzed a more elementary strategy space, the set of all (stochastic) first-order strategies. We find that even in this simpler strategy space, there is an unexpected strategy of indirect reciprocity that can maintain cooperation under appropriate conditions. This strategy is *GSCO* (Generous Scoring). *GSCO* does not require higher order information. All it requires is that players sometimes forgive when they witness a defection.

5.5.7 Appendix

Efficient computation of payoffs

Here we show how the payoffs according to Eq. (5.19) can be computed more efficiently by taking into account that players with the same strategy receive the same expected payoff.

To this end, suppose the population contains players with s different strategies in total. Let k_i denote the number of players using strategy i, such that $\sum_{i=1}^{s} k_i = n$. Slightly abusing our previous notation, we now refer to $x_{ij}(t)$ as the probability that a player with strategy i deems a player with strategy j as good at time t. When rewriting recursion (5.14) for this special case, we get two types of equations, depending on whether a player with strategy i meets a co-player with the same strategy or with a different one,

$$
\begin{aligned}
x_{ii}(t{+}1) &= \Big(1{-}w{-}\,\lambda_i(\bar{w}{-}w) + wr_i\Big(1{+}\lambda_i(1{-}2\varepsilon)(k_i{-}2)\Big)\Big)\cdot x_{ii}(t)\\
&\quad + w\lambda_i r_i(1{-}2\varepsilon)\cdot\sum_{l\neq i} k_l\cdot x_{il}(t) + \Big(w\,q_i + \lambda_i(\bar{w}{-}w)(\varepsilon r_i + q_i)\Big),\\
x_{ij}(t{+}1) &= \Big(1{-}w{-}\,\lambda_i(\bar{w}{-}w)\Big)\cdot x_{ij}(t) + wr_i\Big(1{+}\lambda_i(1{-}2\varepsilon)(k_i{-}1)\Big)\cdot x_{ji}(t)\\
&\quad + w\lambda_i r_i(1{-}2\varepsilon)\cdot\sum_{l\neq i}(k_l{-}1_{jl})\cdot x_{jl}(t) + \Big(w\,q_i + \lambda_i(\bar{w}{-}w)(\varepsilon r_i + q_i)\Big).
\end{aligned}
\tag{5.66}
$$

In the above equation, the symbol 1_{jl} is an indicator function; its value is one if $j=l$ and zero otherwise. Similar to the previous section, we can rewrite the above equation as $\mathbf{x}(t{+}1)=\mathbf{M}\,\mathbf{x}(t)+\mathbf{v}$, where $\mathbf{M}$ is now an $s^2{\times}s^2$ matrix, and $\mathbf{x}(t)$ and $\mathbf{v}$ are s^2-dimensional vectors. From this recursion, we can again compute the expected probability x_{ij} to find an i-player's automaton with respect to a j-player in a good state, using Eq. (5.18). The payoff π_i of a player with strategy i then becomes

$$
\pi_i = \sum_{j=1}^{s}\frac{k_j - 1_{ij}}{n-1}\cdot\Big(x_{ji}\cdot b - x_{ij}\cdot c\Big).
\tag{5.67}
$$

Especially when there are only a few different strategies in a large population, this algorithm can be expected to run considerably faster compared to the algorithm derived in **Section 5.5.3**.

For the limit of rare mutations and for the characterization of Nash equilibria it is useful to have explicit formulas for the special case that only $s=2$ strategies are present in the population (a resident and a deviating mutant). Suppose there are k players with strategy $\sigma_1=(y_1,p_1,q_1,\lambda_1)$ and $n{-}k$ players with strategy $\sigma_2=(y_2,p_2,q_2,\lambda_2)$. Then the matrix $\mathbf{M}$ takes the following form,

$$
\mathbf{M}=\begin{pmatrix}
\begin{matrix}1{-}w{-}\lambda_1(\bar{w}{-}w){+}\\ wr_1(1{+}\lambda_1(1{-}2\varepsilon)(k{-}2))\end{matrix} & w\lambda_1 r_1(1{-}2\varepsilon)(n{-}k) & 0 & 0\\
0 & 1{-}w{-}\,\lambda_1(\bar{w}{-}w) & wr_1(1{+}\lambda_1(1{-}2\varepsilon)(k{-}1)) & w\lambda_1 r_1(1{-}2\varepsilon)(n{-}k{-}1)\\
w\lambda_2 r_2(1{-}2\varepsilon)(k{-}1) & wr_2(1{+}\lambda_2(1{-}2\varepsilon)(n{-}k{-}1)) & 1{-}w{-}\,\lambda_2(\bar{w}{-}w) & 0\\
0 & 0 & w\lambda_2 r_2(1{-}2\varepsilon)k & \begin{matrix}1{-}w{-}\lambda_2(\bar{w}{-}w){+}\\ wr_2(1{+}\lambda_2(1{-}2\varepsilon)(n{-}k{-}2))\end{matrix}
\end{pmatrix}
$$

The vectors $\mathbf{v}$ and $\mathbf{x_0}$ are given by

$$
\mathbf{v}=\Big(wq_1{+}\lambda_1(\bar{w}{-}w)(\varepsilon r_1{+}q_1),\ wq_1{+}\lambda_1(\bar{w}{-}w)(\varepsilon r_1{+}q_1),\ wq_2{+}\lambda_2(\bar{w}{-}w)(\varepsilon r_2{+}q_2),\ wq_2{+}\lambda_2(\bar{w}{-}w)(\varepsilon r_2{+}q_2)\Big)^{\intercal},
$$

$$
\mathbf{x_0} = (y_1,y_1,y_2,y_2)^{\intercal}.
$$

The weighted average probability that an i-player considers a j-player as good again can be written as

$$
\mathbf{x} = (x_{11},x_{12},x_{21},x_{22}) = (\mathbf{1}{-}d\mathbf{M})^{-1}\Big((1{-}d)\mathbf{x_0} + d\mathbf{v}\Big).
$$

The equations for the two payoffs then become

$$\begin{aligned}\pi_1 &= \left(\frac{k-1}{n-1}\cdot x_{11} + \frac{n-k}{n-1}\cdot x_{21}\right) b - \left(\frac{k-1}{n-1}\cdot x_{11} + \frac{n-k}{n-1}\cdot x_{12}\right) c,\\ \pi_2 &= \left(\frac{k}{n-1}\cdot x_{12} + \frac{n-k-1}{n-1}\cdot x_{22}\right) b - \left(\frac{k}{n-1}\cdot x_{21} + \frac{n-k-1}{n-1}\cdot x_{22}\right) c.\end{aligned} \tag{5.68}$$

5.5.8 Proofs of the equilibrium results

Proof of Lemma 1. Let us calculate the converse probability $1-\delta$ that a given focal pair does not interact again after it has interacted in the current round. This case occurs if either the entire population game ends immediately after the current round; or the population game continues for one more round but the focal pair does not interact in that game; or the population game continues for two more rounds without the focal pair interacting, etc. That is, we obtain

$$1-\delta = (1-d) + d(1-d)(1-w) + d^2(1-d)(1-w)^2 + \ldots = \frac{1-d}{1-d(1-w)}. \tag{5.69}$$

Solving this equation for δ and using $w = 2/\big(n(n-1)\big)$ yields Eq. (5.20). □

Proof of Lemma 2. As introduced in Section 5.5.3, let $x_{ij}(t)$ denote the probability that player i considers player j to be good at time t (or equivalently, that player i would cooperate with j in round t). Assume for the moment that players apply arbitrary strategies $(y_i, p_i, q_i, \lambda_i)$. We consider the following quantity,

$$f_{ij}(T) := (1-d)\sum_{\tau=0}^{T} d^\tau \Big(d\cdot x_{ij}(\tau+1) - x_{ij}(\tau)\Big). \tag{5.70}$$

In this formula, we can express $x_{ij}(\tau+1)$ in terms of $x_{ij}(\tau)$ using Equation (5.12). This yields

$$\begin{aligned}f_{ij}(T) = (1-d)\sum_{\tau=0}^{T} d^\tau\cdot\Big[&-\Big(1-d+dw+d\lambda_i(\bar{w}-w)\Big)\cdot x_{ij}(\tau) + dwr_i\cdot x_{ji}(\tau)\\ &+d\lambda_i(1-2\varepsilon)r_i\cdot\sum_{l\neq i,j} wx_{jl}(\tau) + d\Big(wq_i+\lambda_i(\bar{w}-w)(\varepsilon r_i+q_i)\Big)\Big].\end{aligned} \tag{5.71}$$

Taking the limit as T becomes large, Eq. (5.71) becomes

$$\begin{aligned}\lim_{T\to\infty} f_{ij}(T) = &-\Big(1-d+dw+d\lambda_i(\bar{w}-w)\Big)\cdot x_{ij} + dwr_i\cdot x_{ji}\\ &+d\lambda_i(1-2\varepsilon)r_i\cdot\sum_{l\neq i,j} wx_{jl} + d\Big(wq_i+\lambda_i(\bar{w}-w)(\varepsilon r_i+q_i)\Big),\end{aligned} \tag{5.72}$$

where $x_{ij} = (1-d)\sum_{\tau=0}^{\infty} d^\tau x_{ij}(\tau)$ is again the weighted average probability that player i deems player j as good, as defined in Eq. (5.18).

On the other hand, $f_{ij}(T)$ takes the form of a telescopic sum. So we obtain an alternative representation of the limit by calculating

$$\lim_{T\to\infty} f_{ij}(T) = \lim_{T\to\infty} (1-d)\Big(d^T x_{ij}(T) - x_{ij}(0)\Big) = -(1-d)y_i. \tag{5.73}$$

Because the two right hand sides of Eq. (5.72) and Eq. (5.73) need to agree, we obtain

$$\Big(1-d+dw+d\lambda_i(\bar{w}-w)\Big)\cdot x_{ij} - dwr_i\cdot x_{ji} - d\lambda_i(1-2\varepsilon)r_i\cdot \sum_{l\neq i,j} wx_{jl} - d\Big(wq_i+\lambda_i(\bar{w}-w)(\varepsilon r_i+q_i)\Big) = (1-d)y_i. \tag{5.74}$$

Now, due to the assumptions of the lemma we are considering a homogeneous population such that all players employ the same strategy, $(y_i, p_i, q_i, \lambda_i) = (y, p, q, \lambda)$ for all i. Due to symmetry, it follows that $x_{ij} = x_{ji} = x_{jl} =: x$ for all i, j, l. In that case, Eq. (5.74) simplifies to

$$\Big(1-d+dw+d\lambda(\bar{w}-w)\Big)x - dwrx - d\lambda(1-2\varepsilon)r(\bar{w}-w)x - d\Big(wq+\lambda(\bar{w}-w)(\varepsilon r+q)\Big) = (1-d)y. \tag{5.75}$$

Solving this equation for x yields

$$x = \frac{(1-d)y + d\Big(wq + \lambda(\bar{w}-w)(\varepsilon r+q)\Big)}{(1-d) + d\Big(w(1-r) + \lambda(\bar{w}-w)(2\varepsilon r+1-r)\Big)}. \tag{5.76}$$

Plugging in the definitions $w=2/(n(n-1))$ and $\bar{w}=2/n$, and using the expression for δ derived in Lemma 1, Eq. (5.76) simplifies to Eq. (5.21). The second part of the lemma follows by solving Eq. (5.21) for $x=1$ and $x=0$, respectively. □

Proof of Lemma 3. Suppose without loss of generality that the first $n-1$ individuals apply the resident strategy (y, p, q, λ), whereas player n applies some arbitrary strategy (not necessarily reactive). Then each resident enforces a relationship of the form (5.74), with $i \in \{1, \ldots, n-1\}$ and $j=n$,

$$\Big(1-d+dw+d\lambda(\bar{w}-w)\Big)x_{in} - dwrx_{ni} - d\lambda(1-2\varepsilon)rw \sum_{l\neq i,n} x_{nl} = (1-d)y + d\Big(wq+\lambda(\bar{w}-w)(\varepsilon r+q)\Big). \tag{5.77}$$

Multiplying both sides of the equation with $1/(n-1)$ and summing over all $i \in \{1, \ldots, n-1\}$ yields

$$\Big(1-d+dw+d\lambda(\bar{w}-w)\Big)\sum_{i=1}^{n-1}\frac{x_{in}}{n-1} - dwr\sum_{i=1}^{n-1}\frac{x_{ni}}{n-1} - d\lambda(1-2\varepsilon)r(\bar{w}-w)\sum_{i=1}^{n-1}\frac{x_{ni}}{n-1} = (1-d)y + d\Big(wq+\lambda(\bar{w}-w)(\varepsilon r+q)\Big). \tag{5.78}$$

By rearranging the terms in Eq. (5.78), we can calculate how often the first $n-1$ players cooperate on average against the mutant player,

$$\sum_{i=1}^{n-1}\frac{x_{in}}{n-1} = \frac{(1-d)y + d\Big(wq+\lambda(\bar{w}-w)(\varepsilon r+q)\Big)}{\Big(1-d+dw+d\lambda(\bar{w}-w)\Big)} + \frac{dr\Big(w+\lambda(1-2\varepsilon)(\bar{w}-w)\Big)}{\Big(1-d+dw+d\lambda(\bar{w}-w)\Big)}\sum_{i=1}^{n-1}\frac{x_{ni}}{n-1}. \tag{5.79}$$

Note that according to Eq. (5.79), there is a linear relationship between how often residents cooperate on average against the mutant, and how often the mutant cooperates on average against each resident.

According to the payoff formula (5.19), the mutant's payoff is given by

$$\pi_n = \sum_{i=1}^{n-1} \frac{x_{in}}{n-1} b - \sum_{i=1}^{n-1} \frac{x_{ni}}{n-1} c. \tag{5.80}$$

We can then plug Eq. (5.79) into Eq. (5.80). Using the definitions $w = 2/(n(n-1))$ and $\bar{w} = 2/n$, and using the formula for δ derived in Lemma 1, we obtain after some rearranging,

$$\pi_n = A_1 + A_2(r - r^*_\lambda) \sum_{i=1}^{n-1} \frac{x_{ni}}{n-1}, \tag{5.81}$$

where

$$A_1 = \frac{(1-\delta)y + \delta\Big(q + (n-2)(q + r\varepsilon)\lambda\Big)}{1 + (n-2)\delta\lambda} \cdot b,$$

$$A_2 = \frac{1 + (n-2)(1 - 2\varepsilon)\lambda}{1 + (n-2)\delta\lambda} \cdot \delta b, \tag{5.82}$$

$$r^*_\lambda = \frac{1 + (n-2)\delta\lambda}{1 + (n-2)(1 - 2\varepsilon)\lambda} \cdot \frac{c}{\delta b}.$$

□

Proof of Theorem 1. Depending on $r = p - q$, we distinguish three cases:

(a) $r = r^*_\lambda$. In this case, Lemma 3 implies that any strategy yields the same payoff $\pi' = A_1$ against n–1 co-players with strategy σ. In particular, there is no profitable deviation from σ. Hence σ is a Nash equilibrium.

(b) $r < r^*_\lambda$. It follows from Lemma 3 that a strategy is a best response to σ if and only if it defects in every interaction. In particular, σ is a best response to itself if and only if a population of σ players is fully defecting. By Lemma 2, the population is fully defecting if and only if $y = p = q = 0$, or if $y = q = 0$ and either $\lambda = 0$, $n = 2$, or $\varepsilon = 0$.

(c) $r > r^*_\lambda$. For a strategy with $r > r^*_\lambda$, Lemma 3 implies that σ is a best response to itself if and only if a population of σ players is fully cooperative. While $y = p = q = 1$ is a fully cooperative strategy according to Lemma 2, this strategy does not satisfy $r = p - q > r^*_\lambda$. Hence, this case only permits a Nash equilibrium if $y = p = 1$, $q < 1 - r^*_\lambda$, and either $\lambda = 0$, $n = 2$, or $\varepsilon = 0$.

□

Proof of Theorem 2. The first two cases follow directly from Eqs. (5.26) and (5.27) by solving for $q \geq 0$. For the third case, consider a cooperative Nash equilibrium $\sigma = (1, 1, q^*_\lambda, \lambda)$ with $0 < \lambda < 1$. We calculate the partial derivative of $q^*_\lambda = 1 - r^*_\lambda$ with respect to λ,

$$\frac{\partial q^*_\lambda}{\partial \lambda} = \frac{(n-2)(1-\delta-2\varepsilon)}{\Big(1 + (n-2)(1-2\varepsilon)\lambda\Big)^2} \cdot \frac{c}{\delta b}. \tag{5.83}$$

We can distinguish three cases:

(i) $\delta = 1 - 2\varepsilon$. In this case, we obtain $\partial q_\lambda^* / \partial \lambda = 0$ for all values of λ. In particular, if $q_\lambda^* \geq 0$ for some given $\lambda \in (0,1)$, then also $q_0^* \geq 0$ and $q_1^* \geq 0$.

(i) $\delta < 1 - 2\varepsilon$. In this case we have $\partial q_\lambda^* / \partial \lambda > 0$ for all λ. In particular, if $q_\lambda^* \geq 0$ for some given $\lambda \in (0,1)$, then $q_1^* \geq 0$.

(i) $\delta > 1 - 2\varepsilon$. We conclude that $\partial q_\lambda^* / \partial \lambda < 0$ for all λ; if $q_\lambda^* \geq 0$ for $\lambda \in (0,1)$, then $q_0^* \geq 0$. □

Proof of Corollary 1. Suppose $\sigma = (1, 1, q_\lambda^*, \lambda)$ is a cooperative Nash equilibrium for given values of n, δ and ε. We distinguish two cases:

(i) $\delta \geq 1 - 2\varepsilon$. It follows from the proof of Theorem 2 that also $GTFT = (1, 1, q_0^*, 0)$ is a Nash equilibrium for the given parameter values. Moreover, because the threshold value δ_0 in Eq. (5.28) is independent of n and ε, it follows that $GTFT$ is also a cooperative Nash equilibrium for all $n' \geq n$, $\delta' \geq \delta$, and $\varepsilon' \leq \varepsilon$.

(i) $\delta < 1 - 2\varepsilon$. In this case it follows that $GSCO = (1, 1, q_1^*, 1)$ is also a Nash equilibrium for the given parameter values. Because the threshold δ_1 is monotonically increasing in ε and monotonically decreasing in n (for $\delta < 1 - 2\varepsilon$), $GSCO$ is also a cooperative Nash equilibrium for all $n' \geq n$, $\delta' \geq \delta$, and $\varepsilon' \leq \varepsilon$.

□

Proof of Theorem 3. In the limit $\varepsilon \to 0$, it follows from Theorem 2 that there is a cooperative Nash equilibrium in reactive strategies if and only if

$$\delta \geq \delta_1 = \frac{c}{(n-1)b - (n-2)c}. \tag{5.84}$$

Now suppose to the contrary there is some arbitrary other strategy σ that is a cooperative Nash equilibrium for some $\delta < \delta_1$. In particular, players must not have an incentive to deviate by playing *ALLD*. Because players in a cooperative equilibrium need to cooperate against unknown co-players, a player who deviates to *ALLD* obtains at least the payoff of b in the first interaction he participates in. Thus, *ALLD*'s expected deviation payoff from σ satisfies

$$\pi_D \geq (1-d)b \cdot \Big(1 + (1-\bar{w})d + (1-\bar{w})^2 d^2 + \dots\Big) = \frac{(1-d)b}{1-(1-\bar{w})d}. \tag{5.85}$$

For σ to be a Nash equilibrium it thus needs to be the case that $\pi_D \leq b - c$, which implies

$$d \geq \frac{c}{\bar{w}b + (1-\bar{w})c}. \tag{5.86}$$

Using the identities $\bar{w} = 2/n$ and $d = n(n-1)\delta / \big(2 + (n-2)(n+1)\delta\big)$, this inequality simplifies to $\delta \geq \delta_1$, with δ_1 as defined in Eq. (5.84). □

Proof of Theorem 4. The first part of the proof of Theorem 4 is analogous to the proof of Lemma 3 and Theorem 1, and is therefore omitted here. The second part follows from the requirement that the conditional cooperation probability q needs to satisfy $q \geq 0$ for $\lambda = 0$ and $\lambda = 1$, respectively. □

Proof of Theorem 5. If implementation errors occur with a probability e, a player with strategy $\sigma = (y, p, q, \lambda)$ effectively employs the strategy $\hat{\sigma} = (\hat{y}, \hat{p}, \hat{q}, \hat{\lambda})$ where

$$\hat{y} = (1-e)y + e(1-y), \quad \hat{p} = (1-e)p + e(1-p), \quad \hat{q} = (1-e)q + e(1-q), \quad \hat{\lambda} = \lambda. \tag{5.87}$$

For the strategy σ to be a generic Nash equilibrium, it now needs to be the case that

$$\hat{r} := \hat{p} - \hat{q} = r^*_\lambda, \tag{5.88}$$

where r^*_λ is as defined in the model without implementation errors, Eq. (5.23). In addition, for a homogeneous σ-population to be fully cooperative in the absence of errors, we require

$$y = 1 \quad \text{and} \quad p = 1. \tag{5.89}$$

Jointly solving Eqs. (5.87)–(5.89) for q yields

$$q = 1 - \frac{1 + (n-2)\delta\lambda}{(1-2e)\Big(1 + (n-2)(1-2\varepsilon)\lambda\Big)} \frac{c}{\delta b}. \tag{5.90}$$

This proves the first part. The second part again follows from $q \geq 0$ for both $\lambda = 0$ and $\lambda = 1$. □

Proof of Theorem 6. Again, the first part of the proof of Theorem 6 is analogous to the proof of Lemma 3 and Theorem 1; the second part follows from requiring $q \geq 0$ for $\lambda = 0$ and $\lambda = 1$, respectively. □

Proof of Theorem 7. The first part follows immediately from Lemma 3, by replacing λ by the effective probability $\lambda\nu$ to take indirect information into account. The second part again is a consequence of the requirement that $q \geq 0$ for both $\lambda = 0$ and $\lambda = 1$. □

Proof of Theorem 8. As with the previous results, the first part of the proof of Theorem 8 is analogous to the proof of Lemma 3. For example, Eq. (5.77) needs to be replaced by

$$\Big(1 - d + wd(1-\kappa) + d(\bar{w} - w)\kappa\Big) x_{in} - dw(1-\kappa) r x_{ni} - dwr(1-2\varepsilon)\kappa \sum_{l \neq i,n} x_{nl} = (1-d)y + d\Big(w(1-\kappa)q + \kappa(\bar{w} - w)(q + \varepsilon r)\Big). \tag{5.91}$$

As a result, the analogous quantities to A_1, A_2, and r^*_λ in Lemma 3 are given by

$$
\begin{aligned}
A_1 &= \frac{(1-\delta)y+\delta\Big((1-\kappa)q+(n-2)\kappa(q+r\varepsilon)\Big)}{1+(n-3)\delta\kappa}\cdot b\\
A_2 &= \frac{1-\kappa+\kappa(n-2)(1-2\varepsilon)}{1+(n-3)\delta\kappa}\cdot \delta b\\
r^*_\kappa &= \frac{1+(n-3)\delta\kappa}{1-\kappa+\kappa(n-2)(1-2\varepsilon)}\frac{c}{\delta b}.
\end{aligned}
\tag{5.92}
$$

The second part of the Theorem then follows from requiring $q{:=}1{-}r^*_\kappa \geq 0$. □

CHAPTER 6

Conclusions and future perspectives

The contributions to the field of indirect reciprocity presented in this thesis are quite comprehensive. First, we have demonstrated a fundamental shortcoming of the well-known leading eight social norms: while they are highly successful under fully public information, we have shown that they break down under private and imperfect information. Disagreements introduced by small amounts of noise can proliferate and prevent cooperation from emerging when players do not synchronize their opinions about others. This result has given way to two further new findings: on one hand, complex norms of indirect reciprocity such as the leading eight cannot be made more robust against noise by introducing stochastic "generosity", unlike simple strategies of direct reciprocity such as Tit-for-Tat. On the other hand, we have shown that it is possible to sustain cooperation in imperfect information settings with higher order norms by letting players have more nuanced opinions of each other. When reputations can take values beyond "good" and "bad", the effects of noise and incomplete observations is reduced, effectively slowing down or even preventing the proliferation of disagreements.

Finally, we have presented a unified mathematical framework of direct and indirect reciprocity, where we find a new, simple and successful strategy for indirect reciprocity with a natural analogue in direct reciprocity. We have analyzed the equilibria of this model and have characterized the environments in which either of the two modes of reciprocity can sustain cooperation. Additionally, we have explored the evolvability of the two modes when individuals engage in social learning to adapt to their environment. Our study has indicated that in order to sustain cooperation via indirect reciprocity, also under imperfect information, complex social norms are not always required and simple rules can already suffice in contrast.

The results presented in this thesis still naturally open up various avenues for future work. Here, we discuss two open sets of questions which can serve as starting points for further exciting research into indirect reciprocity.

6.1 Indirect reciprocity on networks

A main assumption usually made in previous literature on indirect reciprocity is modeling the population of individuals as well mixed. This means that the underlying topology is a complete graph: everyone is connected to everyone else, and thereby has the potential to observe everyone else as well as interact with everyone else. This is a highly idealized setting, and in a more realistic model, we would need to assume instead that individuals are connected via a less dense underlying topology. Constant-fitness evolutionary dynamics on graphs have already been a topic of interest [LHN05]. Also, direct reciprocity has already been studied in this setting, by modeling players as nodes on networks where only neighbors can interact in social dilemmas (for example public good games) [AN14, SF07, ON07]. Future research on indirect reciprocity with underlying network structure thus faces a rich set of problems that are very relevant for our understanding of human society and cooperation.

First, the most fundamental questions should be answered. How does the success of indirect reciprocity as we know it depend on the underlying topology of the network? Can a topology other than the complete graph sometimes enhance cooperation via indirect reciprocity (especially in noisy settings), or is more structure in a population always detrimental? Can we identify common features that make a topology very suitable or very unsuitable for the evolution of cooperation via indirect reciprocity? The work presented in this thesis already found that incomplete observation can hinder the evolution of cooperation in well-mixed populations - does additional structure mitigate this problem, or worsen it? Answering these questions can ultimately give insight into how sophisticated strategies fare in the realistic interplay between population structure and imperfect information.

On the other hand, there is the possibility to take a different approach by using the unified framework of direct and indirect reciprocity (Chapter 5). Importantly, the main equation to calculate payoffs that we use in this work is a very general one, and can easily be used to transfer and extend this framework to a more complicated topology given by an adjacency matrix. Preliminary work has shown that with this master equation, we can explicitly calculate payoffs of players on an arbitrary network in this model. This implies that we can further use it to explore the stability of simple indirect reciprocity strategies in a more sophisticated setting, consider how the two modes of reciprocity interact and evolve in a structured population, and thus find out how cooperation can spread on networks.

Future research can also take on more extensive questions related to the interplay between network structure and social behavior. One crucial aspect of social networks is the role that communication plays. Some preliminary results have shown that in well-mixed populations, gossip can have a mediating function that enables individuals to deal with unreliable observations, effectively reducing error rates and boosting cooperation. Analyzing gossip in structured populations, and investigating the effect of e.g. liars, who misrepresent information, can be a first step in understanding how communication can help or hinder indirect reciprocity. Another interesting aspect is the effect of inherent network properties that influence the propagation of information itself. That is, one can assume that not only are observations and interactions restricted by a topology, but that the information flow is subject to certain constraints too. Thus, even in a model where all players are faithfully informed about an interaction, propagation delays can lead to different reputation dynamics and disagreement, and remove the (somewhat unnatural) synchronicity assumption made in earlier models. Finding properties of topologies that can cope well with such noise can be an interesting question to explore in this context. Finally, in the most complex version of a model of indirect reciprocity on networks, one could analyze what happens when these networks are dynamic, and links between individuals are not static anymore. This is a very powerful model, and has already been of interest in other areas. In this setting, players can make and break links over time, based on their assessment of others, leading to a co-evolution of behavior and topology. This analysis would give an exciting insight into how social ties in communities evolve.

6.2 Zero-determinant strategies for indirect reciprocity

Players using zero-determinant strategies of direct reciprocity [PD12, HCN18, HRZ15] can enforce a linear relationship between their own and their opponent's payoff in a repeated game by choice of only their own strategy parameters. Thus, zero-determinant players are able to unilaterally set the opponent's expected payoff in an iterated game. This class of strategies includes the so-called "extortionate" strategies, which can set the payoff of the player using them to always be larger than the payoff of their co-player, while forcing their opponent to cooperate in every round. An extortioner can in fact teach their co-player to cooperate if the co-player occasionally changes their strategy in response to the extortioner's fixed behavior. When the co-player tries to increase their own payoff, this also automatically increases the extortioner's payoff. As a result, the co-player becomes more cooperative over time, and the extortioner's payoff increases by more. These strategies have been investigated in various studies [HNS13, SP14a, SP14c, MH16b]. Another special case of zero-determinant strategies are the so-called equalizer strategies [BNS97, HTS15], which give the opponent the exact same payoff, no matter which strategy of arbitrary complexity they use. These equalizer strategies are of particular interest in the context of this thesis, as they are a fundamental part of our theoretical results on a unified framework of direct and indirect reciprocity in Chapter 5. In fact, by extending the theory of zero-determinant strategies, we show that cooperative Nash equilibria of this unified framework take the form of such equalizer strategies. They are equilibria against all possible mutants, no matter the complexity of their strategy. This finding serves to show the relevance of the class of zero-determinant strategies for indirect reciprocity additionally to direct reciprocity. From this, the natural question then follows if we can in fact not only find equalizers, but the entire class of zero-determinant strategies in the realm of indirect reciprocity, and which form they take. Do extortioners in indirect reciprocity exist, and how do they relate to their direct reciprocity equivalents?

Bibliography

[AH81] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211:1390–1396, 1981.

[Aki16] E. Akin. The iterated prisoner's dilemma: Good strategies and their dynamics. In I. Assani, editor, *Ergodic Theory, Advances in Dynamics*, pages 77–107. de Gruyter, Berlin, 2016.

[Ale87] R. Alexander. *The Biology of Moral Systems*. Aldine de Gruyter, New York, 1987.

[AN14] Benjamin Allen and Martin A Nowak. Games on graphs. *EMS Surveys in Mathematical Sciences*, 1(1):113–151, 2014.

[AT09] P. M. Altrock and A. Traulsen. Fixation times in evolutionary games under weak selection. *New Journal of Physics*, 11:013012, 2009.

[BDF15] Aaron T Beck, Denise D Davis, and Arthur Freeman. *Cognitive therapy of personality disorders*. Guilford Publications, 2015.

[Ber11] U. Berger. Learning to cooperate via indirect reciprocity. *Games and Economic Behavior*, 72:30–37, 2011.

[BG16] U. Berger and A. Grüne. On the stability of cooperation under indirect reciprocity with first-order information. *Games and Economic Behavior*, 98:19–33, 2016.

[BJHN16] S. K. Baek, H. C. Jeong, C. Hilbe, and M. A. Nowak. Comparing reactive and memory-one strategies of direct reciprocity. *Scientific Reports*, 6:25676, 2016.

[BKO05] G.E. Bolton, E. Katok, and A. Ockenfels. Cooperation among strangers with limited information about reputation. *Journal of Public Economics*, 89:1457–1468, 2005.

[BL87] R. Boyd and J.M.D. Lorberbaum. No pure strategy is evolutionary stable in the iterated prisoner's dilemma game. *Nature*, 327:58–59, 1987.

[Blu93] L. E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.

[BNS97] M. C. Boerlijst, M. A. Nowak, and K. Sigmund. Equal pay for all prisoners. *American Mathematical Monthly*, 104:303–307, 1997.

[Boy89] R. Boyd. Mistakes allow evolutionary stability in the repeated Prisoner's Dilemma game. *Journal of Theoretical Biology*, 136:47–56, 1989.

[BS95] J. Bendor and P. Swistak. Types of evolutionary stability and the problem of cooperation. *Proceedings of the National Academy of Sciences USA*, 92:3596–3600, 1995.

[BS04] H. Brandt and K. Sigmund. The logic of reprobation: Assessment and action rules for indirect reciprocation. *Journal of Theoretical Biology*, 231:475–486, 2004.

[BS05] H. Brandt and K. Sigmund. Indirect reciprocity, image scoring, and moral hazard. *Proceedings of the National Academy of Sciences USA*, 102:2666–2670, 2005.

[BS06] H. Brandt and K. Sigmund. The good, the bad and the discriminator - errors in direct and indirect reciprocity. *Journal of Theoretical Biology*, 239:183–194, 2006.

[CFW19] D. Clark, D. Fudenberg, and A. Wolitzky. Individualistic records robustly support social cooperation. *Working paper*, 2019.

[CFW20] Daniel Clark, Drew Fudenberg, and Alexander Wolitzky. Indirect reciprocity with simple records. *Proceedings of the National Academy of Sciences*, 117(21):11344–11349, 2020.

[CGLF+15] Jose A Cuesta, Carlos Gracia-Lázaro, Alfredo Ferrer, Yamir Moreno, and Angel Sánchez. Reputation drives cooperative behaviour and network formation in human groups. *Scientific reports*, 5(1):1–6, 2015.

[CGVP16] V. Capraro, F. Giardini, D. Vilone, and M. Paolucci. Partner selection supported by opaque reputation promotes cooperative behavior. *Judgment and Decision Making*, 11(6):589–600, 2016.

[Cor09] T. H. Cormen. *Introduction to algorithms.* MIT Press, Cambridge, MA, 2009.

[CSP06] F. A. C. C. Chalub, F. C. Santos, and J. M. Pacheco. The evolution of norms. *Journal of Theoretical Biology*, 241:233–240, 2006.

[CT08] J. C. Claussen and A. Traulsen. Cyclic dominance and biodiversity in well-mixed populations. *Physical Review Letters*, 100:058104, 2008.

[CZN12] K. Chatterjee, D. Zufferey, and M. A. Nowak. Evolutionary game dynamics in populations with different learners. *Journal of Theoretical Biology*, 301:161–173, 2012.

[DLTZ15] Y. Dong, C. Li, Y. Tao, and B. Zhang. Evolution of conformity in social dilemmas. *PLoS One*, 10(9):e0137435, 2015.

[EF09] D. Engelmann and U. Fischbacher. Indirect reciprocity and strategic reputation building in an experimental helping game. *Games and Economic Behavior*, 67:399–407, 2009.

[Feh04] Ernst Fehr. Don't lose your reputation. *Nature*, 432:449–450, Nov 2004.

[FFG+13] I. Fischer, A. Frid, S. J. Goerg, S. A. Levin, D. I. Rubenstein, and R. Selten. Fusing enacted and expected mimicry generates a winning strategy that promotes the evolution of cooperation. *PNAS*, 110:10229–10233, 2013.

[FGF01] U. Fischbacher, S. Gächter, and E. Fehr. Are people conditionally cooperative? Evidence from a public goods experiment. *Economic Letters*, 71:397–404, Apr 2001.

[FI06] D. Fudenberg and L. A. Imhof. Imitation processes with small mutations. *Journal of Economic Theory*, 131:251–262, 2006.

[FM90] D. Fudenberg and E. Maskin. Evolution and cooperation in noisy repeated games. *American Economic Review*, 80:274–279, May 1990.

[FNTI06] D. Fudenberg, M. A. Nowak, C. Taylor, and L. A. Imhof. Evolutionary game dynamics in finite populations with strong selection and weak mutation. *Theoretical Population Biology*, 70:352–363, 2006.

[FT98] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, Cambridge, 6th edition, 1998.

[FW89] J. Farrell and R. Ware. Evolutionary stability in the repeated prisoner's dilemma. *Theoretical Population Biology*, 36:161–166, 1989.

[GGLM+14] J. Grujic, 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*, 4:4615, 2014.

[GK21] N.E. Glynatsi and V.A. Knight. A bibliometric study of research topics, collaboration and centrality in the field of the Iterated Prisoner's Dilemma. *Humanities and Social Sciences Communications*, 8:45, 2021.

[GT12a] J. García and A. Traulsen. Leaving the loners alone: Evolution of cooperation in the presence of antisocial punishment. *Journal of Theoretical Biology*, 307:168–173, 2012.

[GT12b] J. García and A. Traulsen. The structure of mutations and the evolution of cooperation. *PLoS One*, 7:e35287, 2012.

[GvV16] J. García and M. van Veelen. In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory*, 161:161–189, 2016.

[GvV18] 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*, 5:102, 2018.

[HCN18] C. Hilbe, K. Chatterjee, and M. A. Nowak. Partners and rivals in direct reciprocity. *Nature Human Behaviour*, 2(7):469–477, 2018.

[HHCN19] O.P. Hauser, C. Hilbe, K. Chatterjee, and M.A. Nowak. Social dilemmas among unequals. *Nature*, 572:524—527, 2019.

[HI12] C. Hauert and L. Imhof. Evolutionary games in deme structured, finite populations. *Journal of Theoretical Biology*, 299:106–112, 2012.

[Hil11] C. Hilbe. Local replicator dynamics: A simple link between deterministic and stochastic models of evolutionary game theory. *Bulletin of Mathematical Biology*, 73:2068–2087, 2011.

[HMPS11] The Anh Han, Luís Moniz Pereira, and Francisco C Santos. Intention recognition promotes the emergence of cooperation. *Adaptive Behavior*, 19(4):264–279, 2011.

[HMVCN17] 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 USA*, 114:4715–4720, 2017.

[HNS13] C. Hilbe, M. A. Nowak, and K. Sigmund. The evolution of extortion in iterated prisoner's dilemma games. *Proceedings of the National Academy of Sciences USA*, 110:6913–6918, 2013.

[HNT13] C. Hilbe, M. A. Nowak, and A. Traulsen. Adaptive dynamics of exortion and compliance. *PLoS One*, 8:e77886, 2013.

[HRZ15] D. Hao, Z. Rong, and T. Zhou. Extortion under uncertainty: Zero-determinant strategies in noisy games. *Physical Review E*, 91(5):052803, 2015.

[HS97] 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 B*, 264:513–519, 1997.

[HS98] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK, 1998.

[HST+18] C. Hilbe, L. Schmid, J. Tkadlec, K. Chatterjee, and M. A. Nowak. Indirect reciprocity with private, noisy, and incomplete information. *Proceedings of the National Academy of Sciences USA*, 115:12241–12246, 2018.

[HT12] C. Hilbe and A. Traulsen. Emergence of responsible sanctions without second order free riders, antisocial punishment or spite. *Scientific Reports*, 2:458, 2012.

[HTB+07] C. Hauert, A. Traulsen, H. Brandt, M. A. Nowak, and K. Sigmund. Via freedom to coercion: the emergence of costly punishment. *Science*, 316:1905–1907, 2007.

[HTS15] C. Hilbe, A. Traulsen, and K. Sigmund. Partners or rivals? Strategies for the iterated prisoner's dilemma. *Games and Economic Behavior*, 92:41–52, 2015.

[HvCN18] C. Hilbe, S. Šimsa, K. Chatterjee, and M. A. Nowak. Evolution of cooperation in stochastic games. *Nature*, 559:246–249, 2018.

[HWTN14] C. Hilbe, B. Wu, A. Traulsen, and M. A. Nowak. Cooperation and control in multiplayer social dilemmas. *Proceedings of the National Academy of Sciences USA*, 111:16425–16430, 2014.

[IFN05] L. A. Imhof, D. Fudenberg, and M. A. Nowak. Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences USA*, 102:10797–10800, 2005.

[IFN07] L. A. Imhof, D. Fudenberg, and M. A. Nowak. Tit-for-tat or win-stay, lose-shift? *Journal of Theoretical Biology*, 247(3):574–580, Aug 2007.

[IM18] G. Ichinose and N. Masuda. Zero-determinant strategies in finitely repeated games. *Journal of Theoretical Biology*, 438:61–77, 2018.

[IN10] L. A. Imhof and M. A. Nowak. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B*, 277:463–468, 2010.

[JHTM11] J. Jacquet, C. Hauert, A. Traulsen, and M. Milinski. Shame and honour drive cooperation. *Biology Letters*, 7:899–901, 2011.

[JOS+18] Peter K. Jonason, Atsushi Oshio, Tadahiro Shimotsukasa, Takahiro Mieda, ÃĄrpÃąd CsathÃş, and Maria Sitnikova. Seeing the world in black or white: The dark triad traits and dichotomous thinking. *Personality and Individual Differences*, 120:102–106, 2018.

[Kan92] M. Kandori. Social norms and community enforcement. *Review of Economic Studies*, 59(1):63–80, 1992.

[KH21] Marcus Krellner and The Anh Han. Pleasing enhances indirect reciprocity under private assessment. *Artificial Life*, 2021.

[KK89] D. P. Kraines and V. Y. Kraines. Pavlov and the prisoner's dilemma. *Theory and Decision*, 26:47–79, 1989.

[KT75] S. Karlin and H. M. A. Taylor. *A First Course in Stochastic Processes*. Academic, London, 2nd edition edition, 1975.

[LFS99] A. Lotem, M. A. Fishman, and L. Stone. Evolution of cooperation between individuals. *Nature*, 400:226–227, 1999.

[LH01] O. Leimar and P. Hammerstein. Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society B*, 268:745–753, 2001.

[LHF15] C. Lee, M. Harper, and D. Fryer. The art of war: Beyond memory-one strategies in population games. *PLoS One*, 10(3):e0120625, 2015.

[LHN05] E. Lieberman, C. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433:312–316, 2005.

[LIDS19] Y. Lee, Y. Iwasa, U. Dieckmann, and K. Sigmund. Social evolution leads to persistent corruption. *Proceedings of the National Academy of Sciences USA*, 116(27):13276–13281, 2019.

[Lin97] K. Lindgren. Evolutionary dynamics in game-theoretic models. In W. B. Arthur, Steven N. Durlauf, and David A. Lane, editors, *The Economy as an Evolving Complex System II*, pages 337–368. Addison-Wesley, Reading MA, 1997.

[LJS12] Howard G Lavine, Christopher D Johnston, and Marco R Steenbergen. *The ambivalent partisan: How critical loyalty promotes democracy*. Oxford University Press, 2012.

[LLXH15] J. Liu, Y. Li, C. Xu, and P. M. Hui. Evolutionary behavior of generalized zero-determinant strategies in iterated prisoner's dilemma. *Physica A*, 430:81–92, 2015.

[LMB21] Sanghun Lee, Yohsuke Murase, and Seung Ki Baek. Local stability of cooperation in a continuous model of indirect reciprocity. *Scientific Reports*, 11, 2021.

[LZS+04] Klaus Lieb, Mary C Zanarini, Christian Schmahl, Marsha M Linehan, and Martin Bohus. Borderline personality disorder. *The Lancet*, 364(9432):453–461, 2004.

[McA15] A. McAvoy. Comment on "Imitation processes with small mutations". *J. Econ. Theory*, 159:66–69, 2015.

[MH16a] A. McAvoy and C. Hauert. Autocratic strategies for alternating games. *Theoretical Population Biology*, 113:13–22, 2016.

[MH16b] A. McAvoy and C. Hauert. Autocratic strategies for iterated games with arbitrary action spaces. *Proceedings of the National Academy of Sciences*, 113(13):3573–3578, 2016.

[Mol85] P. Molander. The optimal level of generosity in a selfish, uncertain environment. *Journal of Conflict Resolution*, 29:611–618, 1985.

[MS82] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, 1982.

[MS06] G. J. Mailath and L. Samuelson. *Repeated games and reputations*. Oxford Univ. Press, Oxford, UK, 2006.

[MS10] A. P. Melis and D. Semmann. How is human cooperation different? *Philosophical Transactions of the Royal Society B*, 365:2663–2674, 2010.

[MSBK01] M. Milinski, D. Semmann, T. C. Bakker, and H.-J. Krambeck. Cooperation through indirect reciprocity: image scoring or standing strategy? *Proceedings of the Royal Society B*, 268:2495–2501, 2001.

[MSK02] M. Milinski, D. Semmann, and H. J. Krambeck. Reputation helps solve the 'tragedy of the commons'. *Nature*, 415:424–426, 2002.

[MSP73] J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246:15–18, 1973.

[MVC13] L. A. Martinez-Vaquero and J. A. Cuesta. Evolutionary stability and resistance to cheating in an indirect reciprocity model based on reputation. *Physical Review E*, 87:052810, 2013.

[MVCS12] L. A. Martinez-Vaquero, J. A. Cuesta, and A. Sanchez. Generosity pays in the presence of direct reciprocity: A comprehensive study of 2x2 repeated games. *PLoS ONE*, 7(4):E35135, 2012.

[MvdBE13] L. Molleman, E. van den Broek, and M. Egas. Personal experience and reputation interact in human decisions to help reciprocally. *Proceedings of the Royal Society B*, 28:20123044, 2013.

[MVHPL15] Luis A Martinez-Vaquero, The Anh Han, Luís Moniz Pereira, and Tom Lenaerts. Apology and forgiveness evolve to resolve failures in cooperative agreements. *Scientific reports*, 5:10639, 2015.

[NK04] Mayuko Nakamaru and Masakado Kawata. Evolution of rumours that discriminate lying defectors. *Evolutionary Ecology Research*, 6:261–283, 2004.

[NM07] Lisa A Napolitano and Dean McKay. Dichotomous thinking in borderline personality disorder. *Cognitive Therapy and Research*, 31(6):717–726, 2007.

[NM11] M. Nakamura and N. Masuda. Indirect reciprocity under incomplete observation. *PLoS Computational Biology*, 7:e1002113, 2011.

[Now06] M. A. Nowak. Five rules for the evolution of cooperation. *Science*, 314:1560–1563, 2006.

[NPSH15] H. H. Nax, M. Perc, A. Szolnoki, and D. Helbing. Stability of cooperation under image scoring in group interactions. *Scientific Reports*, 5, 2015.

[NRTV07] N. Nissan, T. Roughgarden, E. Tardos, and V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007.

[NS90] M. A. Nowak and K. Sigmund. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematicae*, 20:247–265, 1990.

[NS92] M. A. Nowak and K. Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355:250–253, 1992.

[NS98a] M. A. Nowak and K. Sigmund. The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, 194:561–574, 1998.

[NS98b] M. A. Nowak and K. Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393:573–577, 1998.

[NS05] M. A. Nowak and K. Sigmund. Evolution of indirect reciprocity. *Nature*, 437:1291–1298, 2005.

[NSES95] M. A. Nowak, K. Sigmund, and E. El-Sedy. Automata, repeated games and noise. *Journal of Mathematical Biology*, 33:703–722, 1995.

[NSTF04] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428:646–650, 2004.

[OFP95] M. Okuno-Fujiwara and A. Postlewaite. Social norms and random matching games. *Games and Economic Behavior*, 9:79–109, Sep 1995.

[Oht04] H. Ohtsuki. Reactive strategies in indirect reciprocity. *Journal of Theoretical Biology*, 227:299–314, 2004.

[OI04] H. Ohtsuki and Y. Iwasa. How should we define goodness? – Reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology*, 231:107–20, 2004.

[OI06] H. Ohtsuki and Y. Iwasa. The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, 239:435–444, 2006.

[OIN09] Hisashi Ohtsuki, Yoh Iwasa, and Martin A. Nowak. Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature*, 457:79–82, Feb 2009.

[OIN15] H. Ohtsuki, Y. Iwasa, and M. A. Nowak. Reputation effects in public and private interactions. *PLoS Computational Biology*, 11(11):e1004527, 2015.

[Oka20] I. Okada. A review of theoretical studies on indirect reciprocity. *Games*, 11(3):27, 2020.

[ON07] Hisashi Ohtsuki and Martin A Nowak. Direct reciprocity on graphs. *Journal of Theoretical Biology*, 247:462–470, 2007.

[ON08] H. Ohtsuki and M. A. Nowak. Evolutionary stability on graphs. *Journal of Theoretical Biology*, 251:698–707, 2008.

[Osh09] Atsushi Oshio. Development and validation of the dichotomous thinking inventory. *Social Behavior and Personality: an international journal*, 37(6):729–741, 2009.

[Osh12] Atsushi Oshio. An all-or-nothing thinking turns into darkness: Relations between dichotomous thinking and personality disorders1. *Japanese Psychological Research*, 54(4):424–429, 2012.

[OSI13] K. Oishi, T. Shimada, and N. Ito. Group formation through indirect reciprocity. *Physical Review E*, 87:030801, 2013.

[OSN17] I. Okada, T. Sasaki, and Y. Nakai. Tolerant indirect reciprocity can boost social welfare through solidarity with unconditional cooperators in private monitoring. *Scientific Reports*, 7:9737, 2017.

[OSN18] I. Okada, T. Sasaki, and Y. Nakai. A solution for private assessment in indirect reciprocity using solitary observation. *Journal of Theoretical Biology*, 455:7–15, 2018.

[OYS+18] I. Okada, H. Yamamoto, Y. Sato, S. Uchida, and T. Sasaki. Experimental evidence of selective inattention in reputation-based cooperation. *Scientific Reports*, 8:14813, 2018.

[OYU20] Isamu Okada, Hitoshi Yamamoto, and Satoshi Uchida. Hybrid assessment scheme based on the stern- judging rule for maintaining cooperation under indirect reciprocity. *Games*, 11(1), 2020.

[PB03] K. Panchanathan and R. Boyd. A tale of two defectors: the importance of standing for evolution of indirect reciprocity. *Journal of Theoretical Biology*, 224:115–126, 2003.

[PD92] G. Pollock and L. A. Dugatkin. Reciprocity and the emergence of reputation. *Journal of Theoretical Biology*, 159:25–37, 1992.

[PD12] W. H. Press and F. D. Dyson. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *PNAS*, 109:10409–10413, 2012.

[PHRZ15] L. Pan, D. Hao, Z. Rong, and T. Zhou. Zero-determinant strategies in iterated public goods game. *Scientific Reports*, 5:13096, 2015.

[Pos99] M. Posch. Win-stay, lose-shift strategies for repeated games - memory length, aspiration levels and noise. *Journal of Theoretical Biology*, 198:183–195, 1999.

[PSC06] J. M. Pacheco, F. C. Santos, and F. A. C. C. Chalub. Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Computational Biology*, 2:1634–1638, 2006.

[PVSP14] 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 Comput Biol*, 10(11):e1003945, 2014.

[RHR+18] 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*, 9:555, 2018.

[RN12] D. G. Rand and M. A. Nowak. Human cooperation. *Trends in Cogn. Sciences*, 117:413–425, 2012.

[Rob07] G. Roberts. Evolution of direct and indirect reciprocity. *Proceedings of the Royal Society B*, 275(1631):173–179, Oct 2007.

[RSC15] 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*, 10(7):e0134128, 2015.

[RSP19] A. L Radzvilavicius, A. J. Stewart, and J. B. Plotkin. Evolution of empathetic moral evaluation. *eLife*, 8:e44269, 2019.

[RvdR12] M. Restivo and A. van de Rijt. Experimental study of informal rewards in peer production. *PLoS One*, 7(3):e34358, 2012.

[RW90] W. Raub and J. Weesie. Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology*, 96(3):626–654, 1990.

[RZSL06] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on eBay: A controlled experiment. *Experimental Economics*, 9:79–101, 2006.

[SA07] S Suzuki and Eizo Akiyama. Evolution of indirect reciprocity in groups of various sizes and comparison with direct reciprocity. *Journal of Theoretical Biology*, 245(3):539–552, Apr 2007.

[SATJ82] MD SALMAN AKHTAR and J Anderson Thomson Jr. Overview: Narcissistic personality disorder. *Am J Psychiatry*, 139(1), 1982.

[SCHN21] L. Schmid, K. Chatterjee, C. Hilbe, and M.A. Nowak. A unified framework of direct and indirect reciprocity. *Nature Human Behaviour*, 2021.

[SDSTH10] K. Sigmund, H. De Silva, A. Traulsen, and C. Hauert. Social learning promotes institutions for governing the commons. *Nature*, 466:861–863, 2010.

[Sel75] R. Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55, 1975.

[SF07] G. Szabó and G. Fáth. Evolutionary games on graphs. *Physics Reports*, 446:97–216, 2007.

[Sig10] K. Sigmund. *The Calculus of Selfishness*. Princeton Univ. Press, Princeton, NJ, 2010.

[Sig12] K. Sigmund. Moral assessment in indirect reciprocity. *Journal of Theoretical Biology*, 299:25–30, 2012.

[SKM04] Dirk Semmann, Hans-Jürgen Krambeck, and Manfred Milinski. Strategic investment in reputation. *Behavioral Ecology and Sociobiology*, 56(3):5, Jul 2004.

[SKSM07] R. D. Sommerfeld, H.-J. Krambeck, D. Semmann, and M. Milinski. Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences USA*, 104(44):17435–17440, 2007.

[SMJ+14] A. Szolnoki, M. Mobilia, L. Jiang, B. Szczesny, A. M. Rucklidge, and M. Perc. Cyclic dominance in evolutionary games: A review. *Journal of The Royal Society Interface*, 11:20140735, 2014.

[SMUE16] V. Swakman, L. Molleman, A. Ule, and M. Egas. Reputation-based cooperation: Empirical evidence for behavioral strategies. *Evolution and Human Behavior*, 37:230–235, 2016.

[SN16] Motohide Seki and Mayuko Nakamaru. A model for gossip-mediated evolution of altruism with various types of false information by speakers and assessment by listeners. *Journal of Theoretical Biology*, 407:90–105, 2016.

[SON17] T. Sasaki, I. Okada, and Y. Nakai. The evolution of conditional moral assessment in indirect reciprocity. *Scientific Reports*, 7:41870, 2017.

[SP12] A. J. Stewart and J. B. Plotkin. Extortion and cooperation in the prisoner's dilemma. *Proceedings of the National Academy of Sciences USA*, 109:10134–10135, 2012.

[SP13] A. J. Stewart and J. B. Plotkin. From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proceedings of the National Academy of Sciences USA*, 110(38):15348–15353, 2013.

[SP14a] A. J. Stewart and J. B. Plotkin. Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences USA*, 111(49):17558 – 17563, 2014.

[SP14b] A. Szolnoki and M. Perc. Defection and extortion as unexpected catalysts of unconditional cooperation in structured populations. *Scientific Reports*, 4:5496, 2014.

[SP14c] A. Szolnoki and M. Perc. Evolution of extortion in structured populations. *Physical Review E*, 89:022804, 2014.

[SP15] A. J. Stewart and J. B. Plotkin. The evolvability of cooperation under local and non-local mutations. *Games*, 6(3):231–250, 2015.

[SP16] A. J. Stewart and J. B. Plotkin. Small groups and long memories promote cooperation. *Scientific Reports*, 6:26889, 2016.

[SPP16] A. J. Stewart, T. L. Parsons, and J. B. Plotkin. Evolutionary consequences of behavioral diversity. *Proceedings of the National Academy of Sciences USA*, 113(45):E7003–E7009, 2016.

[SS04] A. Szolnoki and G. Szabó. Phase transitions for rock-scissors-paper game on different networks. *Physical Review E*, 70:037102, 2004.

[SS06] I. Seinen and A. Schram. Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review*, 50:581–602, 2006.

[SSHC21] Laura Schmid, Pouya Shati, Christian Hilbe, and Krishnendu Chatterjee. Evolution of indirect reciprocity under assessment and action generosity. *Scientific Reports*, 11, 2021.

[SSP16] F. P. Santos, F. C. Santos, and J. M. Pacheco. Social norms of cooperation in small-scale societies. *PLoS Computational Biology*, 12(1):e1004709, 2016.

[SSP18] F. P. Santos, F. C. Santos, and J. M. Pacheco. Social norm complexity and past reputations in the evolution of cooperation. *Nature*, 555:242–245, 2018.

[ST98] G. Szabó and C. Tőke. Evolutionary Prisoner's Dilemma game on a square lattice. *Physical Review E*, 58:69–73, 1998.

[Sug86] Robert Sugden. *The economics of rights, co-operation and welfare*. Blackwell, Oxford, UK, 1986.

[TH09] A. Traulsen and C. Hauert. Stochastic evolutionary game dynamics. In Heinz Georg Schuster, editor, *Reviews of Nonlinear Dynamics and Complexity*, pages 25–61. Wiley-VCH, Weinheim, 2009.

[THDS+09] A. Traulsen, C. Hauert, H. De Silva, M. A Nowak, and K. Sigmund. Exploration dynamics in evolutionary games. *Proceedings of the National Academy of Sciences USA*, 106:709–712, 2009.

[TNP06] A. Traulsen, M. A. Nowak, and J. M. Pacheco. Stochastic dynamics of invasion and fixation. *Physical Review E*, 74:011909, 2006.

[TPN07] A. Traulsen, J. M. Pacheco, and M. A. Nowak. Pairwise comparison and selection temperature in evolutionary game dynamics. *Journal of Theoretical Biology*, 246:522–529, 2007.

[Tri71] R. L. Trivers. The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46:35–57, 1971.

[TSM13] S. Tanabe, H. Suzuki, and N. Masuda. Indirect reciprocity with trinary reputations. *Journal of Theoretical Biology*, 317:338–347, 2013.

[Uch10] S. Uchida. Effect of private information on indirect reciprocity. *Physical Review E*, 82:036111, 2010.

[US09] S. Uchida and K. Sigmund. The competition of assessment rules for indirect reciprocity. *Journal of Theoretical Biology*, 263(1):13–19, 2009.

[US13] S. Uchida and T. Sasaki. Effect of assessment error and private information on stern-judging in indirect reciprocity. *Chaos, Solitons and Fractals*, 56:175–180, 2013.

[UYOS18] S. Uchida, H. Yamamoto, I. Okada, and T. Sasaki. A theoretical approach to norm ecosystems: Two adaptive architectures of indirect reciprocity show different paths to the evolution of cooperation. *Frontiers in Physics*, 6:14, 2018.

[VA00] Gerthe Veen and Arnoud Arntz. Multidimensional dichotomous thinking characterizes borderline personality disorder. *Cognitive Therapy and Research*, 24(1):23–45, 2000.

[vAS16] J. van Apeldoorn and A. Schram. Indirect reciprocity; A field experiment. *PLoS One*, 11(4):e0152076, 2016.

[vNM53] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 3rd edition, 1953.

[vSPLS12] S. van Segbroeck, J. M. Pacheco, T. Lenaerts, and F. C. Santos. Emergence of fairness in repeated group interactions. *Physical Review Letters*, 108:158104, 2012.

[vVGRN12] M. van Veelen, J. García, D. G. Rand, and M. A. Nowak. Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences USA*, 109:9929–9934, 2012.

[WB02] Claus Wedekind and Victoria A Braithwaite. The long-term benefits of human generosity in indirect reciprocity. *Current Biology*, 12:1012–1015, Jun 2002.

[WBGT15] Bin Wu, Benedikt Bauer, Tobias Galla, and Arne Traulsen. Fitness-based models and pairwise comparison models of evolutionary games are typically different—even in unstructured populations. *New Journal of Physics*, 17:023043, 2015.

[WD19] Z. Wang and R. Durrett. Extrapolating weak selection in evolutionary games. *Journal of Mathematical Biology*, 78:135–154, 2019.

[WGHT13] B. Wu, J. García, C. Hauert, and A. Traulsen. Extrapolating weak selection in evolutionary games. *PLoS Computational Biology*, 9:e1003381, 2013.

[WGWT12] B. Wu, C. S. Gokhale, L. Wang, and A. Traulsen. How small are small mutation rates? *Journal of Mathematical Biology*, 64:803–827, 2012.

[WM00] C. Wedekind and M. Milinski. Cooperation through image scoring in humans. *Science*, 288:850–852, 2000.

[XGH19] J. Xu, J. García, and T. Handfield. Cooperation with bottom-up reputation dynamics. *AAMAS'19*, 2019.

[XRW+17] X. Xu, Z. Rong, Z.-X. Wu, T. Zhou, and K. T. Chi. Extortion provides alternative routes to the evolution of cooperation in structured populations. *Physical Review E*, 95:052302, 2017.

[YD05] Carrie Yurica and Robert DiTomasso. *Cognitive Distortions*, pages 117–122. Springer US, 01 2005.