Intention Recognition Promotes The Emergence of Cooperation

Han The Anh, Luís Moniz Pereira,[*] Francisco C. Santos

Centro de Inteligência Artificial (CENTRIA)
Departamento de Informática, Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

Emails: h.anh@fct.unl.pt, lmp@di.fct.unl.pt, fcsantos@fct.unl.pt

Short title: Intention Recognition and Emergence of Cooperation

[*]Corresponding author. Email address: lmp@di.fct.unl.pt; Phone number (office): +351 212948533; Mobile number: +351 936778968; Fax: +351 212948541

Abstract

Few problems have created the combined interest of so many unrelated areas as the evolution of cooperation. As a result, several mechanisms have been identified to work as catalyzers of cooperative behavior. Yet, these studies, mostly grounded on evolutionary dynamics and game theory, have neglected the important role played by intention recognition in behavioral evolution. Here we address explicitly this issue, characterizing the dynamics emerging from a population of intention recognizers. We derive a Bayesian Network model for intention recognition in the context of repeated social dilemmas and evolutionary game theory, by assessing the internal dynamics of trust between intention recognizers and their opponents. Intention recognizers are then able to predict the next move of their opponents based on past direct interactions, which, in turn, enables them to prevail over the most famous strategies of repeated dilemmas of cooperation, even in presence of noise. Overall, our framework offers new insights on the complexity and beauty of behavioral evolution driven by elementary forms of cognition.

Keywords: Evolution of Cooperation, Intention Recognition, Bayesian Networks, Evolutionary Game Theory.

1 Introduction

Intention recognition can be found abundantly in many kinds of interactions and communications, not only in Human but also many other species (Tomasello, 2008). The knowledge about intention of others in a situation could enable to plan in advance, either to secure a successful cooperation or to deal with potential hostile behaviours (van Hees and Roy, 2008; Roy, 2009). Given the advantage of knowing the intentions of others and the abundance of intention recognition among different species, it is clear that intention recognition should be taken into account when studying or modeling collective behavior. This issue becomes even more relevant when the achievement of a goal by an individual does not depend uniquely on its own actions, but also on the decisions and actions of others, namely when individuals cooperate or have to coordinate their actions to achieve a task, especially when the possibility of communication is limited (Kraus, 1997; Heinze, 2003; Van Segbroeck et al., 2010). For instance, in population-based artificial intelligence applications (Bonabeau et al., 1999; Ampatzis et al., 2008; Gutierrez et al., 2009), such as collective robotics and others, the inherent problem of lack of intention recognition due to the simplicity of the agents is often solved by assuming homogeneous populations, in which each agent has a perfect image of the other as a copy of their own self. Yet, the problem remains in heterogeneous agent systems where it is likely that agents speak different languages, have different designs or different levels of intelligence; hence, intention recognition may be the only way agents understand each other to secure successful cooperation or coordination among heterogeneous agents. Moreover, in more realistic settings where deceiving may offer additional profits, individuals often attempt to hide their real intentions and make others believe in pretense ones (Robson, 1990; Tomasello, 2008; Skyrms, 2010; Pereira and Han, 2011; Santos et al., 2011).

Intention recognition is defined, in general terms, as the process of becoming aware of the intention of another agent and, more technically, as the problem of inferring an agent's intention through its actions and their effects on the environment (Kautz and Allen, 1986; Charniak and Goldman, 1993; Heinze, 2003). For the recognition task, several issues can be raised grounded on the eventual distinction between the model an agent creates about himself and the one used to describe others, often addressed in the context of the "Theory of Mind" theory, which neurologically reposes in part on "mirror neurons", at several cortical levels, as supporting evidence (Iacoboni et al., 2005; Rizzolatti and Craighero, 2004; Nakahara and Miyashita, 2005). The problem of intention recognition has been paid much attention in AI, Philosophy and Psychology for several decades (Kautz and Allen, 1986; Charniak and Goldman, 1993; Bratman, 1987, 1999; Geib and Goldman, 2009). Whereas intention recognition has been extensively studied in small scale interactive settings, there is an absolute lack of modelling research with respect to large scale social contexts; namely the evolutionary roles and aspects of intention recognition.

In this work, we study the role of intention recognition for one of the most challenging but intriguing issues, traversing areas as diverse as Biology, Economics, Artificial Intelligence, Political Science, or Psychology: the problem of *evolution of cooperation* (Hardin, 1968; Axelrod, 1984; Sigmund, 2010). In its simplest form, a cooperative act is metaphorically described as the act of paying a cost to convey a benefit to someone else. If two players simultaneously decide to cooperate or not, the best possible response will be to try to receive the benefit without paying the cost. In an evolutionary setting, we may also wonder why would natural selection equip selfish individuals with altruistic tendencies while it incites competition between individuals and thus apparently rewards only selfish behavior? Several mechanisms responsible for promoting cooperative behavior have been recently identified (Sigmund, 2010; Nowak, 2006). From kin and group ties (West et al., 2007; Traulsen and Nowak, 2006), to different forms of reciprocity (Nowak and Sigmund, 1992a; Imhof et al., 2005; Trivers, 1971; Pacheco et al., 2006; Nowak and Sigmund, 2005) and networked populations (Santos and Pacheco, 2005; Santos et al., 2006; Szabó and Fáth, 2007; Santos et al., 2008; Lindgren and Nordahl, 1994), several aspects have been shown to play an important role in the emergence of cooperation. Differently, here we shall describe how cooperation may emerge from the interplay between population dynamics and individuals' cognitive abilities, namely the ability to perform intention recognition.

Our study is carried out within the framework of Evolutionary Game Theory (EGT) (Hofbauer and Sigmund, 1998). Here, individual success (or fitness) is expressed in terms of the outcome of a 2-person game, which, in turn, is used by individuals to copy others whenever these appear to be more successful. Comparative accumulated payoffs are used to update the population: more successful individuals produce more offspring, which inherit their strategy. Equivalently, the same process can be seen as if, instead of inheriting strategies, new individuals adapt by copying strategies from acquaintances that did better. Overall, this type of dynamics can be conveniently described as an ordinary differential equation – the replicator equation (Hofbauer and Sigmund, 1998)–, which nicely describes any simple evolutionary process. This framework is however more general one could initially foresee, as the ensuing dynamics may be also shown to be equivalent to finite-action learning automata (Börgers et al., 1997; Van Segbroeck et al., 2010), in which agents revise their strategies by means of incipient reinforcement learning techniques (Narendra and Thathachar, 1989).

In this work we model intention recognition within the framework of repeated interactions. In the context of direct reciprocity (Trivers, 1971) intention recognition is being performed using the information about past *direct* interactions. We study this issue using the repeated Prisoner's Dilemma (PD), i.e., intentions are inferred from past individual experiences. Naturally, the same principles could be extended to cope with indirect information, as in indirect reciprocity (Nowak and Sigmund, 2005; Pacheco et al., 2006; Ohtsuki and Iwasa, 2006). This eventually introduces moral judgment and concern for individual reputation, which constitutes "per se" an important area where intention recognition may play a pivotal role. Here, however, we shall concentrate on the simpler case of intention recognition from past experiences.

Contrary to other approaches dealing with the integration of (direct or indirect) information about the past in individual decisions, e.g. in (Masuda and Ohtsuki, 2009; Ohtsuki and Iwasa, 2006; Wang et al., 2008; Vukov et al., 2011), intention recognition is performed using a Bayesian Network (BN) model. BNs have proven to be one of the most successful approaches for intention recognition (Charniak and Goldman, 1993; Tahboub, 2006; Pereira and Han, 2011; Geib and Goldman, 2009). Their flexibility for representing probabilistic dependencies as well as causal relations, and the efficiency of inference methods have made them an extremely powerful tool for problem solving under uncertainty

(Pearl, 1988, 2000), and appropriate to deal with several probabilistic as well as causal dependencies occurring in intention recognition.

We derive a Bayesian Network model for intention recognition in the context of social dilemmas, taking into account mutual trusts between the intention recognizer and his opponent. Trusts are accumulated through past interactions, assuming that intention recognizers have a memory. Greater memory sizes enable to build longer-term mutual trusts, and therefore allow better tolerance to the errors of intended actions. We study analytically the case of small memory size, and experimentally the effect of having greater memory sizes. In addition, we compare the performance of intention recognizers with the most famous strategies of the repeated PD.

The repeated (or iterated) PD is usually known as a story of tit-for-tat (TFT), which won both Axelrod's tournaments (Axelrod, 1984; Axelrod and Hamilton, 1981). *TFT* starts by cooperating, and does whatever the opponent did in the previous round. It will cooperate if the opponent cooperated, and will defect if the opponent defected. But if there are erroneous moves because of noise (i.e. an intended move is wrongly performed with a given execution error, referred here as "noise"), the performance of *TFT* declines, in two ways: (i) it cannot correct errors and (ii) a population of *TFT* players is undermined by random drift when *AllC* (always cooperate) mutants appear (which allows exploiters to grow). Tit-for-tat is then advantageously replaced by generous tit-for-tat (GTFT), a strategy that cooperates if the opponent cooperated in the previous round, but sometimes cooperates even if the opponent defected (with a fixed probability $p > 0$) (Nowak and Sigmund, 1992b). *GTFT* can correct mistakes, but remains suffering the random drift; in addition, it deals with pure defectors worse than *TFT*.

Subsequently, *TFT* and *GTFT* were replaced by win-stay-lose-shift (WSLS) as the winning strategy chosen by evolution (Nowak and Sigmund, 1993). *WSLS* repeats the previous move whenever it did well, but changes otherwise. *WSLS* corrects mistakes better than *GTFT* and does not suffer random drift. However, it is exploited seriously by pure defectors.

Here we show that our innovative intention recognition strategy (IR) can correct mistakes even better than *WSLS*, and not be exploited by pure defectors. We compare the performance of *TFT, WSLS* and *IR* under mutation-selection dynamics in finite populations (Imhof et al., 2005; Hauert et al., 2007), in a well-mixed population of pure cooperators (AllC), pure defectors (AllD) and, additionally, of individuals of either of the above other three strategies, as well as all these strategies together. The results show that *IR* performs best, in the sense that populations spend more time in a homogenous state of *IR*s, even in the presence of noise.

2 Materials and Methods

We consider a population of constant size N. At each evolution step, a random pair of players are chosen to play with each other. The population consists of pure cooperators, pure defectors plus either of *TFT*s or of *WSLS*s or of intention recognizers who, being capable of recognizing another's intention based on the past interactions, seek the cooperators to cooperate with and to defect toward detected defectors.

2.1 Interaction between Individuals

Interactions are modeled as symmetric two-player games defined by the payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R, R & S, T \\ T, S & P, P \end{pmatrix} \end{array}$$

A player who chooses to cooperate (C) with someone who defects (D) receives the sucker's payoff S, whereas the defecting player gains the temptation to defect, T. Mutual cooperation (resp., defection) yields the reward R (resp., punishment P) for both players. Depending on the ordering of these four payoffs, different social dilemmas arise (Macy and Flache, 2002; Santos et al., 2006; Sigmund, 2010). Namely, in this work we are concerned with the Prisoner's Dilemma (PD), where $T > R > P > S$. In a single round, it is always best to defect, but cooperation may be rewarded if the game is repeated. In repeated PD, it is also required that mutual cooperation is preferred over an equal probability of unilateral cooperation and defection ($2R > T + S$); otherwise alternating between cooperation and defection would lead to a higher payoff than mutual cooperation. For convenience and a clear representation of results, we later mostly use the Donation game (Sigmund, 2010)—a famous special case of the PD—where $T = b,\ R = b - c,\ P = 0,\ S = -c$, satisfying that $b > c > 0$, where b and c stand respectively for "benefit" and "cost" (of cooperation).

In a population of N individuals interacting via a repeated (or iterated) Prisoner's dilemma, whenever two specific strategies are present in the population, say **A** and **B**, the fitness of an individual with a strategy **A** in a population with k **A**s and $(N - k)$ **B**s can be written as

$$\Pi_A(k) = \frac{1}{r(N-1)} \sum_{j=1}^{r} [(k-1)\pi_{A,A}(j) + (N-k)\pi_{A,B}(j)] \tag{1}$$

where $\pi_{A,A}(j)$ ($\pi_{A,B}(j)$) stands for the payoff obtained from a round j as a result of their mutual behavior of an **A** strategist in an interaction with a **A** (**B**) strategist (as specified by the payoff matrix above), and r is the total number of rounds of the Prisoner's dilemma. As usual, instead of considering a fixed number of rounds, upon completion of each round, there is a probability w that yet another round of the game will take place, resulting in an average number of $< r >= (1 - w)^{-1}$ rounds per interaction (Sigmund, 2010). In the following, all values of Π will be computed analytically. When this is not possible, we shall use numerical simulations, as stated below.

2.2 Bayesian Networks

Definition 2.1 *A Bayesian Network (BN) is a pair consisting of a directed acyclic graph (DAG) whose nodes represent variables and missing edges encode conditional independencies between the variables, and an associated probability distribution satisfying the Markov assumption of conditional independence, saying that variables are independent of non-descendants given their parents in the graph (Pearl, 1988, 2000).*

In a BN, associated with each node of its DAG is a specification of the distribution of its variable, say A, conditioned on its parents in the graph (denoted

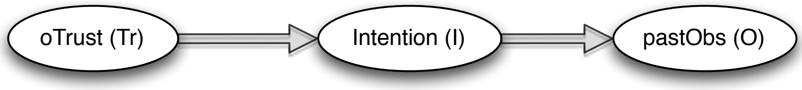

Figure 1: Bayesian Network for Intention Recognition in Social Dilemmas. Pre-intentional level has one node, *oTrust (Tr)*, receives Boolean values, t (*true*) or f (*false*), representing the other's trust on us (the intention recognizers). Intentional level has one node, *Intention (I)*, receiving value C or D, corresponding to more cooperative and more defective, respectively, in the past. It is causally affected by *oTrust*. Activity level has one node, *pastObs (O)*, causally affected by *Intention* node. Its value is a pair (n_C, n_D) where n_C and n_D are the number of times the recognized player cooperated and defected, respectively, in the recent M (memory size) steps. *pastObs* is the only observed (evidence) node.

by $pa(A)$)—i.e., $P(A|pa(A))$ is specified. If $p(A) = \emptyset$ (A is called root node), its unconditional probability distribution, $P(A)$, is specified. These distributions are called Conditional Probability Distribution (CPD) of the BN.

The joint distribution of all node values can be determined as the product of conditional probabilities of the value of each node on its parents $P(X_1, ..., X_N) = \prod_{i=1}^{N} P(X_i|pa(X_i))$, where $V = \{X_i | 1 \leq i \leq N\}$ is the set of nodes of the DAG.

Suppose there is a set of evidence nodes (i.e. their values are observed) in the DAG, say $O = \{O_1, ..., O_m\} \subset V$. We can determine the conditional probability distribution of a variable X given the observed value of evidence nodes by using the conditional probability formula

$$P(X|O) = \frac{P(X, O)}{P(O)} = \frac{P(X, O_1, ..., O_m)}{P(O_1, ..., O_m)} \tag{2}$$

where the numerator and denominator are computed by summing up the joint probabilities over all absent variables with respect to V.

2.3 Intention Recognition in Social Dilemmas

In (Pereira and Han, 2011), a general BN model for intention recognition is presented and justified based on Heinze's intentional model (Heinze, 2003; Tahboub, 2006). Basically, the BN consists of three layers: cause/reason nodes in the first layer (called *pre-intentional*), connecting to intention nodes in the second one (called *intentional*), in turn connecting to action nodes in the third (called *activity*). Intuitively, the observed actions of an agent are causally affected by his/her intentions, which are in turn causally affected by the causes/reasons for which he committed to the intentions (Bratman, 1987, 1999). The interested readers are referred to (Pereira and Han, 2011; Heinze, 2003; Tahboub, 2006) for detailed discussions.

Based on this general model, we present an intention recognition model in the context of the social dilemmas, taking into account the past *direct* interactions (Figure 1). The model is described from the view of an intention recognizer (denoted by $\mathcal{I}$) with respect to a co-player (denoted by $\mathcal{J}$), whose intention (C or D) is to be recognized. A player's intentions here can be understood as the characters or types of the player: how cooperative or defective he is in general when playing with me. Saying that the co-player has intention C (resp., D) means that, in general, he intends to cooperate with me (resp., exploit or defect towards me). Thus, if he has been cooperative in the past, it is likely he will continue to cooperate in the current interaction.

$\mathcal{J}$'s intention in a given interaction is causally affected by the trust he holds towards his opponent ($\mathcal{I}$), which is accumulated over their past (observed) interactions. $\mathcal{J}$'s intention in turn has given rise to his past actions. Let $M > 0$ be the memory size of intention recognizers, i.e. they can remember their moves and their opponents' moves in the last M rounds of interaction with any specific players.

For this Bayesian Network, we need to determine the prior probability of the node *oTrust*, i.e. $P(Tr)$; the CPD table of node *Intention*—specifying the conditional probability of $\mathcal{J}$ having an intention (C or D) given the trust he holds towards his opponent ($\mathcal{I}$), i.e. $P(I|Tr)$; and the CPD table of the node *pastObs*—specifying the conditional probability of the past observations given $\mathcal{J}$'s intention (C or D), i.e. $P(O|I)$.

To begin with, let $n_C(\mathcal{X}, \mathcal{Y})$ and $n_D(\mathcal{X}, \mathcal{Y})$ denote the numbers of times a player $\mathcal{X}$ cooperated and defected, respectively, in the last M interactions with another player $\mathcal{Y}$. Note that $n_C(\mathcal{X}, \mathcal{Y}) + n_D(\mathcal{X}, \mathcal{Y}) \leq M$, and the equality occurs only when the two players have interacted with each other at least M times.

Trust Distribution. The probability that $\mathcal{J}$ trusts $\mathcal{I}$ is given by how often $\mathcal{I}$ cooperated with $\mathcal{J}$. This can be written as

$$P(Tr = t) = \frac{1}{2} + \frac{n_C(\mathcal{I}, \mathcal{J}) - n_D(\mathcal{I}, \mathcal{J})}{2M} \tag{3}$$

It is easily seen that $0 \leq P(Tr = t) \leq 1$, and $P(Tr = t) = 0$ (resp., 1) if $\mathcal{I}$ defected (resp., cooperated) in all recent M interactions. These correspond to the extremes that $\mathcal{I}$ lost all his/her trust (resp., gained complete trust) concerning $\mathcal{J}$. We further assume that, in the first interaction, the trust level is neutral: $P(Tr = t) = \frac{1}{2}$.

Definition of $\mathbf{P(I|Tr)}$. We use the following CPD table

$$P(I = C|Tr = t) = P(I = D|Tr = f) = h$$

$$P(I = C|Tr = f) = P(I = D|Tr = t) = 1 - h$$

where h is the probability the intention recognizer ($\mathcal{I}$) thinks the co-player ($\mathcal{J}$) has intention C given that he ($\mathcal{J}$) completely trusts me ($\mathcal{I}$).

As the intention recognizers are cooperative, i.e. they seek the cooperators to cooperate with and generously start by cooperating with everybody, we assume $h \geq 0.5$. This probability reflects the intention recognizers' **optimistic level**—assumed fixed for their entire life cycle (generation).

Definition of $\mathbf{P(O|I)}$. The conditional probability of the past observations about the co-player given his intention. It can be interpreted as how trustful or cooperative the intention recognizer ($\mathcal{I}$) thinks his co-player ($\mathcal{J}$) is, and can be defined as how often $\mathcal{J}$ cooperated with $\mathcal{I}$ in an interaction. It can be given as follows

$$P(O = (n_1, n_2)|I = C) = \frac{1}{2} + \frac{n_1 - n_2}{2M}$$
$$P(O = (n_1, n_2)|I = D) = \frac{1}{2} + \frac{n_2 - n_1}{2M} \quad (4)$$

where $n_1 = n_C(\mathcal{J}, \mathcal{I})$ and $n_2 = n_D(\mathcal{J}, \mathcal{I})$.

In a nutshell, the intention recognition model presented takes into account the past direct interactions, in terms of mutual trusts, which are encoded into a Bayesian Network.

Intention Recognizer. In an interaction, the probabilities of the co-player having intention C and having intention D, given his M recent past actions $o = (n_C, n_D)$, are computed using Eq. (2)

$$p(I = C|O = o) = \frac{p(I = C, O = o)}{p(O = o)}$$
$$p(I = D|O = o) = \frac{p(I = D, O = o)}{p(O = o)} \quad (5)$$

The intention recognizer plays C if he recognizes that the co-player is more cooperative (thus more likely to play C than to play D), i.e. $p(I = C|O = o) \geq p(I = D|O = o)$. Otherwise, he plays D. Players using this intention recognition strategy are henceforth referred to as *IR* players.

With the trust functions given in Eq. (3) and (4), we obtain that the intention recognizer ($\mathcal{I}$) cooperates with his co-player ($\mathcal{J}$) iff (see Appendix A for details)

$$\Delta = sP + Q \geq 0 \quad (6)$$

where $s = 2h - 1$ $(0 \leq s \leq 1)$; $P = n_C(\mathcal{I}, \mathcal{J}) - n_D(\mathcal{I}, \mathcal{J})$ and $Q = n_C(\mathcal{J}, \mathcal{I}) - n_D(\mathcal{J}, \mathcal{I})$.

2.4 Evolutionary Dynamics

The accumulated payoff from all interactions (see Eq. (1)) emulates the individual *fitness* or social *success* and the most successful individuals will tend to be imitated by others, implementing a simple form of social learning (Szabó and Toke, 1998; Traulsen et al., 2006; Sigmund, 2010). Any player (including *IR*) can change its strategy by adopting another players strategy with a probability defined by the Fermi distribution below. If a strategy has a higher (average) payoff or fitness than another, it tends to be imitated more by the other. The *IR* strategy in general has higher fitness than all others, thus it tends to by imitated by them, thereby dominating the population most of the time.

A strategy update event is defined in the following way, corresponding to the so-called pairwise comparison (Szabó and Toke, 1998; Traulsen et al., 2006). At each time-step, one individual i with a fitness f_i is randomly chosen for behavioral revision. i will adopt the strategy of a randomly chosen individual j with fitness f_j with a probability given by the Fermi function (from statistical physics)

$$p(f_i, f_j) = \left(1 + e^{-\beta(f_j - f_i)}\right)^{-1}$$

where the quantity β, which in physics corresponds to an inverse temperature, controls the intensity of selection. When $\beta = 0$ we obtain the limit of neutral drift, and with the increasing of β one enhances the role played by the game payoff in the individual fitness, and behavioral evolution (Traulsen et al., 2006, 2007).

In the absence of mutations, the end states of evolution are inevitably monomorphic, as a result of the stochastic nature of the evolutionary dynamics and update rule. As we are interested in a global analysis of the population dynamics with multiple strategies, we further assume that with a small probability μ individuals switch to a randomly chosen strategy, freely exploring the space of possible behaviors. By introducing a small probability of mutation or exploration, the eventual appearance of a single mutant in a monomorphic population, this mutant will fixate or will become extinct long before the occurrence of another mutation and, for this reason, the population will spend all of its time with a maximum of two strategies present simultaneously (Fudenberg and Imhof, 2005; Imhof et al., 2005; Hauert et al., 2007; Santos et al., 2011). This allows one to describe the evolutionary dynamics of our population in terms of a reduced Markov Chain of a size equal to the number of different strategies, where each state represents a possible monomorphic end-state of the population associated with a given strategy, and the transitions between states are defined by the fixation probabilities of a single mutant of one strategy in a population of individuals who adopt another strategy. The resulting stationary distribution characterizes the average time the population spends in each of these monomorphic states, and can be computed analytically (Karlin and Taylor, 1975; Fudenberg and Imhof, 2005; Imhof et al., 2005; Hauert et al., 2007; Santos et al., 2011) (see below).

In the presence of two strategies the payoffs of each are given by Eq. (1), whereas the probability to change the number k of individuals with a strategy **A** (by $\pm$ one in each time step) in a population of $(N-k)$ **B**-strategists is

$$T^{\pm}(k) = \frac{N-k}{N}\frac{k}{N}\left[1 + e^{\mp\beta[\Pi_A(k)-\Pi_B(k)]}\right]^{-1}$$

The fixation probability of a single mutant with a strategy **A** in a population of $(N-1)$ **B**s is given by (Traulsen et al., 2006)

$$\rho_{B,A} = \left(\sum_{i=0}^{N-1}\prod_{j=1}^{i}\lambda_j\right)^{-1} \tag{7}$$

where $\lambda_j = T^-(j)/T^+(j)$.

In the limit of neutral selection ($\beta = 0$), $\lambda_j = 1$. Thus, $\rho_{B,A} = 1/N$. Considering a set $\{1, ..., n_S\}$ of different strategies, these fixation probabilities determine a transition matrix $[T_{ij}]_{i,j=1,...,n_S}$, with $T_{ii} = 1 - \sum_{k=1,k\neq i}^{n_S}\rho_{k,i}/(n_S - 1)$ and $T_{ij,j\neq i} = \rho_{ji}/(n_S - 1)$, of a Markov Chain. The normalized eigenvector associated with the eigenvalue 1 of the transposed of M provides the stationary distribution described above (Karlin and Taylor, 1975; Fudenberg and Imhof, 2005; Imhof et al., 2005; Hauert et al., 2007; Santos et al., 2011), describing the relative time the population spends adopting each of the strategies.

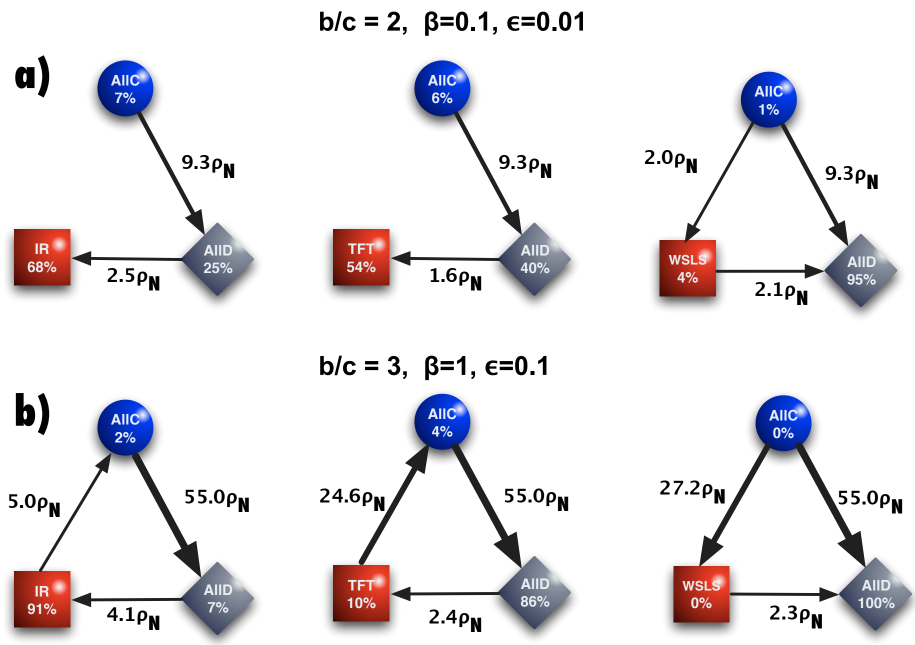

Figure 2: Transition probabilities and stationary distributions (in percentage), computed analitically for two distinct sets of parameters. We consider a population of *AllC*s, *AllD*s and either *WSLS*s, *TFT*s or *IR*s ($M = 2$). The black arrows are only shown for the transition directions that are rather more likely than neural. The strongest transition is from *AllC* to *AllD*. The transition from *AllD* to *IR* is stronger than to *TFT*. This is reversed in case of *WSLS*, where the most probable transition occurs from *WSLS* to *AllD*. For slow intensity of selection β (panel a), the transitions between *AllC* and *IR* and *AllC* and *TFT* are near neutral, and for strong selection (panel b), there is a transition from *TFT* to *AllC*, which is stronger than from *IR* to *AllC*. Also, the greater β, the stronger the transition from *AllC* to *AllD* and to *WSLS*. The calculations in both cases are made with $N = 100$ and $\omega = 0.9$; $\rho_N = 1/N$ denotes the neutral fixation.

3 Results

We shall start by considering a finite population consisting of *AllC*, *AllD* and *IR* players. It is easily seen that in the absence of noise, a player adopting a *IR* strategy performs similarly to a *TFT* player, i.e. always cooperates with an *AllC*, always defects with *AllD* after cooperating in the first round, and always cooperate amongst themselves. In the sequel we study the performance of *IR* in the presence of noise, and compare with *TFT* and *WSLS*. First we study analytically the case of memory two, i.e. $M = 2$.

3.1 Evolution of short-memory intention recognizers

In the presence of noise, let us assume that an intended action (C or D) can fail with probability $\epsilon \in [0, 1]$. We obtain the following payoff matrix for *AllC*, *AllD* and *IR*, where all terms of order $O(\epsilon^2)$ have been ignored (see Appendix B.1)

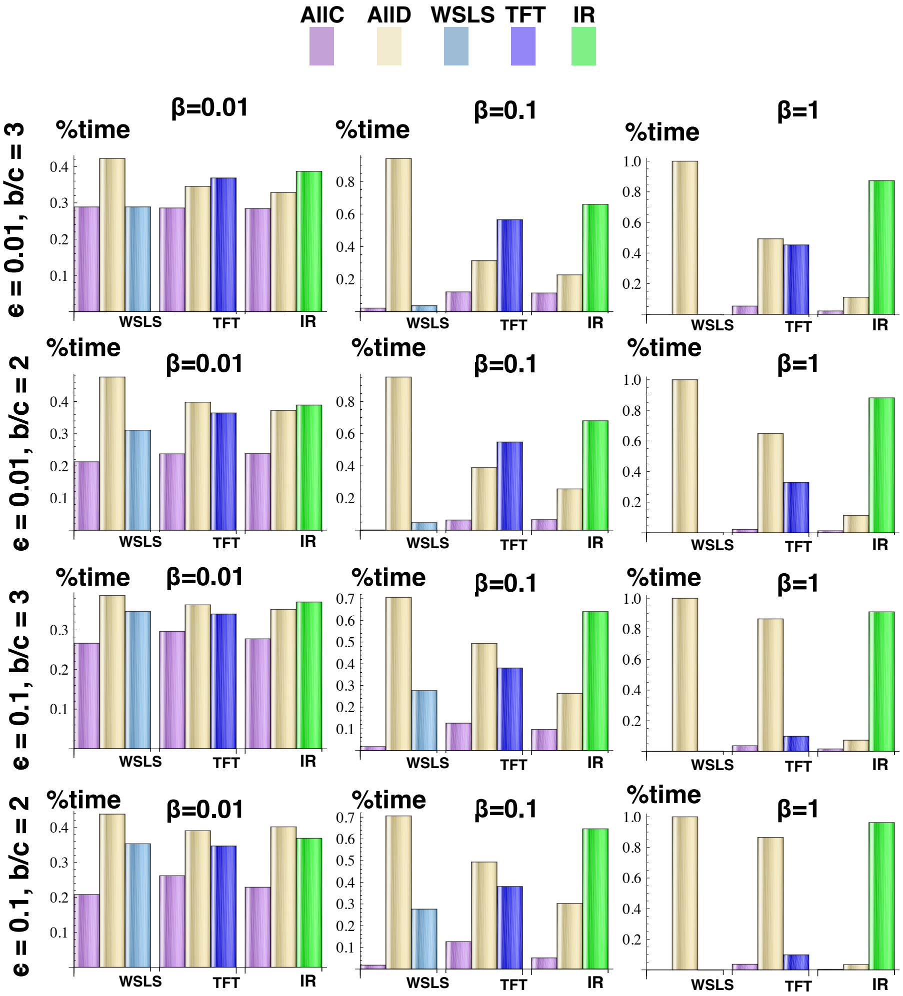

Figure 3: Percentage of time spent at *AllC*, *AllD* and either *WSLS*, *TFT* or *IR*. We compare *IR* with *WSLS* and *TFT* when interacting with *AllC*s and *AllD*s. We consider different selection intensities β (0.01, 0,1 and 1), noise levels ϵ (0.01 and 0.1) and PD benefit-to-cost ratios b/c (2 and 3). In all cases, *IR* is better than *TFT* and *WSLS*. In a population of *AllC*s, *AllD*s and *IR*s, the system spends more time in homogeneous state of *IR*s, especially when selection is strong. *TFT* and *WSLS* perform poorly at strong selection intensities, as the population spends most of the time at *AllD*s. The calculations in all cases are made with $N = 100$ and $\omega = 0.9$.

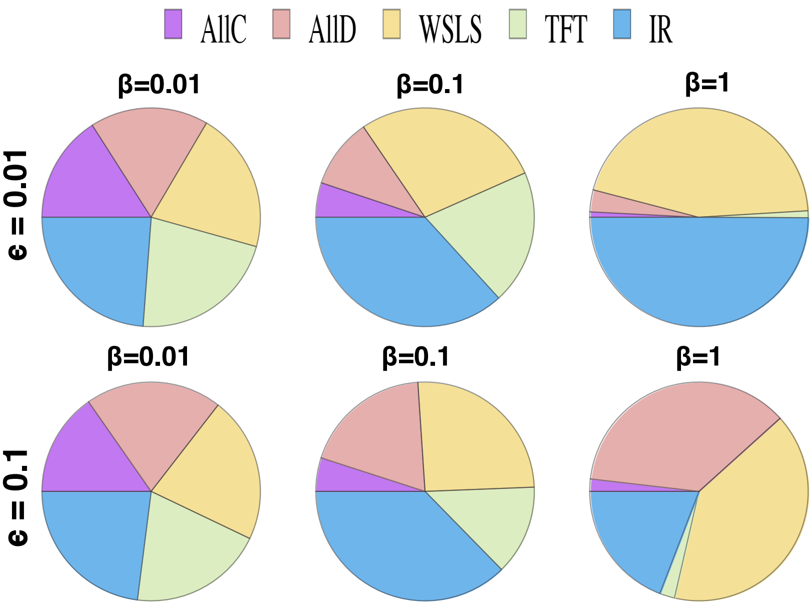

Figure 4: Stationary distribution in a population of five possible strategies *AllC*, *AllD*, *WSLS*, *TFT* and *IR* ($M = 2$). The population spends most of the time in a homogeneous state of *IR*s. *WSLS* also performs well in this 5-strategy setting, reconfirming that it needs other catalyzers such as *TFT* to perform well. On the contrary, these results show that *IR* performs well in either case (the other setting in Figure 3). The calculations in all cases are made with $N = 100$, $b/c = 3$, $\omega = 0.9$ and $h = 0.6$. The average fitness of *IR*, *WSLS* and *TFT* players interacting with each other are obtained by averaging over 10^7 simulated interactions.

$$\begin{pmatrix} (b-c)(1-\epsilon) & -c+(b+c)\epsilon & \frac{c\left(\omega-1+\epsilon\left(1-4\omega+\omega^2\right)\right)}{1-\omega+3\epsilon\omega-\epsilon\omega^2} + \frac{b\left(\omega-1+2\epsilon\left(1-3\omega+\omega^2\right)\right)}{1-\omega+3\epsilon\omega-\epsilon\omega^2} \\ b-b\epsilon-c\epsilon & (b-c)\epsilon & \frac{\epsilon\left(b\left(2+\omega-\omega^2\right)-c\right)}{1-\epsilon\omega} \\ \frac{b\left(1-\epsilon-\epsilon\omega^2\right)-c\left(1-\epsilon\left(2-\omega+\omega^2\right)\right)}{1-\epsilon\omega^2} & \frac{\epsilon\left(b-c\left(2+\omega-\omega^2-\omega^3\right)\right)}{1-\epsilon\omega^2(1+\omega)} & \frac{(b-c)\left(1+\epsilon\left(-2+4\omega+\omega^2+\omega^3\right)\right)}{1+\epsilon\omega(3+\omega+\omega^2)} \end{pmatrix}$$

Let $A(X, Y)$ be the payoff of strategist X when playing against strategist Y. We can show that (Appendix B.2)

$$A(IR, IR) > A(WSLS, WSLS) > A(TFT, TFT)$$
$$A(AllD, IR) < A(AllD, TFT) < A(AllD, WSLS)$$
$$A(IR, AllD) > A(TFT, AllD) > A(WSLS, AllD)$$
$$A(AllC, WSLS) < A(AllC, TFT) < A(AllC, IR)$$
$$A(WSLS, AllC) > A(TFT, AllC) > A(IR, AllC)$$

The first inequality implies that *IR* deals with noise better than *TFT* and *WSLS*, when interacting with individuals alike. As a result, a homogeneous population

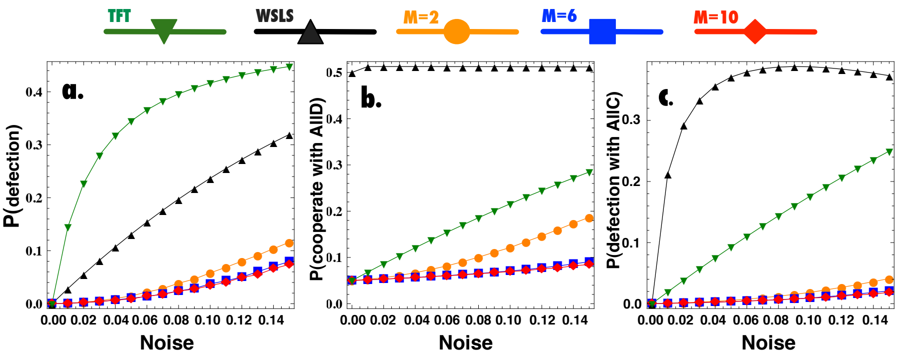

Figure 5: Plot of probabilities of reactions computed numerically for *IR*s of different memory sizes ($M = 2, 6, 10$), and analogous quantities for *WSLS* and *TFT* strategists, as a function of the level of noise. a) Probability of defection of *IR*, *TFT* and *WSLS* strategists when playing with themselves. The probability of defection between *IR*s decreases with the memory size. b) Probability that each of the three strategies—*IR*, *TFT* and *WSLS*—cooperates with a *AllD*. The probability that *IR* cooperates with *AllD* decreases when memory size increases. c) Probability that a *IR*, *TFT* and *WSLS* player defects with a *AllC*. The probability that *IR* defects with *AllC* decreases when memory size increases. In all cases, $\omega = 0.9$, $h = 0.6$ and the probabilities are obtained by averaging over 10^7 simulated interactions.

of *IR*s has a higher level of cooperation (thus, greater average payoff) than the ones of *WSLS*s and *TFT*s [1]. The next two inequalities imply that *IR* deals with *AllD* better than *TFT*, which is in turn better than *WSLS*. The fixation probability of an *IR* taking over a population of *AllD* is greater than those of *TFT* and *WSLS* (Figure 2). Finally, the last two inequalities imply that *IR* is more cooperative to *AllC* (i.e. it is more tolerant of noise originated by *AllC*).

Figures 2 and 3 show a comparison among *WSLS*, *TFT* or *IR* in terms of the percentage of time the population spends at their homogenous state, in a setting of either strategy and *AllC* and *AllD* individuals. In all cases, *IR* is better than *TFT* and *WSLS*. In a population of *AllC*s, *AllD*s and *IR*s, the population spends more time in the homogeneous state of *IR*s, especially when the intensity of selection is strong. *TFT* and *WSLS* poorly perform at strong selection intensities—the population spends most of the time at *AllD*s. The poor performance of *WSLS* is not surprising, as *WSLS* needs *TFT* players as a catalyst to perform well, as discussed in (Sigmund, 2010).

Figure 4 shows the results for the setting where all the five strategies *AllC*, *AllD*, *TFT*, *WSLS* and *IR* are in the population. Again, the population will be most likely found in the homogeneous state of *IR*s. *WSLS* also performs well in this setting (since *TFT*s are present).

In short, in these two commonly used settings, *IR* always outperforms *TFT* and *WSLS*. The population spends more time in the homogeneous state of *IR*s. Furthermore, since a population of *IR*s is highly cooperative, it is clear that the introduction of intention recognition significantly increases the cooperation level of the population, leading to a greater social welfare.

[1] Recall that a homogeneous population or homogeneous state is a population with just a single strategy, i.e., all individuals of the population adopt that strategy.

3.2 The role of memory size

The impressive results obtained in the previous section addressed the evolutionary chances of very short-memory intention recognizers. Yet, it is reasonable to suppose that individuals may record a larger number of rounds and use it at their profit. As shown in Figure 5a, a greater memory size enables a better tolerance to noise. It allows the build-up of long-term mutual trusts, which enables a better assessment of errors. The intention recognizers become more generous (see Fig. 5a) and, above all, more tolerant to other players' errors. As a result, a homogenous population of large memory *IR*s can reach closer to the state of perfect cooperation, even in the presence of noise.

In addition, they are more resilient to change opinion about untrustworthy players. Namely, the greater M the smaller probability that an *IR* defects with another *IR* and with an *AllC* (see Fig. 5c), and the smaller the probability of cooperating with *AllD* (see Fig. 5b).

3.3 The role of IRs optimism and their tolerance to noise

In the following, we study the relation of the optimistic level h of an *IR* with his/her noise tolerance, towards a cooperative and towards a defective co-player. For simplicity, we consider the following two cases.

First of all, suppose that at the m-th round, we have $P = P_0 > 0$ and $Q = Q_0 > 0$ (i.e. the players were cooperative), and the co-player will constantly defect (with probability $1 - \epsilon$). Let k be the expected number of rounds *IR* is tolerant to his co-player's defection. Clearly, $k \leq M$ (see Appendix A).

If $m \geq M$, from Eq. (6) and the fact that *IR*'s memory size is M we have that k must satisfy the following equation

$$sP_k + Q_k = 0 \tag{8}$$

where $P_{i+1} = P_i + (1 - \epsilon - P_i/M)$ and $Q_{i+1} = Q_i - (1 - \epsilon + Q_i/M)$ for all $1 \leq i \leq k - 1$. This can be explained as follows. With probability $(1 - \epsilon)$, *IR* cooperates, thus the value of P_i is increased by $(1 - \epsilon)$. But since $m \geq M$, the value of P_i is also decreased, on average, by P_i/M (one round is "forgotten"). A similar argument can be used for Q_i.

By a simple computation[2] we obtain that (8) is equivalent to

$$\left(\frac{M-1}{M}\right)^k = \frac{M(1-s)(1-\epsilon)}{M(1-s)(1-\epsilon) + sP_0 + Q_0} = \frac{1}{1 + \frac{sP_0+Q_0}{M(1-s)(1-\epsilon)}}$$

The right-hand side is clearly a decreasing function of s. Thus, k is an increasing function of s. Now if $m < M$, by putting $k = k_1 + M - m$ and applying the same method we can show that k_1 is also an increasing function of s.

In short, in both cases k is an increasing function of s (hence, also of h). It means that the more optimistic the *IR* is (i.e. the greater h), the more tolerant/generous he is towards a cooperative co-player. It also means that the more optimistic a *IR*, he might become more generous to defective players if errors occur more frequently at the beginning of their interaction (defective players show up to be cooperative).

[2] Consider the sequence X_i, where $X_i = sP_i + Q_i$. It satisfies the recursive equation $X_{i+1} = \frac{M-1}{M} X_i + (1 - s)(1 - \epsilon)$.

Similarly, let us consider the case where at some round, both players were defective, i.e. $P_0 < 0$ and $Q_0 < 0$, and the co-player starts cooperating (either by mistakes or because the co-player is actually a cooperative one). The number of rounds that *IR* keeps on defecting while the co-player cooperates, by using the same method, can be shown to be an increasing function of h. Hence, the more optimistic a *IR* is, the less generous he is towards a defective player. However, it also means that the more optimistic a *IR* is, the less generous he might become to cooperative players if errors occur more frequently at the beginning of their interaction (cooperative players show up to be defective).

In short, when errors are not frequent at the beginning, a more optimistic *IR* becomes more tolerant to noise, since he/she is more generous towards cooperative players and less so to defective ones. On the other hand, if errors are frequent at the beginning (cooperative players show up to be defective, and vice-versa, defective players show up to be cooperative), the more optimistic a *IR* is, more rounds he/she would take to recognize correctly the opponents' intentions. Overall, this suggests that, ideally, a *IR* should not be optimistic at the beginning of an interaction since otherwise an error could create a wrong bad impression which is hard to recover. When more interactions are made, a higher optimism increases the tolerance to noise of a *IR* strategist. As a result, it suggests that h can be expressed as an increasing function of time.

4 Conclusions

Using the tools of EGT, we have addressed the role played by intention recognition in the evolution of cooperation. In this work, we have shown, in a novel way, the role of intention recognition for the emergence of cooperation within the framework of the repeated Prisoner's Dilemma. Intention recognition is performed using a Bayesian Network model via computing mutual trusts between the intention recognizers and their opponents. Given the broad spectrum of problems which are addressed using this cooperative metaphor, our result indicates how intention recognition can be pivotal in social dynamics. We have shown that the intention recognition strategy prevails over the most successful existent strategies (*TFT*, *WSLS*) of the repeated PD, even when players have a very limited memory. *IR* deals with *AllD* better than *TFT* – the best known defector-dealer, and correct mistake better than *WSLS* – the best known mistake-corrector (Nowak, 2006; Sigmund, 2010). As a result, a homogenous population of *IR*s has a higher level of cooperation than the ones of *WSLS*s and *TFT*s, resisting the invasion of other strategies.

In (Imhof et al., 2005), it has been shown that in absence of noise, in a population of *AllC*s, *AllD*s and *TFT*s, the population spends most of the time in a homogeneous state of *TFT*s. However, as we have shown here, it is not the case if noise is present, especially under strong selection. In absence of noise, *IR* behaves the as well as *TFT*. Moreover, *IR*s are selected by evolution in the latter case where noise is present. We have shown that in a population of *AllC*s, *AllD*s and *IR*s, the population spends most of the time in homogeneous state of *IR*s in a broad range of scenarios and parameters, especially when the intensity of selection is strong. We have also exhibited experimentally that in a population where all the five strategies *AllC*, *AllD*, *TFT*, *WSLS* and *IR* are present, *IR*s still prevail most of the time. Therefore, together with the fact that

*IR*s can correct mistakes better than *WSLS*s and *TFT*s, the presence of *IR*s would significantly increase the overall level of cooperation of the population.

Additionally, we have shown the role of a large memory size in recognizing/correcting errors. Having a greater memory size allows to build longer-term mutual trusts/distrusts, and hence enables to better recognize erroneous moves. It then enables to better tolerate of a selfish act made by cooperative trustful individuals, and refuses to cooperate after an erroneous cooperation made by a defective untrustworthy ones. Indeed, intention recognition gives rise to an incipient mechanism of commitment formation, from which future behaviors may be assessed and trust bonds established.

Overall, our work provides new insights on the complexity and beauty of behavioral evolution driven by elementary forms of cognition.

5 Acknowledgments

HTA and FCS acknowledge the support from FCT-Portugal
(grant SFRH/BD/62373/2009 and R&D project PTDC/FIS/101248/2008, respectively).

References

Ampatzis, C., Tuci, E., Trianni, V., and Dorigo, M. (2008). Evolution of signaling in a multi-robot system: Categorization and communication. *Adaptive Behavior*, 16(1):5–26.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books, ISBN 0-465-02122-2, New York.

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

Bonabeau, E., Dorigo, M., and Theraulaz, G. (1999). *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, USA.

Börgers, T., Sarin, R., and Street, G. (1997). Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77:1–14.

Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. The David Hume Series, CSLI, Harvard University Press, Cambridge, Massachusetts.

Bratman, M. E. (1999). *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge University Press.

Charniak, E. and Goldman, R. (1993). A Bayesian model of plan recognition. *Artificial Intelligence*, 64(1):53–79.

Fudenberg, D. and Imhof, L. A. (2005). Imitation processes with small mutations. *Journal of Economic Theory*, 131:251–262.

Geib, C. W. and Goldman, R. P. (2009). A probabilistic plan recognition algorithm based on plan tree grammars. *Artificial Intelligence*, 173(2009):1101–1132.

Gutierrez, A., Campo, A., Santos, F. C., Monasterio-Huelin, and Dorigo, M. (2009). Social odometry: Imitation based odometry in collective robotics. *International Journal of Advanced Robotics Systems*, 2(6):129–136.

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

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

Heinze, C. (2003). *Modeling Intention Recognition for Intelligent Agent Systems*. PhD thesis, The University of Melbourne, Australia.

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

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). *PLoS Biol*, 3(3):e79.

Imhof, L. A., Fudenberg, D., and Nowak, M. A. (2005). Evolutionary cycles of cooperation and defection. *Proc Natl Acad Sci USA*, 102:10797–10800.

Imhof, L. A., Fudenberg, D., and Nowak, M. A. (2007). Tit-for-tat or win-stay, lose-shift? *Journal of Theoretical Biology*, 247(3):574 – 580.

Karlin, S. and Taylor, H. E. (1975). *A First Course in Stochastic Processes*. Academic Press, New York.

Kautz, H. and Allen, J. (1986). Generalized plan recognition. In Kehler, T. and Rosenschein, S., editors, *Proceedings of the Conference of the American Association of Artificial Intelligence (AAAI-86)*, pages 32–38. AAAI.

Kraus, S. (1997). Negotiation and cooperation in multi-agent environments. *Artificial Intelligence*, 94(1-2):79–98.

Lindgren, K. and Nordahl, M. G. (1994). Evolutionary dynamics of spatial games. *Physica D: Nonlinear Phenomena*, 75(1-3):292 – 309.

Macy, M. W. and Flache, A. (2002). Learning dynamics in social dilemmas. *Proc. Natl. Acad. Sci. USA*, 99:7229–7236.

Masuda, N. and Ohtsuki, H. (2009). A theoretical analysis of temporal difference learning intheiterated prisoners dilemma game. *Bulletin of Mathematical Biology*, 71:1818–1850.

Nakahara, K. and Miyashita, Y. (2005). Understanding intentions: Through the looking glass. *Science*, 308(5722):644–645.

Narendra, K. and Thathachar, M. (1989). *Learning Automata: An introduction*. Prentice-Hall International, Boston, USA.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805):1560. DOI: 10.1126/science.1133755.

Nowak, M. A. and Sigmund, K. (1992a). Tit for tat in heterogeneous populations. *Nature*, 355:250–253.

Nowak, M. A. and Sigmund, K. (1992b). Tit for tat in heterogeneous populations. *Nature*, 355:250–253.

Nowak, M. A. and Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit for tat in prisoners dilemma. *Nature*, 364:56–58.

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

Ohtsuki, H. and Iwasa, Y. (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, 239(4):435 – 444.

Pacheco, J. M., Santos, F. C., and Chalub, F. A. C. C. (2006). Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Comput. Biol.*, 2(12):e178.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, Palo Alto.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge U.P.

Pereira, L. M. and Han, T. A. (2011). Intention recognition with evolution prospection and causal Bayesian networks. In Madureira, A., Ferreira, J., and Vale, Z., editors, *Computational Intelligence for Engineering Systems: Emergent Applications*, volume 46, pages 1–33. Springer.

Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192.

Robson, A. (1990). Efficiency in evolutionary games: Darwin, Nash, and the secret handshake. *Journal of Theoretical Biology*, 144(3):379–396.

Roy, O. (2009). Intentions and interactive transformations of decision problems. *Synthese*, 169(2):335–349.

Santos, F. C. and Pacheco, J. M. (2005). Scale-free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. Lett.*, 95:098104.

Santos, F. C., Pacheco, J. M., and Lenaerts, T. (2006). Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proc. Natl. Acad. Sci. U.S.A.*, 103:3490–3494.

Santos, F. C., Pacheco, J. M., and Skyrms, B. (2011). Co-evolution of pre-play signaling and cooperation. *Journal of Theoretical Biology*, 274(1):30–35.

Santos, F. C., Santos, M. D., and Pacheco, J. M. (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature*, 454:214–216.

Sigmund, K. (2010). *The Calculus of Selfishness*. Princeton U. Press.

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford University Press.

Szabó, G. and Fáth, G. (2007). Evolutionary games on graphs. *Phys Rep*, 97-216(4-6).

Szabó, G. and Toke, C. (1998). Evolutionary prisoners dilemma game on a square lattice. *Phys Rev E*, 58:69–73.

Tahboub, K. A. (2006). Intelligent human-machine interaction based on dynamic Bayesian networks probabilistic intention recognition. *J. Intell. Robotics Syst.*, 45:31–52.

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Traulsen, A. and Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. *Proc. Natl. Acad. Sci. U.S.A.*, 103(29):10952.

Traulsen, A., Nowak, M. A., and Pacheco, J. M. (2006). Stochastic dynamics of invasion and fixation. *Phys. Rev. E*, 74:11909.

Traulsen, A., Pacheco, J. M., and Nowak, M. A. (2007). Pairwise comparison and selection temperature in evolutionary game dynamics. *J. Theor. Biol.*, 246:522–529.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quaterly Review of Biology*, 46:35–57.

van Hees, M. and Roy, O. (2008). Intentions and plans in decision and game theory. In Verbeek, B., editor, *Reasons and intentions*, pages 207–226. Ashgate Publishers, Dartmouth.

Van Segbroeck, S., de Jong, S., Nowé, A., Santos, F. C., and Lenaerts, T. (2010). Learning to coordinate in complex networks. *Adaptive Behavior*, 18:416–427.

Vukov, J., Santos, F. C., and Pacheco, J. M. (2011). Incipient Cognition Solves the Spatial Reciprocity Conundrum of Cooperation. *PLoS ONE*, 6(3):e17939.

Wang, S., Szalay, M. S., Zhang, C., and Csermely, P. (2008). Learning and innovative elements of strategy adoption rules expand cooperative network topologies. *PLoS ONE*, 3(4):e1917.

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

A Decision making with intention recognition

He we derive a simplified expression for intention recognizers' decision making, i.e., when to cooperate and when to defect. Recall that an *IR* cooperates iff he recognizes that the co-player is more likely to cooperate than to defect, i.e. $p(I = C|O = o) \geq p(I = D|O = o)$. From Eq. (5) (main text), we have that the inequality holds if an only if

$$
\begin{aligned}
& p(I = C, O = o) \geq p(I = D, O = o) \\
\iff & p(I = C, O = o, Tr = t) + p(I = C, O = o, Tr = f) \geq \\
& \qquad p(I = D, O = o, Tr = t) + p(I = D, O = o, Tr = f) \\
\iff & p(O = o|I = C)p(I = C|Tr = t)p(Tr = t) + \\
& p(O = o|I = C)p(I = C|Tr = f)p(Tr = f) \geq \\
& p(O = o|I = D)p(I = D|Tr = t)p(Tr = t) + \\
& p(O = o|I = D)p(I = D|Tr = f)p(Tr = f) \\
\iff & mtr\,[otr.h + (1 - otr)(1 - h)] \geq (1 - mtr)\,[otr(1 - h) + (1 - otr)h]
\end{aligned}
$$

where $otr = P(Tr = t)$ and $mtr = P(O = o|I = C)$.

Simplifying both sides we obtain

$$h(2.otr - 1) \geq otr - mtr \tag{9}$$

From Eqs. (3) and (4) (main text), (9) can be rewritten as

$$\Delta = sP + Q \geq 0 \tag{10}$$

where $s = 2h - 1$ $(0 \leq s \leq 1)$; $P = n_C(\mathcal{I}, \mathcal{J}) - n_D(\mathcal{I}, \mathcal{J})$ and $Q = n_C(\mathcal{J}, \mathcal{I}) - n_D(\mathcal{J}, \mathcal{I})$.

In short, (10) provides a simple decision making model for the intention recognizer, taking into account the co-player's M recent moves as well as the intention recognizer own's M recent moves, linking by a factor of $(2h-1)$ where h is the optimistic level of the intention recognizer.

We have that $-M \leq P, Q \leq M$. Let us look at some extreme cases.

- If $Q = M$, i.e. the co-player cooperates in all M recent steps, then $\Delta > 0$. The *IR* cooperates in the next round.
- If $Q = -M$, i.e. the co-player defects in all M recent steps, then $\Delta < 0$. The *IR* defects in the next round.
- If $h = 1/2$, then $\Delta = Q$. The *IR* only considers the co-player's past actions to decide his next move: if the co-player cooperated at least as much as defected in the M recent rounds, then *IR* cooperates in the next round, and defects otherwise. We henceforth consider $h > 1/2$: the *IR* also takes its own moves into account.

B Memory-Two Intention Recognizers

To provide a simple mathematical analysis, let us consider the simplest case where *IR* players has a very short memory $M = 2$. By following a similar method as described in (Sigmund, 2010), we derive the analytical payoff matrix for *AllC*, *AllD* and *IR* in the presence of noise. Based on that, we then compare analytically *IR* with other famous strategies, including *TFT* and *WSLS*.

To begin with, since $s > 0$, a memory-two *IR* decides his current move depending on the state of his own and his co-player's last two moves. There are 16 possible states, forming by all the combinations of four possible game situations (R, S, T, P) in each of the last two encounters. We enumerate these states by $state_i$, with $1 \le i \le 16$.

We consider stochastic strategies $(f, l, q_1, q_2, ..., q_{16}) \in [0, 1]^{18}$ where f and l are the propensities to play C in the initial and second rounds, respectively, and q_i are propensities to play C after having been at $state_i$, $1 \le i \le 16$.

Let us assume that player 1 using $(f_1, s_1, p_1, p_2, ..., p_{16})$ encounters a co-player 2 using $(f_2, s_2, q_1, q_2, ..., q_{16})$. We have a Markov chain in the state space $\{state_1, ..., state_{16}\}$. The transition probabilities are given by the stochastic matrix Q below. Note that one player's S is the other player's T.

$$
\mathrm{Q} = \begin{pmatrix}
p_1q_1 & ... & (1-p_1)(1-q_1) & 0 & ... & 0 & & 0 \\
0 & ... & 0 & p_2q_3 & ... & 0 & ... & 0 \\
... & & & ... & & & ... & \\
p_5q_9 & ... & (1-p_5)(1-q_9) & 0 & ... & 0 & & 0 \\
. & & & & & & & \\
. & & & & & & & \\
. & & & & & & & \\
0 & ... & 0 & 0 & .... & 0 & ... & (1-p_{16})(1-q_{16})
\end{pmatrix}
$$

There are several zeros in the matrix. The state with the second component $X \in \{R, S, T, P\}$ can only go to the states with the first component being X. For example, in the first row, $state_1$, i.e. (R, R), can only go to states with the first component being R, i.e. $(R, R), (R, T), (R, S), (R, P)$ (i.e. $state_i, 1 \le i \le 4$).

The initial probabilities for the sixteen states are given by the vector

$$
\begin{aligned}
\mathbf{f} =& \{f_1f_2s_1s_2,\ f_1f_2s_1(1-s_2),\ f_1f_2(1-s_1)s_2,\ f_1f_2(1-s_1)(1-s_2), \\
& f_1(1-f_2)s_1s_2,\ f_1(1-f_2)s_1(1-s_2),\ f_1(1-f_2)(1-s_1)s_2, \\
& f_1(1-f_2)(1-s_1)(1-s_2),\ (1-f_1)f_2s_1s_2,\ (1-f_1)f_2s_1(1-s_2), \\
& (1-f_1)f_2(1-s_1)s_2,\ (1-f_1)f_2(1-s_1)(1-s_2), \\
& (1-f_1)(1-f_2)s_1s_2,\ (1-f_1)(1-f_2)s_1(1-s_2), \\
& (1-f_1)(1-f_2)(1-s_1)s_2,\ (1-f_1)(1-f_2)(1-s_1)(1-s_2)\}
\end{aligned}
$$

In the next round, these probabilities are given by $\mathbf{f}Q$, and in the round n by $\mathbf{f}Q^n$. We denote by $\mathbf{g}$ the vector $\{R, S, T, P, R, S, T, P, R, S, T, P, R, S, T, P\}$, then the payoff for player 1 in round n is given by

$$A(n) = \mathbf{g} \ . \ \mathbf{f}Q^n \tag{11}$$

For $\omega < 1$ the average payoff per round is $(1-\omega)\sum w^n A(n)$ (Sigmund, 2010), i.e.,

$$(1-\omega)\mathbf{g} \ . \ \mathbf{f}(Id - \omega Q)^{-1} \tag{12}$$

where Id is the identity matrix of size 16.

B.1 Payoff Matrices in Presence of Noise

We now derive the payoff matrix for the 3-strategy game involving *AllC*, *AllD* and *IR* in the presence of noise, i.e., an intended action (C or D) can fail with probability $\epsilon \in [0,1]$. The strategies *AllC* and *AllD* are given by $\{1-\epsilon, 1-\epsilon, 1-\epsilon, ..., 1-\epsilon\}$, $\{\epsilon, \epsilon, \epsilon, ..., \epsilon\}$, respectively. Strategy *IR* is given by $\{1-\epsilon, l, q_1, ..., q_{16}\}$ where $l = 2\epsilon(1-\epsilon)$ when playing with *AllD* and $\epsilon^2 + (1-\epsilon)^2$ when playing with *AllC* or *IR* [3]; and $q_i = 1-\epsilon$ for $i \in \{1..5, 7, 9..11, 13\}$ and ϵ for $i \in \{6, 8, 12, 14..16\}$. Note that $X..Y$, where $X \leq Y$ are two natural numbers, denote the sequence $X, X+1, ..., Y$. Basically, *IR* will cooperate in the next round (with probability $1-\epsilon$) iff there is at least a mutual cooperation (i.e. R) in the last two steps (i.e. $i \in \{1..5, 13\}$) or there is at least one T and no P (i.e. $i \in \{7, 9..11\}$).

Considering the PD game with $T = b, R = b-c, P = 0, S = -c$, the payoff matrix for *AllC*, *AllD* and *IR*, applying Eq. (12) for each pair of strategies, is approximately given (where all terms of order $O(\epsilon^2)$ have been ignored),

$$\begin{pmatrix} (b-c)(1-\epsilon) & -c+(b+c)\epsilon & \Lambda \\ b-b\epsilon-c\epsilon & (b-c)\epsilon & \frac{\epsilon\left(b\left(2+\omega-\omega^2\right)-c\right)}{1-\epsilon\omega} \\ \frac{b\left(1-\epsilon-\epsilon\omega^2\right)-c\left(1-\epsilon\left(2-\omega+\omega^2\right)\right)}{1-\epsilon\omega^2} & \frac{\epsilon\left(b-c\left(2+\omega-\omega^2-\omega^3\right)\right)}{1-\epsilon\omega^2(1+\omega)} & \frac{(b-c)\left(1+\epsilon\left(-2+4\omega+\omega^2+\omega^3\right)\right)}{1+\epsilon\omega(3+\omega+\omega^2)} \end{pmatrix}$$

where $\Lambda = \frac{c\left(\omega-1+\epsilon\left(1-4\omega+\omega^2\right)\right)+b\left(\omega-1+2\epsilon\left(1-3\omega+\omega^2\right)\right)}{1-\omega+3\epsilon\omega-\epsilon\omega^2}$ [4].

By a similar method, we derive the payoff matrixes *AllC*, *AllD* and either *TFT* or *WSLS* (the results for the general case of Prisoner's Dilemma can be found in (Imhof et al., 2007)). First, for *AllC*, *AllD* and *WSLS*

$$\begin{pmatrix} (b-c)(1-\epsilon) & b\epsilon-c(1-\epsilon) & c(-1+\epsilon)-\frac{b(1-\epsilon)(1-(1-2\epsilon)\omega)}{-1+(1-2\epsilon)^2\omega} \\ b(1-\epsilon)-c\epsilon & (b-c)\epsilon & -c\epsilon+\frac{b\left(1+2\epsilon^2\omega-\epsilon(1+\omega)\right)}{1+(1-2\epsilon)^2\omega} \\ \Lambda' & \frac{b\epsilon\left(1+(1-2\epsilon)^2\omega\right)-c\left(1-\epsilon-\epsilon\omega+2\epsilon^2\omega\right)}{1+(1-2\epsilon)^2\omega} & (b-c)(1-\epsilon)\left(1-2\epsilon\omega+4\epsilon^2\omega\right) \end{pmatrix}$$

where $\Lambda' = b(1-\epsilon) + \frac{c(1-\epsilon)(1-(1-2\epsilon)\omega)}{-1+(1-2\epsilon)^2\omega}$.

Now, for *AllC*, *AllD* and *TFT*, the payoff matrix reads

$$\begin{pmatrix} (b-c)(1-\epsilon) & b\epsilon-c(1-\epsilon) & c(-1+\epsilon)+b\left(1+2\epsilon^2\omega-\epsilon(1+\omega)\right) \\ b(1-\epsilon)-c\epsilon & (b-c)\epsilon & -c\epsilon-b(-1+\epsilon)(1+(-1+2\epsilon)\omega) \\ \Lambda'' & b\epsilon-c(1-\epsilon)(1-(1-2\epsilon)\omega) & \frac{(b-c)(1-\omega+\epsilon(-1+2\omega))}{1+(-1+2\epsilon)\omega} \end{pmatrix}$$

where $\Lambda'' = b(1-\epsilon) - c\left(1-\epsilon-\epsilon\omega+2\epsilon^2\omega\right)$.

B.2 Comparing IR, WSLS and TFT

Let $A(X,Y)$ be the payoff of strategist X when playing with strategist Y (from the payoff matrices). In the sequel we will show that, for $b \geq (1+\omega)c$, $\omega >$

[3] These can be easily seen from the fact that in the second round *IR*'s decision only depends on what the co-player did in the first round (cf. (10)).

[4] This notation is added only for a better alignment.

3/4 and $\epsilon < 1/3$:

$$A(IR, IR) > A(WSLS, WSLS) > A(TFT, TFT)$$
$$A(AllD, IR) < A(AllD, TFT) < A(AllD, WSLS)$$
$$A(IR, AllD) > A(TFT, AllD) > A(WSLS, AllD)$$
$$A(AllC, WSLS) < A(AllC, TFT) < A(AllC, IR)$$
$$A(WSLS, AllC) > A(TFT, AllC) > A(IR, AllC)$$

The first condition always holds by the usual assumption in Donation game $b \geq 2c$; the second condition means the game is repeated at least 4 rounds.

B.2.1 TFT, WSLS, IR: Amongst themselves

Comparing the bottom rightmost element of each payoff matrix we have

$$A(WSLS, WSLS) - A(TFT, TFT) =$$

$$\frac{(b-c)(1-2\epsilon)^2\epsilon\omega(2\omega - 2\epsilon\omega - 1)}{1+(-1+2\epsilon)\omega} > 0$$

(for small enough ϵ and big enough ω, namely $\epsilon < 1 - \frac{1}{2\omega}$)

$$A(IR, IR) - A(WSLS, WSLS) =$$

$$\frac{(b-c)\epsilon\left(-1+3\omega+\epsilon\left(-3\omega+7\omega^2+3\omega^3+2\omega^4\right)\right)}{1+\epsilon\omega\left(3+\omega+\omega^2\right)} > 0$$

(all the terms of order $O(\epsilon^3)$ in the numerator have been ignored)
In short, for $\epsilon < 1 - \frac{1}{2\omega}$ we have

$$A(IR, IR) > A(WSLS, WSLS) > A(TFT, TFT)$$

B.2.2 TFT, WSLS, IR: With AllD

$$A(AllD, WSLS) - A(AllD, TFT) = \frac{b(1-2\epsilon)^2\omega\left(\epsilon+\omega-3\epsilon\omega+2\epsilon^2\omega\right)}{1+(1-2\epsilon)^2\omega} > 0$$
$$A(AllD, TFT) - A(AllD, IR) = \frac{b(1-\omega)(1-3\epsilon-2\epsilon\omega)}{1-\epsilon\omega} > 0$$

(all the terms of order $O(\epsilon^2)$ in the numerator have been ignored)

$$A(TFT, AllD) - A(WSLS, AllD) = \frac{c(1-2\epsilon)^2\omega\left(\epsilon+\omega-3\epsilon\omega+2\epsilon^2\omega\right)}{1+(1-2\epsilon)^2\omega} > 0$$

$$A(IR, AllD) - A(TFT, AllD) = \frac{c\epsilon\omega^3 + c(1-\omega)\left(1-3\epsilon-\epsilon\omega-\epsilon\omega^2\right)}{1-\epsilon\omega^2(1+\omega)} > 0$$

(all the terms of order $O(\epsilon^2)$ in the numerator have been ignored)
In short, it always holds that

$$A(AllD, WSLS) > A(AllD, TFT) > A(AllD, IR)$$
$$A(IR, AllD) > A(TFT, AllD) > A(WSLS, AllD)$$

B.2.3 TFT, WSLS, IR: With AllC

$A(AllC, WSLS) - A(AllC, TFT) = \frac{-b(1-2\epsilon)^2\epsilon\omega(1+(1-2\epsilon)\omega)}{1-(1-2\epsilon)^2\omega} < 0$

$A(AllC, TFT) - A(AllC, IR) =$

$$\frac{-\epsilon\left(b(1-\omega)(2\omega-1)+\epsilon\omega\left(b-b\omega^2+4b\omega-3c+c\omega\right)\right)}{1-\omega+3\epsilon\omega-\epsilon\omega^2} < 0$$

(all the terms of order $O(\epsilon^2)$ in the numerator have been ignored; $\omega > 3/4$).

$A(WSLS, AllC) - A(TFT, AllC) = \frac{c(1-2\epsilon)^2\epsilon\omega(1+\omega-2\epsilon\omega)}{1-(1-2\epsilon)^2\omega} > 0$

$A(TFT, AllC) - A(IR, AllC) =$

$$\frac{\epsilon\left(2c\epsilon^2\omega^3+c(2\omega(1-\epsilon)-1)+\epsilon\left(b\omega^2-c\omega^2-c\omega^3\right)\right)}{1-\epsilon\omega^2} > 0$$

(for big enough ω and small enough ϵ, namely $\epsilon < 1-\frac{1}{2\omega}$, and $b \geq (1+\omega)c$)
In short, for $b \geq (1+\omega)c,\ \omega > 3/4;$ and $\epsilon < 1/3$ we have

$$A(AllC, WSLS) < A(AllC, TFT) < A(AllC, IR)$$

$$A(WSLS, AllC) > A(TFT, AllC) > A(IR, AllC)$$