AI Communications 00 (20xx) 1–7
DOI 10.3233/AIC-150672
IOS Press

Emergence of cooperation via intention recognition, commitment and apology – A research summary

The Anh Han [a,b,c], Luís Moniz Pereira [d,*], Francisco C. Santos [e,f] and Tom Lenaerts [a,b]
[a] *AI-lab, Vrije Universiteit Brussel, Brussels, Belgium*
E-mails: T.Han@tees.ac.uk, Tom.Lenaerts@ulb.ac.be
[b] *MLG, Département d'Informatique, Université Libre de Bruxelles, Brussels, Belgium*
[c] *School of Computing, Teesside University, Middlesbrough, UK*
[d] *NOVA Laboratory for Computer Science and Informatics (NOVA LINCS), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal*
E-mail: lmp@fct.unl.pt
[e] *INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal*
[f] *ATP-group, CMAF, Instituto para a Investigação Interdisciplinar, Lisbon, Portugal*
E-mail: franciscocsantos@ist.utl.pt

Abstract. The mechanisms of emergence and evolution of cooperation in populations of abstract individuals, with diverse behavioral strategies in co-presence, have been undergoing mathematical study via evolutionary game theory, inspired in part on evolutionary psychology. Their systematic study resorts to simulation techniques, thus enabling the study of aforesaid mechanisms under a variety of conditions, parameters and alternative virtual games. The theoretical and experimental results have continually been surprising, rewarding and promising. In our recent work, we initiated the introduction, in such groups of individuals, of cognitive abilities inspired on techniques and theories of Artificial Intelligence, namely those pertaining to Intention Recognition, Commitment and Apology (separately and jointly), encompassing errors in decision-making and communication noise. As a result, both the emergence and stability of cooperation become reinforced comparatively to the absence of such cognitive abilities. This holds separately for Intention Recognition, for Commitment and for Apology, and even more so when they are jointly engaged.

Our presentation aims to sensitize the reader to these evolutionary game theory based issues, results and prospects, which are accruing in importance for the modeling of minds with machines, with impact on our understanding of the evolution of mutual tolerance and cooperation. Recognition of someone's intentions, which may include imagining the recognition others have of our own intentions, and may comprise not just some error tolerance, but also a penalty for unfulfilled commitment though allowing for apology, can lead to evolutionary stable win/win equilibriums within groups of individuals, and perhaps amongst groups. The recognition and the manifestation of intentions, plus the assumption of commitment – even whilst paying a cost for putting it in place – and the acceptance of apology, are all facilitators in that respect, each of them singly and, above all, in collusion.

Keywords: Intention recognition, commitments, evolution of cooperation, apology, multi-agent systems, evolutionary game theory

1. Introduction

In collective strategic interaction, wherein multiple agents pursue individual strategies, conflicts will arise because the actions of individual agents may have an effect on the welfare of others and on their own in return. Hence, in these situations the need arises for the regulation of individual and collective behavior, traditionally having followed two distinct approaches, well-known in the Economics and Artificial Intelligence literature [3,17,31,36,38,40,41,44,51,53]: the spontaneous emergence of order approach, which studies how

*Corresponding author: Luís Moniz Pereira, NOVA Laboratory for Computer Science and Informatics, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal. E-mail: lmp@fct.unl.pt.

norms result from endogenous agreements among rational individuals, and the mechanism by design approach, which studies how norms are exogenously imposed in order to attain desirable properties of the whole.

In this short research summary, we describe the main results we obtained following essentially the former approach, but crucially complementing it in instilling some individual agents with cognitive abilities that can and will induce cooperation in the population. These abilities enable such individuals to recognize the opportunity whether to decide to cooperate outright, or possibly propose costly cooperation commitments, susceptible to compensation on defaulting, and to accept apology redressing dues. In consequence, cooperation can evolve and emerge.

The problem of evolution of cooperation and of the emergence of collective action – cutting across areas as diverse as Biology, Economy, Artificial Intelligence, Political Science or Psychology – is one of the greatest interdisciplinary challenges science faces today [2,32,45,60]. To understand the evolutionary mechanisms that promote and keep cooperative behaviour is all the more complex as increasingly intricate is the intrinsic complexity of the partaking individuals.

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 [46,60]. From kin and group ties, to different forms of reciprocity and networked populations, several aspects have been shown to play an important role in the emergence of cooperation (see survey in [46,60]).

Moreover, more complex strategies based on the evaluation of interactions between third parties allow the emergence of kinds of cooperation that are immune to exploitation because then interactions are channelled to just those who cooperate. Questions of justice and trust, with their negative (punishment) and positive (help) incentives, are fundamental in games with large diversified groups of individuals gifted with intention recognition capabilities. In allowing them to choose amongst distinct behaviours based on suggestive information about the intentions of their interaction partners – these in turn influenced by the behaviour of the individual himself – individuals are also influenced by their tolerance to error or noise in the communication. One hopes that, to start with, understanding these capabilities can be transformed into mechanisms for spontaneous organization and control of swarms of autonomous robotic agents [4], these being envisaged as large populations of agents where cooperation can emerge, but not necessarily to solve a priori given goals, as in distributed Artificial Intelligence (AI).

With these general objectives, we have specifically studied the way players' strategies adapt in populations involved in cooperation games. We used the techniques of Evolutionary Game Theory (EGT) [35,60], considered games such as the Prisoner's Dilemma and Public Goods Game [35,60], and showed how the actors participating in repeated iterations in these games can benefit from having the ability to recognize the intentions of other actors, to apologize when making mistakes, or to establish commitments, or to combine some of them, thereby leading to an evolutionary stable increase in cooperation [18,19,24,26–28], compared to extant best strategies.

In this paper, we summarize our recent publications on how intention recognition, commitment arrangement and apology can, separately and jointly, lead to the evolution of high levels of cooperation. We discuss how these works provide useful insights for mechanism design in multi-agent systems for regulative purposes. Evolutionary emergent futures are what we have studied, tied to the co-presence of fixed strategies in agents, though an agent may replace its strategy by a more advantageous one on occasion (social learning). We have not yet made a strategy also evolve by adopting features of other strategies into its own, through rule-defined strategies updating, which could be a direction for multi-agent systems (MAS).

2. Intention recognition promotes the evolution of cooperation

The ability of recognizing (or reading) intentions of others has been observed and shown to play an important role in many cooperative interactions, both in humans and primates [39,52,61]. However, most studies on the evolution of cooperation, grounded on evolutionary dynamics and game theory, have neglected the important role played by a basic form of intention recognition in behavioral evolution. In our work

[24,26], we have addressed explicitly this issue, characterizing the dynamics emerging from a population of intention recognizers.

In that work, intention recognition (IR) was implemented using Bayesian Networks (BN) [24,49,50], taking into account the information of current signals of intent, as well as the mutual trust and tolerance accumulated from previous one-on-one play experience – including how my previous defections may influence another's intent – but without resorting to information gathered regarding players' overall reputation in the population.

A player's present intent can be understood here as how he's going to play the next round with me, whether by cooperating or defecting [24]. Intention recognition can also be learnt from a corpus of prior interactions among game strategies [25,26], where each strategy can be envisaged and detected as players' (possibly changing) intent to behave in a certain way [20,21]. In both cases, we experimented with populations with different proportions of diverse strategies in order to calculate, in particular, what is the minimum fraction of individuals capable of intention recognition for cooperation to emerge, invade, prevail and persist.

It is noteworthy that the notions of intentions used in our two models have been specialized for the concrete game-theoretical contexts in place. A more general definition, e.g. as described in Bratman's seminal work [5–7], can accommodate for intention changes or even abandonment. For instance, a player can change his intention or strategy before the next interaction or round of game takes place. That aspect is generally not considered in EGT models, as players can change their strategy only at the end of a generation [60]. However, we envisage a convenient extension to that direction since our intention recognition methods – performed through Bayesian Network inference techniques, as described in [21] – can cope with intention changes and abandonment; even more, they have been tested with a benchmark generated in the EGT context itself [20]. Further discussion of the intention recognition models in the EGT context and possible extensions can be found in [19, Chapter 6].

Intention recognition techniques have been studied actively in AI for several decades [10,23,56], with several applications such as for improving human–computer interactions, assistive living, moral reasoning and team work [22,30,34,37,50,55]. In most of these applications the agents engage in repeated interactions with each other. Our results, both analytically and through extensive agent-based simulations, suggest that equipping the agents with an ability to recognize intentions of others can improve their cooperation and reduce misunderstanding that can result from noise and mistakes. The simulations are carried out in the course of the repeated Prisoner's Dilemma, in the presence of noise, and for varying configurations of the game as well as for a large number of runs. For further details of the simulation setups see [24–26].

3. Commitments promote the emergence of cooperation

Agents make commitments towards others when they give up options in order to influence others. Most commitments depend on some incentive that is necessary to ensure that an action (or even an intention) is in the agent's interest and thus will be carried out in the future [16]. Asking for prior commitments can just be used as a strategy to clarify the intentions of others, whilst at the same time manifesting our own. All parties then clearly know to what they commit and can refuse such a commitment whenever the offer is made. A classical example of such an agreement is marriage. In that case mutual commitment ensures some stability in the relationship, reducing the fear of exploitation and providing security against potential cataclysms.

In our recent works [18,27] we investigate analytically and numerically whether costly commitment strategies, in which players propose, initiate and honor a deal, are viable strategies for the evolution of cooperative behavior, using the symmetric one-shot Prisoner's Dilemma (PD) game to model a social dilemma. Next to the traditional cooperate (C) and defect (D) options, a player can propose its co-player to commit to cooperation before playing the PD game, willing to pay a personal cost to make the proposal credible. If the co-player accepts the arrangement and also plays C, they both receive their rewards for mutual cooperation. Yet if the co-player plays D, then he or she will have to provide the proposer with a compensation at a personal cost. The cost of compensation is agreed upon in advance, and is assumed to be enforceable, for instance, through the deposit-refund scheme [11] or external parties [42], although our work allows partial enforcement too [18]. Finally, when the co-player does not accept the deal, the game is not played and hence both obtain no payoff. Several free-riding strategies were included in the model, including (i) the fake committers, who accept a commitment proposal yet defect when playing the game, assuming that they can exploit the proposers without suffering a too severe consequence; and (ii) the commitment free-riders, who defect unless being proposed a commitment, which they then accept and cooperate afterwards in the PD game. In other words, these latter players are willing to cooperate when a commitment is proposed but are not prepared to pay the cost of setting it up.

Resorting the EGT mathematical analysis and simulations, we have shown that when the cost of arranging a commitment is justified with respect to the benefit of cooperation, substantial levels of cooperation can be achieved, especially when one insists on sharing the arrangement cost [18]. On the one hand, such commitment proposers can get rid of fake committers by proposing a strong enough compensation cost. On the other hand, they can maintain a sufficient advantage over the commitment free-riders, because a commitment proposer will cooperate with players alike herself, while the latter defect among themselves. We have also compared the commitment strategy with the simple costly punishment strategy – an important one for the evolution of cooperation [13] – where no prior agreements are made. The results show that the first strategy leads to a higher level of cooperation than the latter one. For further details of the simulation setups and mathematical analysis see [18,27].

Note that in our work we specify commitments in the context of one-shot interactions. When considering repeated games, such as the repeated PD, we might extend the specification by having conditional commitments, such as conditional promises and threats [57]. It would be interesting to ask in this context whether a long-term relationship is better sustained through direct reciprocity or through proposing long-term commitments, and how those two mechanisms interact.

There has been an extensive literature of AI and MAS research on commitment, e.g., [8,9,57,64,65], but the main concern therein is how to formalize different aspects of commitment and how a commitment mechanism can be implemented in multi-agent interactions to enhance these (e.g. for improved collaborative problem solving [65]), especially in the context of (non-evolutionary) game theory. Our work would provide insights into the design of MAS that rely on commitments or punishment in order to incentivize cooperation among agents, see for example [33].

4. Economical use of costly commitment via intention recognition

Commitments have been shown to promote cooperation if the cost of arranging them is justified with respect to the benefit of cooperation. But commitment may be quite costly, which leads to the possible prevalence of commitment free-riders [18]. Hence, it should be avoided when necessary. On the other hand, there are many cases where it is difficult to recognize the intention of another agent with sufficient confidence to make any decision based on it. One may have insufficient information for making the prediction (not enough actions being observed, such as in the first interaction scenario), or even one may know the agent well, but also know that the agent is very unpredictable. In such cases, the strategy of proposing a commitment, or manifesting an intention, can help to impose or clarify intentions of others. In addition, intention is usually defined as choice with commitment [5,12,54]. That is, once the agent intends to do something, it must settle on some state of affairs for which to aim, because of its resource limitation and in order to coordinate its future actions. Deciding what to do establishes a personal form of commitment [12,54]. Proposing a commitment deal to another agent consists in asking it to express or clarify its intended decisions.

In a marriage commitment, by giving up the option to leave the other, spouses gain security and an opportunity for a much deeper relationship that would be impossible otherwise [15,43], as it might be risky to assume a partner's intention of staying faithful without the commitment of marriage. A contract is another popular kind of commitment, e.g. for an apartment lease [15]. When it is risky to assume another agent's intention of being cooperative, arranging an appropriate contract provides incentives for cooperation. However, for example in accommodation rental, a contract is not necessary when the cooperative intention is of high certainty, e.g. when the business affair is between close friends or relatives. It said arranging a commitment deal can be useful to encourage cooperation whenever intention recognition is difficult, or cannot be performed with sufficiently high certainty. On the other hand, arranging commitments is not free, and requires a specific capacity to set it up within a reasonable cost (for the agent to actually benefit from it) [42,43] – therefore it should be avoided when opportune to do so.

With such motivations in mind, in our work [19,28] we showed that if the player first predicts the intentions of a co-player and proposes commitment only when they are not confident about their intention prediction, it can significantly facilitate the conditions for cooperation to emerge. The improvement (in level of cooperation) is most significant when it is costly to arrange commitments and when the cooperation is highly beneficial. The results were obtained using agent-based simulations within the context of the one-shot PD.

In short, it seems to us that intention recognition, and its use in the scope of commitment, is a foundational cornerstone where we should begin at, naturally followed by the capacity to establish and honor commitments, as a tool towards the successive construction of collective intentions and social organization [58,59]. Finally, one hopes that understanding these capabilities can be useful in the design of efficient self-organized and distributed engineering applications [4], from bio- and socio-inspired computational algorithms, to swarms of autonomous robotic agents.

5. Apology in committed vs. commitment-free repeated interactions

Apology is perhaps the most powerful and ubiquitous mechanism for conflict resolution [1,14,48], especially among individuals involving in long-term repeated interactions (such as a marriage). An apology can resolve a conflict without having to involve external parties (e.g. teachers, parents, courts), which may cost all sides of the conflict significantly more. Evidence supporting the usefulness of apology abounds, ranging from medical error situations to seller–customer relationships [1]. Apology has been implemented in several computerized systems such as human–computer interaction and online markets so as to facilitate users' positive emotions and cooperation [62,63].

The iterated Prisoner's Dilemma (IPD) has been the standard model to investigate conflict resolution and the problem of the evolution of cooperation in repeated interaction settings [2,60]. This IPD game is usually known as a story of tit-for-tat (TFT), which won both Axelrod's tournaments [2]. TFT cooperates if the opponent cooperated in the previous round, and defects if the opponent defected. But if there can be erroneous moves due to noise (i.e. an intended move is wrongly performed), the performance of TFT declines, because an erroneous defection by one player leads to a sequence of unilateral cooperation and defection. A generous version of TFT, which sometimes cooperates even if the opponent defected [47], can deal with noise better, yet not thoroughly. For these TFT-like strategies, apology is modeled implicitly as one or more cooperative acts after a wrongful defection.

In our recent work [29], we describe a model containing strategies that explicitly apologize when making an error between rounds. An apologizing act consists in compensating the co-player an appropriate amount (the higher the more sincere), in order to ensure that this other player cooperates in the next actual round. As such, a population consisting of only apologizers can maintain perfect cooperation. However, other behaviors that exploit such apology behavior could emerge, such as those that accept apology compensation from others but do not apologize when making mistakes (fake apologizers), destroying any benefit of the apology behavior. To study which behaviour emerges from this game, we consider a population of different possible strategies, including those that apologize when making mistakes and those which are not willing to do so. Also, we have strategies that accept apology (hence, compensation) from others and those which do not. Resorting to EGT methods [60], we show that when the apology occurs in a system where the players first ask for a commitment before engaging in the interaction [18,19,27,28], this exploitation can be avoided. Our results lead to the following conclusions, having been validated both analytically and through extensive agent-based simulations: (i) Apology alone is insufficient to achieve high levels of cooperation; (ii) Apology supported by prior commitment leads to significantly higher levels of cooperation; (iii) Apology needs to be sincere to function properly, whether in a committed relationships or commitment-free ones (which is in accordance with existing experimental studies, e.g. in [48]); (iv) A much costlier apology tends to be used in committed relationships than in commitment-free ones, as it can help better identify free-riders such as fake apologizers: '*commitments bring about sincerity*'.

As apology [62,63] and commitment [64,65] have been widely studied in AI and Computer Science, for example, about how these mechanisms can be formalized, implemented, and used to enhance cooperation in human–computer interactions and online market systems [62,63], as well as general multi-agent systems [64,65], our study would provide important insights for the design and deployment of such mechanisms; for instance, what kind of apology should be provided to customers when making mistakes, and whether apology can be enhanced when complemented with commitments to ensure better cooperation, e.g. compensation from customer's for wrongdoing.

6. Conclusions

We have argued that the study of the aforementioned issues has come of age and is ripe with research opportunities, having communicated some of the inroads we explored, and pointed to the more detailed published results of what we have achieved, with respect to intention recognition, commitment and mutual tolerance through apology, within the overarching evolutionary game theory context.

Acknowledgements

TAH and TL acknowledges the support from the FWO Vlaanderen (Postdoctoral fellowship). TL also acknowledges the support of the *Fonds de la Recherche Scientifique* – FNRS. FCS acknowledges the support of FCT-Portugal through projects EXPL/EEI-SII/2556/2013 and PEst-OE/EEI/LA0021/2013.

References

[1] J. Abeler, J. Calaki, K. Andree and C. Basek, The power of apology, *Economics Letters* **107**(2) (2010), 233–235.
[2] R. Axelrod, *The Evolution of Cooperation*, Basic Books, New York, 1984, ISBN 0-465-02122-2.
[3] R. Axelrod, An evolutionary approach to norms, *American Political Science Review* **80**(4) (1986), 1095–1111.
[4] E. Bonabeau, M. Dorigo and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford Univ. Press, New York, NY, 1999.
[5] M.E. Bratman, *Intention, Plans, and Practical Reason*, The David Hume Series, CSLI, 1987.
[6] M.E. Bratman, Planning and the stability of intention, *Minds and Machines* **2**(1) (1992), 1–16.
[7] M.E. Bratman, D.J. Israel and M.E. Pollack, Plans and resource-bounded practical reasoning, *Computational Intelligence* **4**(3) (1988), 349–355.
[8] C. Castelfranchi and R. Falcone, Principles of trust for MAS: Cognitive anatomy, social importance, and quantification, in: *Proceedings of International Conference on Multi Agent Systems*, IEEE, 1998, pp. 72–79.
[9] C. Castelfranchi and R. Falcone, *Trust Theory: A Socio-Cognitive and Computational Model*, Wiley Series in Agent Technology, Wiley, 2010.
[10] E. Charniak and R.P. Goldman, A Bayesian model of plan recognition, *Artificial Intelligence* **64**(1) (1993), 53–79.
[11] T.L. Cherry and D.M. McEvoy, Enforcing compliance with environmental agreements in the absence of strong institutions: An experimental analysis, *Environmental and Resource Economics* **54**(1) (2013), 63–77.
[12] P.R. Cohen and H.J. Levesque, Intention is choice with commitment, *Artificial Intelligence* **42**(2,3) (1990), 213–261.
[13] E. Fehr and S. Gachter, Altruistic punishment in humans, *Nature* **415** (2002), 137–140.
[14] U. Fischbacher and V. Utikal, On the acceptance of apologies, *Games and Economic Behavior* **82** (2013), 592–608.
[15] R.H. Frank, Cooperation through emotional commitment, in: *Evolution and the Capacity for Commitment*, R.M. Nesse, ed., Russell Sage, New York, 2001, pp. 55–76.
[16] H. Gintis, Beyond selfishness in modeling human behavior, in: *Evolution and the Capacity for Commitment*, R.M. Nesse, ed., Russell Sage, New York, 2001.
[17] T. Groves, Incentives in teams, *Econometrica* **41**(4) (1973), 617–631.
[18] T. Han, L. Pereira, F. Santos and T. Lenaerts, Good agreements make good friends, *Scientific Reports* **3** (2013), 2695.
[19] T.A. Han, *Intention Recognition, Commitments and Their Roles in the Evolution of Cooperation: From Artificial Intelligence Techniques to Evolutionary Game Theory Models*, SAPERE, Vol. 9, Springer, 2013.
[20] T.A. Han and L.M. Pereira, Context-dependent incremental intention recognition through Bayesian network model construction, in: *Proceedings of the Eighth UAI Bayesian Modeling Applications Workshop (UAI-AW 2011)*, A. Nicholson, ed., Vol. 818, CEUR Workshop Proceedings, 2011, pp. 50–58.
[21] T.A. Han and L.M. Pereira, Context-dependent incremental decision making scrutinizing the intentions of others via Bayesian network model construction, *Intelligent Decision Technologies* **7**(4) (2013), 293–317.
[22] T.A. Han and L.M. Pereira, Intention-based decision making and its applications, in: *Human Behavior Recognition Technologies: Intelligent Applications for Monitoring and Security*, H. Guesgen and S. Marsland, eds, IGI Global, 2013, pp. 174–211 (Chapter 9).
[23] T.A. Han and L.M. Pereira, State-of-the-art of intention recognition and its use in decision making – A research summary, *AI Communication Journal* **26**(2) (2013), 237–246.
[24] T.A. Han, L.M. Pereira and F.C. Santos, Intention recognition promotes the emergence of cooperation, *Adaptive Behavior* **19**(3) (2011), 264–279.
[25] T.A. Han, L.M. Pereira and F.C. Santos, The role of intention recognition in the evolution of cooperative behavior, in: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'2011)*, T. Walsh, ed., AAAI, 2011, pp. 1684–1689.
[26] T.A. Han, L.M. Pereira and F.C. Santos, Corpus-based intention recognition in cooperation dilemmas, *Artificial Life Journal* **18**(4) (2012), 365–383.
[27] T.A. Han, L.M. Pereira and F.C. Santos, The emergence of commitments and cooperation, in: *AAMAS'2012*, 2012, pp. 559–566.
[28] T.A. Han, L.M. Pereira and F.C. Santos, Intention recognition, commitment, and the evolution of cooperation, in: *Proceedings of IEEE Congress on Evolutionary Computation*, IEEE, 2012, pp. 1–8.
[29] T.A. Han, L.M. Pereira, F.C. Santos and T. Lenearts, Why is it so hard to say sorry: The evolution of apology with commitments in the iterated Prisoner's Dilemma, in: *IJCAI'2013*, 2013, pp. 177–183.
[30] T.A. Han, A. Saptawijaya and L.M. Pereira, Moral reasoning under uncertainty, in: *Proceedings of the 18th International Conference on Logic for Programming, Artificial Intelligence and Reasoning (LPAR-18)*, LNCS, Vol. 7180, Springer, 2012, pp. 212–227.
[31] T.A. Han, L. Tran-Thanh and N.R. Jennings, The cost of interference in evolving systems, in: *The 17th International Workshop on Coordination, Organisations, Institutions and Norms (COIN 2014)*, 2014.
[32] G. Hardin, The tragedy of the commons, *Science* **162** (1968), 1243–1248.

[33] M.R. Hasan and A. Raja, Emergence of cooperation using commitments and complex network dynamics, in: *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Vol. 2, IEEE, 2013, pp. 345–352.

[34] C. Heinze, Modeling intention recognition for intelligent agent systems, PhD thesis, The University of Melbourne, Australia, 2003.

[35] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge Univ. Press, 1998.

[36] M.O. Jackson, Mechanism theory, in: *The Encyclopedia of Life Support Systems*, 2000.

[37] N. Lesh, Scalable and adaptive goal recognition, PhD thesis, University of Washington, 1998.

[38] R.P. McAfee, Mechanism design by competing sellers, *Econometrica* **61**(6) (1993), 1281–1312.

[39] A.N. Meltzoff, Imitation and other minds: The "like me" hypothesis, in: *Perspectives on Imitation: From Neuroscience to Social Science. Imitation, Human Development, and Culture*, MIT Press, Cambridge, MA, 2005, pp. 55–77.

[40] R. Myerson, Incentive compatibility and the bargaining problem, *Econometrica* **47**(1) (1979), 61–73.

[41] M. Naor, B. Pinkas and R. Sumner, Privacy preserving auctions and mechanism design, in: *Proceedings of the 1st ACM Conference on Electronic Commerce*, ACM, 1999, pp. 129–139.

[42] R.M. Nesse, *Evolution and the Capacity for Commitment*, Russell Sage Foundation Series on Trust, Russell Sage, 2001.

[43] R.M. Nesse, Natural selection and the capacity for subjective commitment, in: *Evolution and the Capacity for Commitment*, R.M. Nesse, ed., Russell Sage, New York, 2001, pp. 1–44.

[44] N. Nisan and A. Ronen, Algorithmic mechanism design, in: *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, ACM, 1999, pp. 129–140.

[45] M.A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life*, Harvard Univ. Press, Cambridge, MA, 2006.

[46] M.A. Nowak, Five rules for the evolution of cooperation, *Science* **314**(5805) (2006), 1560–1563. doi:10.1126/science.1133755.

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

[48] Y. Ohtsubo and E. Watanabe, Do sincere apologies need to be costly? Test of a costly signaling model of apology, *Evolution and Human Behavior* **30**(2) (2009), 114–123.

[49] L.M. Pereira and T.A. Han, Intention recognition via causal Bayes networks plus plan generation, in: *Progress in Artificial Intelligence, Proceedings of 14th Portuguese International Conference on Artificial Intelligence (EPIA'09)*, LNAI, Vol. 5816, Springer, 2009, pp. 138–149.

[50] L.M. Pereira and T.A. Han, Intention recognition with evolution prospection and causal Bayesian networks, in: *Computational Intelligence for Engineering Systems 3: Emergent Applications*, Springer, 2011, pp. 1–33.

[51] S. Phelps, P. McBurney and S. Parsons, Evolutionary mechanism design: A review, *Autonomous Agents and Multi-Agent Systems* **21**(2) (2010), 237–264.

[52] D.G. Rand, D. Fudenberg and A. Dreber, It's the thought that counts: The role of intentions in reciprocal altruism, 2013, available at: SSRN 2259407.

[53] D. Ross, *Economic Theory and Cognitive Science: Microexplanation*, MIT Press, 2005.

[54] O. Roy, Thinking before acting: Intentions, logic, rational choice, PhD thesis, ILLC Dissertation Series DS-2008-03, Amsterdam, 2009.

[55] P. Roy, B. Bouchard, A. Bouzouane and S. Giroux, A hybrid plan recognition model for Alzheimer's patients: Interleaved-erroneous dilemma, in: *Proceedings of IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, 2007, pp. 131–137.

[56] F. Sadri, Logic-based approaches to intention recognition, in: *Handbook of Research on Ambient Intelligence: Trends and Perspectives*, 2011, pp. 346–375.

[57] T.C. Schelling, *The Strategy of Conflict*, Oxford Univ. Press, London, 1990.

[58] J.R. Searle, *The Construction of Social Reality*, The Free Press, New York, 1995.

[59] J.R. Searle, *Making the Social World: The Structure of Human Civilization*, Oxford Univ. Press, 2010.

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

[61] M. Tomasello, *Origins of Human Communication*, MIT Press, 2008.

[62] J.-Y. Tzeng, Toward a more civilized design: Studying the effects of computers that apologize, *International Journal of Human–Computer Studies* **61**(3) (2004), 319–345.

[63] S. Utz, U. Matzat and C. Snijders, On-line reputation systems: The effects of feedback comments and reactions on building and rebuilding trust in on-line auctions, *International Journal of Electronic Commerce* **13**(3) (2009), 95–118.

[64] M. Winikoff, Implementing commitment-based interactions, in: *AAMAS'2007*, 2007, pp. 868–875.

[65] M. Wooldridge and N.R. Jennings, The cooperative problem-solving process, *Journal of Logic and Computation* **9**(4) (1999), 563–592.