Engineering Pro-sociality with Autonomous Agents

Ana Paiva and **Fernando P. Santos** and **Francisco C. Santos**
INESC-ID and Instituto Superior Técnico, University of Lisbon
Av. Prof. Cavaco Silva, Tagus Park, 2744-016 Porto Salvo,
Portugal

Abstract

This paper envisions a future where autonomous agents are used to foster and support pro-social behavior in a hybrid society of humans and machines. Pro-social behavior occurs when people and agents perform costly actions that benefit others. Acts such as helping others voluntarily, donating to charity, providing informations or sharing resources, are all forms of pro-social behavior. We discuss two questions that challenge a purely utilitarian view of human decision making and contextualize its role in hybrid societies: *i)* What are the conditions and mechanisms that lead societies of agents and humans to be more pro-social? *ii)* How can we engineer autonomous entities (agents and robots) that lead to more altruistic and cooperative behaviors in a hybrid society? We propose using social simulations, game theory, population dynamics, and studies with people in virtual or real environments (with robots) where both agents and humans interact. This research will constitute the basis for establishing the foundations for the new field of *Pro-social Computing*, aiming at understanding, predicting and promoting pro-sociality among humans, through artificial agents and multiagent systems.

Introduction

Everyday we are inundated with reports of situations that challenge our belief in humanity. The aim of moving towards more humane and fair societies appears to have been forgotten, as anti-social behavior dominates the headlines. According to analysts, journalists and even some politicians, the world seems to be lacking empathy, compassion and caring[1]. When famous and influential people exhibit clear signs of not esteeming others, acting without conscience or guilt over the unearned privileges they often enjoy, we should indeed be worried. They are our society's role models. Similar concerns occur when established social norms (Nyborg and others 2016) are unable to provide escape to Hardin's tragedy of the commons (Hardin 1968), resulting in undesirable situations such as antibiotic resistance, climate change, or overexploitation of natural resources (Levin 2006; Nyborg and others 2016).

[1]https://www.theguardian.com/science/2013/jan/04/barack-obama-empathy-deficit

One can question if the advent of autonomous technology is in itself contributing to the adverse situations that we are witnessing. It is undeniable that the rise of technological giants has promoted a society that is less equal, and more divided (Stockhammer 2015). Perhaps the perception of autonomy and intelligence in current systems is also a factor leading to a decrease in our sense of responsibility towards others and thus, make us, humans, less humane. It is hard to know what role increasingly autonomous technology will play in this new society. However, since we are on the brink of an "autonomy" revolution (the named fourth industrial revolution), with autonomous cars already in our streets and drones at our doorsteps, we must address these questions. Social psychology and behavioral economics have been researching how constructs, such as altruism or empathy, affect decision-making and cooperation. Findings in these areas have rarely been taken into consideration by computer scientists, engineers and technology developers in general. In fact, the dominant view of human decision making is based on the *homo economicus* principle of utility maximization, and this is already the backbone of several approaches to model behavior in autonomous machines.

Despite the negative examples and the predictions of mainstream economic models, humans often act in ways that benefit others: people behave pro-socially when giving money to charity, donating blood, sharing food, offering one's seat in the bus, helping a co-worker with some problem or informing an outsider about the direction to a city location. Beyond small gestures, cooperation is the building block of complex social behavior and it underlies the evolutionary origins of human societies. It is thereby fundamental to understand – and engineer – the contexts that prevent selfishness and conflict, while allowing pro-sociality to be sustained (or induced, when absent). It is not by chance that the evolution of cooperation has been identified by Science's invited panel of scientists as one the major scientific challenges of our century (Pennisi 2005).

In this context, we would like to ask how autonomous agents can be used to nurture or nudge (Thaler and Sunstein 2008) cooperation and pro-sociality in a society of humans and machines. How can we design autonomous agents which, immersed within humans, can promote collective action in situations where it may not naturally arise? How can we foster cooperation in organizations, help people to address cyber-bullying when they witness it, combat the bystander problem, make people engaged in social good, promote sustainable habits, fight climate change, and so on? Can autonomous systems play a role there? To address these problems, several mechanisms have been identified as supportive of cooperation, in situation ranging from two-person dilemmas to large-scale collective action problems (Ostrom 2015; Nowak 2006; Rand and Nowak 2013).

Here we defend a complementarity between such mechanisms and autonomous machines in order to improve prosocial behaviour within human groups. This approach is particularly relevant in cooperation problems involving large populations, especially in situations where a minority of carefully engineered artificial agents may produce a regime shift towards pro-social behaviors. In fact, the introduction of artificial agents may offer the means to overcome large-scale coordination barriers (Santos and Pacheco 2011) and tipping points (Scheffer 2009) towards a more pro-social environment. Similarly, it may create novel tipping points, initially absent from human social dynamics. This can be achieved by designing autonomous agents that could influence others to behave in a certain way, by increasing the visibility of actions, advertising reputations or collective risks, indirectly enforcing pre-defined social norms, introducing previously absent behaviors, or simply creating empathic relations with humans – among many other possibilities.

The recent interest in AI applications for the good of society is not new, and there has been a surge of new developments and events over the past few years. Competitions or workshops, like the AAAI'17 WS on "AI and Operations Research for Social Good", whose purpose is to explore and promote the application of artificial intelligence for social good, are among many examples that we can find nowadays. In fact, the United Nations[2] together with the XPRIZE Foundation organized the AI for Good Global Summit in Geneva in 2017. Among other topics, these events address technical AI approaches for creating more sustainable cities, deal with disaster response, address the impact of inequality, or improve public health. The work here proposed goes in that direction, having the potential to cause impact in some of these application areas.

This paper, therefore, proposes a vision where autonomous systems pro-actively act, foster and promote prosociality, instead of passively allowing or supporting the delegation of responsibility into the technology. We believe that this new type of computing will be linked with aspects of transparency, accountability and participation, which are all timely and urgently needed in our society.

To begin with, we define *Pro-social Computing* as "computing directed at supporting and promoting actions that benefit the society and others at the cost of one's own". This is a broad notion that may encompass different alternative views of how to engineer pro-social computing. To make it more concrete, we will start by proposing simple scenarios where pro-social computing can be used. Then we will give a glimpse of research agenda for engineering pro-social autonomous agents and discuss the future of this area.

Application Cases

Just to place this area into perspective, let us illustrate three simple situations where pro-social computing and, more specifically, pro-social agents, may play a role in changing the prevailing non-cooperative social dynamics, in a hybrid society of humans and robots.

Fighting the bystander effect

The well-known case of Kitty Genovese's murder more than five decades ago is without a doubt the most publicized case of the infamous "bystander effect". In this horrific case, several witnesses were "caught, fascinated, distressed, unwilling to help but unwilling to turn away" (Darley and Latane 1968), while Kitty was attacked. Witnesses did not intervene, and Kitty Genovese was brutally murdered. The term bystander effect was actually coined after this event. In spite of controversies surrounding the role of the bystanders in that particular situation, many studies have been conducted over the years, where the bystander effect is repeatedly observed. This effect verifies that, as the number of people witnessing a distressing event increases, their willingness to help decreases (thus reducing pro-social behavior). In computer-mediated scenarios (e.g., social media) we can also observe the bystander effect, as it was shown that the amount of time for an intervention increases with number of people witnessing the situation (the virtual bystanders). In fact, the growth of cyber-bullying in social media can be clearly related to the bystander effect.

Why do people witness, condemn, and yet do not help? According to the theory proposed by Darley and Latane, three processes may occur before there is an action by the bystander to aid the victim:

- **1) Audience inhibition**, that is, individuals may not act as the risk of embarrassment arises if others are watching and it turns out that the situation did not require any help;
- **2) Social influence**, whereby inaction becomes the established behavior as individuals are observing others and take their inaction as a guideline for their own behavior;
- **3) Diffusion of responsibility**, that is, the costs of non-intervention are shared in the presence of other people.

Finally, if there is partial observability and uncertainty about what the others are doing, any bystander can even assume that one of the observers is already acting and helping, therefore disregarding the need to offer any assistance. From a technological standpoint one can ask if this bystander effect may be addressed, and in particular if:

- Can autonomous machines and agents (particularly if they are embodied in the physical world) be considered "audience" in this "bystander" effect? That is, would these autonomous machines increase the bystander effect?
- In particular, does the diffusion of responsibility also occur when, instead of humans, we have "autonomous machines"?

[2] See http://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx

- And social influence? Can machines/agents exhibit behaviors (either by acting or non acting) that influence other's (and humans') behaviors?
- If agents can have social influence on humans, would they be able to counter-act the bystander effect? If so, how can we build technology for that?

Sustaining fairness and preventing inequality

Human decision-making is often driven by fair and equalitarian motives (Camerer 2003). Factors such as the cultural setting (Oosterbeek, Sloof, and Van De Kuilen 2004), engagement in large-scale institutions (Henrich and others 2010), or even the socio-economic class of the individuals (Piff et al. 2010), provide clues regarding the propensity to be fair. In fact, the influence of fairness is often strong enough to overcome rationality and selfishness, which poses important challenges to disciplines aiming to justify fair behavior (Thaler 1988). In this realm, the experiments with the Ultimatum Game (UG) are particularly illuminating (Güth, Schmittberger, and Schwarze 1982). In this interaction paradigm, two agents interact with each other: the Proposer is endowed with some resource and has to propose a division with the Responder. If the Responder rejects the proposal, none of the players earn anything. If the proposal is accepted, they will divide the resource as it was proposed. In the context of UG, only the egalitarian division, in which both the Proposer and the Responder earn a similar reward, is considered a fair result. Multiple studies attest that people are fair when playing the UG (Camerer 2003). Interestingly, seemingly irrational decisions rely on a complex neural architecture: when facing unfair proposals by other humans, the areas of the brain that get activated are those associated with negative emotional states, such as anger and disgust (Sanfey et al. 2003). Introducing machines and artificial agents in the game may thus result in different responses, as the attribution of causality shifts (Blount 1995). Designing artificial agents that incorporate the mechanisms responsible for the levels of fairness observed in human interactions is non-trivial. Will humans infer causes and assign responsibilities to artificial agents? Will artificial agents blame humans (or other agents) for unfair behaviors? How to escape the *Computer Says No*[3] paradigm of unaccountable decision-making when being unfair, immortalized in the British sketch show Little Britain?

Besides economic games, the relationship between AI, fairness and equality has often been written with a negative connotation. AI was associated with unemployment due to the automation of low qualified job positions as well as with a decrease in social mobility given the inability to re-train individuals in order to positively engage in a hybrid human-agent society.

Notwithstanding, and despite skepticism, we believe that pro-social computing can bring the opportunity to engineer fair systems, using the lessons from, *e.g.*, psychology and evolutionary biology. For example, just as noisy bots aid coordination in populations of humans and agents (Shirado and Christakis 2017), specific behaviors, hard-coded in selected agents, may potentiate the ensuing levels of fairness in a hybrid society (Santos et al. 2016).

In what concerns inequality and fairness, several technical questions may be raised, regarding the challenges posed by a human-agent society. In particular:

- Can autonomous machines and agents undermine (or strengthen) the social and cultural ties existing in a society and deplete (or increase) the ensuing levels of fairness?
- Will "autonomous machines" lack human causal attribution, leading them to be excused from unfair behaviors?
- Will machines be able to engage in sanctioning and/or reciprocal arrangements, often pointed as sustaining fairness in human societies?

Promoting cooperation in complex multiagent systems

The problems discussed above may be seen as part of the broader discipline of cooperation studies (Sigmund 2010; Genesereth, Ginsberg, and Rosenschein 1986). Cooperation is one of the major elements of human social behavior, acting as the glue for the whole society. Essential institutions such as welfare provision, national defense, public health systems and courts depend on the willingness of citizens to contribute to a public good, i.e., to cooperate. Without our capacity to cooperate, we would not survive as a species. And yet, altruistic cooperation involves a cost to provide a benefit to others, challenging evolutionary and economic theories.

The dynamics of cooperation can be conveniently described as a complex adaptive system (Miller and Page 2009; Levin 2006), where macroscopic cooperative patterns emerge from the complex interplay of decisions, peer-influence and social norms adopted at the microscopic level. In this context, experimental economics combined with multiagent simulations grounded on game theory – and its population-based counterpart, evolutionary game theory – provide a powerful approach to model and understand the complex ecology of choices that characterizes this type of problems. This combination of tools has successfully identified key mechanisms associated with the emergence of cooperation, from kin and reciprocity mechanisms (Nowak 2006; Rand and Nowak 2013), to the positive impact of social norms (Axelrod 1986; Fehr and Fischbacher 2004; Nyborg and others 2016; Ohtsuki and Iwasa 2004; Santos, Santos, and Pacheco 2016), networks of interaction (Santos, Santos, and Pacheco 2008), signaling (Skyrms 2010), among others.

Cooperation among humans has further peculiarities: a meta-analysis performed on more than 100 experiments involving over 5,000 subjects found that, in general, opportunities for human-human communication significantly raised cooperation rates (Sally 1995). The idiosyncrasies of human deliberation process also impact the observed levels of cooperation. When people make rapid and intuitive decisions in a collaborative scenario, there is more cooperation than when people make their decisions after a time for deliberation and reflection (Rand, Greene, and Nowak 2012; Bear and Rand 2016; Jagau and van Veelen 2017); the ten-

[3]https://en.wikipedia.org/wiki/Computer_says_no

dency to be pro-social is intuitive, and subjects who reach their decisions more quickly are more cooperative.

There is therefore an opportunity to employ this knowledge about human cooperation dynamics in the design of human-agent systems in which cooperation emerges and is sustained over time. To do this, it is important to identify the environmental conditions that, combined with the presence of artificial "influential agents", would provide a paradigm shift in situations in which purely selfish behaviors are the expected outcomes. These conditions are naturally dynamic, as dilemmas change in time and/or depend on the frequency of behaviors in the population (Sigmund 2010; Stone and Veloso 2000).

Moreover, evidence shows that pro-social computing will be confronted with similar tools that aim at supporting the interests of just a few, instead of benefiting the society as a whole (think about twitter bots spreading misinformation). In this context, the use of frequency-dependent models may provide important clues on how to successively adapt and prevail in a complex ecology of competing strategies – pro-social and selfishly designed agents – stemming a Red Queen dynamics (Van Valen 1973) that is common to a wide range of self-organized systems.

Being able to do this successfully would provide advances in several domains. Overexploitation of natural resources, voluntary vaccination, climate agreements and city planning, overuses and resistance to antibiotics, are just a few examples of the most important collective challenges in which, today, humans often act in their self-interest. Moreover, the ubiquitous nature of these problems (Levin 2006; Santos and Pacheco 2011; Tavoni and Levin 2014; Nyborg and others 2016) will turn any new principle discovered in these topics into a valuable contribution to a wide range of areas and applications addressing the interplay between technological, social and ecological systems. Again, this is a challenge for the area of pro-social computing, which should aim at understanding, predicting, and influencing human behavior. Several questions lie ahead, for instance:

- Can human-agent cooperation rely on the same mechanism (e.g., reciprocity, social norms, signaling, networks) that sustain cooperation in human societies?
- Can "influential agents" be used to elicit cooperation in scenarios where defection is today observed, providing a paradigm shift in situations in which purely selfish behaviors are the expected outcomes?
- How can the particularities of human (dual-process) deliberation and human communication be used to design artificial agents that both cooperate and elicit cooperation from humans and/or other agents?

Engineering pro-sociality: a research agenda

The application cases here presented provide a first glimpse of what problems pro-social computing can address. Research in pro-social computing involves many different areas, including not only AI, but also economics, sociology, psychology, human-agent interaction, information sciences and evolutionary biology. In fact, pro-social computing must carefully take into account the genetic and social pathways of pro-sociality. On one way, scientists like Frans de Waal take a positive stance and, based on large studies with some of our most closest primate relatives, provide evidence on the biological origin of kindness, compassion, cooperation and helping behaviors, which seem to underlie our most innate actions (De Waal 1996). On the other hand, one should regard pro-sociality as a social construct, often nurtured by intergenerational education processes (Dixit and Levin 2017). Studying and engineering pro-sociality is thus a multidisciplinary endeavor that must be addressed at different scales: We will need to study the effects at the macro society level but also at the micro individual level.

From a methodological point of view, for the area of pro-social computing to develop, we propose research in the following sub-areas:

- Understanding the emergence of pro-social behavior in populations using large-scale simulation of multiagent systems;
- Performing Experimental Studies with Humans and agents using social dilemmas to understand the conditions and situations where pro-social behaviors emerge;
- Engineering specific (even perhaps pathological) behaviors in the initial scenarios for social simulation to study the effects on populations;
- Performing studies with humans and Virtual Agents in Virtual Worlds. Agents can be built as pro-social (given the previous results), triggering pro-social behavior.
- Engineering Social Robots as pro-social agents in order to test the them in natural physical spaces, where humans and agents co-exist.

Engineering agents in a hybrid social environment (where both humans and agents co-exist) will involve not only pro-social agents in their behavior, but also agents that reason about the pro-sociality level of others, how that can be influenced, and how to act accordingly. It is known, from several studies on "social control", that the presence of others influences people's deviant behaviors. A disapproval look may suffice to prevent anti-social actions (Bateson, Nettle, and Roberts 2006). Similarly, the artificial look of a robot may elicit altruism (Burnham and Hare 2007). These nuances will become fundamental once we start combining humans and agents in pro-social computing. As such, to engineer pro-sociality, we will need to address a set of other cognitive capabilities, and in particular Empathy, Morality and Theory of Mind. Empathy has been largely used by the media as a persuasive tool to make people imagine themselves in the place of a suffering other and to motivate help (Coke, Batson, and McDavis 1978). In fact, we see empathy as essential to foster pro-social actions and, as such, empathy will need to be synthesized in agents (Paiva et al. 2004; 2017) as well as modeled as a heuristic to understand others – potentially in group interaction (Santos et al. 2015). Morality, that is, the capability to act following a given code of conduct, should also play a central role in pro-social computing. Agents should have the capacity to distinguish between *good* and *bad* behaviors in specific contexts, must be considered both during their own decision-making process and when judging the actions of surrounding agents. Finally, Theory of Mind can be used to create models of the internal state of other agents and humans and reason about them. All these capabilities will constitute the building blocks of the agents that will allow them to determine the desirability of an event for others and the society, as well as their subjective individual appraisals (Dias, Mascarenhas, and Paiva 2014).

We believe Pro-social computing to be a promising new area that will support positively the role of AI in the decisions made by future societies. For centuries, the investigation into human nature has tried to answer whether humans are either fundamentally good or fundamentally bad. Luckily, despite human nature being guided mostly by self-serving motivations, it is also known that we help each other at our own cost. Our AI systems should also do that, and take advantage of this characteristic of human nature to promote pro-sociality in general.

Acknowledgements

This work was supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, PhD fellowship SFRH/BD/94736/2013 and through the following FCT funded projects: AMIGOS (PTDC/EEISII/7174/2014), STOCA (PTDC/MAT-STA/3358/2014) and eCoop (PTDC/EEI-SII/5081/2014), as well as by the EC H2020 project LAW TRAIN project, with grant agreement n. 653587.

References

Axelrod, R. 1986. An evolutionary approach to norms. *Am. Polit. Sci. Rev.* 80(4):1095–1111.

Bateson, M.; Nettle, D.; and Roberts, G. 2006. Cues of being watched enhance cooperation in a real-world setting. *Biol. Lett* 2(3):412–414.

Bear, A., and Rand, D. G. 2016. Intuition, deliberation, and the evolution of cooperation. *Proc. Natl. Acad. Sci. USA* 113(4):936–941.

Blount, S. 1995. When social outcomes aren't fair: The effect of causal attributions on preferences. *Organ Behav Hum Decis Process.* 63(2):131–144.

Burnham, T. C., and Hare, B. 2007. Engineering human cooperation. *Hum Nat.* 18(2):88–108.

Camerer, C. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton Univ. Press.

Coke, J. S.; Batson, C. D.; and McDavis, K. 1978. Empathic mediation of helping: A two-stage model. *J Pers Soc Psychol.* 36(7):752.

Darley, J. M., and Latane, B. 1968. Bystander intervention in emergencies: diffusion of responsibility. *J Pers Soc Psychol.* 8(4p1):377.

De Waal, F. B. 1996. *Good natured*. Harvard Univ. Press.

Dias, J.; Mascarenhas, S.; and Paiva, A. 2014. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*. Springer International Publishing. 44–56.

Dixit, A., and Levin, S. 2017. Social creation of pro-social preferences for collective action. In *The Theory of Externalities and Public Goods*. Springer. 127–143.

Fehr, E., and Fischbacher, U. 2004. Social norms and human cooperation. *Trends Cogn. Sci.* 8(4):185–190.

Genesereth, M. R.; Ginsberg, M. L.; and Rosenschein, J. S. 1986. Cooperation without communication. In *AAAI'86*. AAAI Press.

Güth, W.; Schmittberger, R.; and Schwarze, B. 1982. An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3(4):367–388.

Hardin, G. 1968. The tragedy of the commons. *Science* 162(3859):1243–1248.

Henrich, J., et al. 2010. Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327(5972):1480–1484.

Jagau, S., and van Veelen, M. 2017. A general evolutionary framework for the role of intuition and deliberation in cooperation. *Nat Hum Behav* 1(8):s41562–017.

Levin, S. A. 2006. Learning to live in a global commons: socioeconomic challenges for a sustainable environment. *Ecol. Res.* 21(3):328–333.

Miller, J. H., and Page, S. E. 2009. *Complex adaptive systems: An introduction to computational models of social life*. Princeton Univ Press.

Nowak, M. A. 2006. Five rules for the evolution of cooperation. *Science* 314(5805):1560–1563.

Nyborg, K., et al. 2016. Social norms as solutions. *Science* 354(6308):42–43.

Ohtsuki, H., and Iwasa, Y. 2004. How should we define goodness?reputation dynamics in indirect reciprocity. *J Theor Biol* 231(1):107–120.

Oosterbeek, H.; Sloof, R.; and Van De Kuilen, G. 2004. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Exp. Econ.* 7(2):171–188.

Ostrom, E. 2015. *Governing the commons*. Cambridge Univ. Press.

Paiva, A.; Dias, J.; Sobral, D.; Aylett, R.; Sobreperez, P.; Woods, S.; Zoll, C.; and Hall, L. 2004. Caring for agents and agents that care: Building empathic relations with synthetic agents. In *AAMAS'04*, 194–201. IFAAMAS.

Paiva, A.; Leite, I.; Boukricha, H.; and Wachsmuth, I. 2017. Empathy in virtual agents and robots: A survey. *ACM Trans Interact Intell Syst* 7(3):11.

Pennisi, E. 2005. How did cooperative behavior evolve? *Science* 309(5731):93–93.

Piff, P. K.; Kraus, M. W.; Côté, S.; Cheng, B. H.; and Keltner, D. 2010. Having less, giving more: the influence of social class on prosocial behavior. *J Pers Soc Psychol.* 99(5):771.

Rand, D. G., and Nowak, M. A. 2013. Human cooperation. *Trends Cogn. Sci.* 17(8):413–425.

Rand, D. G.; Greene, J. D.; and Nowak, M. A. 2012. Spontaneous giving and calculated greed. *Nature* 489(7416):427.

Sally, D. 1995. Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. *Ration. Soc.* 7(1):58–92.

Sanfey, A. G.; Rilling, J. K.; Aronson, J. A.; Nystrom, L. E.; and Cohen, J. D. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300(5626):1755–1758.

Santos, F. C., and Pacheco, J. M. 2011. Risk of collective failure provides an escape from the tragedy of the commons. *Proc. Natl. Acad. Sci. USA* 108(26):10421–10425.

Santos, F. P.; Santos, F. C.; Paiva, A.; and Pacheco, J. M. 2015. Evolutionary dynamics of group fairness. *J Theor Biol* 378:96–102.

Santos, F. P.; Santos, F. C.; Paiva, A.; and Pacheco, J. M. 2016. Execution errors enable the evolution of fairness in the ultimatum game. In *22nd European Conference on Artificial Intelligence (ECAI 2016)*, 1592–1593.

Santos, F. C.; Santos, M. D.; and Pacheco, J. M. 2008. Social diversity promotes the emergence of cooperation in public goods games. *Nature* 454(7201):213.

Santos, F. P.; Santos, F. C.; and Pacheco, J. M. 2016. Social norms of cooperation in small-scale societies. *PLoS Comput Biol* 12(1):e1004709.

Scheffer, M. 2009. *Critical transitions in nature and society*. Princeton University Press.

Shirado, H., and Christakis, N. A. 2017. Locally noisy autonomous agents improve global human coordination in network experiments. *Nature* 545(7654):18.

Sigmund, K. 2010. *The calculus of selfishness*. Princeton Univ. Press.

Skyrms, B. 2010. *Signals: Evolution, learning, and information*. Oxford Univ. Press.

Stockhammer, E. 2015. Rising inequality as a cause of the present crisis. *Cambridge Journal of Economics* 39(3):935.

Stone, P., and Veloso, M. 2000. Multiagent systems: A survey from a machine learning perspective. *Auton Robots* 8(3):345–383.

Tavoni, A., and Levin, S. 2014. Managing the climate commons at the nexus of ecology, behaviour and economics. *Nat. Clim. Change* 4(12):1057.

Thaler, R. H., and Sunstein, C. R. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale Univ. Press.

Thaler, R. H. 1988. Anomalies: The ultimatum game. *J. Econ. Perspect* 2(4):195–206.

Van Valen, L. 1973. A new evolutionary law. *Evol Theory* 1:1–30.