Chapter 11
(Social) Norms and Agent-Based Simulation

Giulia Andrighetto, Stephen Cranefield, Rosaria Conte, Martin Purvis, Maryam Purvis, Bastin Tony Roy Savarimuthu, and Daniel Villatoro

11.1 Introduction

This chapter aims to identify the main relevant steps in the evolution of norms as well as some of the factors or determinants of such a process, and to discuss the most urgent scientific tasks to be fulfilled within a community of scientists committed to the study of norms. It is clearly the case that the scientific study of norms needs innovation and opening up to new instruments, new tools, new competencies, and especially new perspectives and approaches. In the last 50 years or so, the issue of norms has been of growing concern for moral and analytical philosophers and for several sub-communities within the social and behavioral sciences (see also Chap. 12). Our understanding of norms did not make significant progress until the advent of computational and simulation-based social science. The formal study of prosocial behavior accomplished within evolutionary game theory produced the most interesting results when scientists deducing macroscopic properties, such as norm-based societies, from properties at the microscopic level, started to look at the

G. Andrighetto (✉)
Institute of Cognitive Science and Technologies, CNR, Rome, Italy
e-mail: giulia.andrighetto@istc.cnr.it

S. Cranefield • M. Purvis • M. Purvis • B.T.R. Savarimuthu
Information Science Department, University of Otago, Dunedin, New Zealand
e-mail: scranefield@infoscience.otago.ac.nz; mpurvis@infoscience.otago.ac.nz; tehrany@infoscience.otago.ac.nz; TonyR@infoscience.otago.ac.nz

R. Conte
Institute of Cognitive Science and Technologies, CNR, Rome, Italy
e-mail: rosaria.conte@istc.cnr.it

D. Villatoro
IIIA – CSIC, Barcelona, Spain
e-mail: dvillatoro@iiia.csic.es

S. Ossowski (ed.), *Agreement Technologies*, Law, Governance and Technology Series 8, DOI 10.1007/978-94-007-5583-3_11,

conditions at which prosocial equilibria emerge. Rather than deducing equilibria from the bottom up, they inverted the methodological procedure: they started to wonder what minimal conditions are required for a certain effect to occur. This is not a cost-free procedure. It brought about a number of counterproductive effects, the most important being a certain degree of arbitrariness of the models developed (for a survey, see Conte and Paolucci 2002). The simulation-based study of the emergence of cooperation and norms started by Axelrod (1984) led to a myriad of agent-based models that are almost totally ad hoc. But interdisciplinarity helps. By the time the simulation-based study of social phenomena had become prominent, another computational field at the intersection between the social sciences and artificial intelligence had already come to the front stage of science, and that is (Multi) Agent Systems. This field is strongly indebted to the logic-based study of action, mental states, and social facts. Unlike game theory, this formal tradition is concerned with the mechanisms of agency at different levels of reality, more than its products. As argued in this chapter, the merging of Agent-Based Modelling (ABM) and Multi-Agent Systems (MAS) appears as a promising direction to give a strong, innovative boost to the study of norms.

11.2 Norm Learning

Research in Normative Multi-Agent Systems often assumes that norms are specified by the institution and all the agents in the society know about these norms ahead of time (Aldewereld et al. 2006; Shoham and Tennenholtz 1995). These works aim to study mechanisms for norm enforcement. On the other hand, researchers interested in the emergence of norms do not assume that agents know what the norms are a priori, but investigate how agents can derive norms from interactions. However, in studying how these norms emerge, most works model interactions based on simple cooperation or coordination games (Sen and Airiau 2007). Agents using these models can undertake few actions (e.g. cooperate and defect). Game-based interaction models do not capture rich interactions that take place in real life and also do not consider the large action-space of agents. So, there is a need for studying mechanisms that take into account the large number of actions that an agent is capable of performing. Some work in this direction has begun. For example, the work of Savarimuthu et al. makes use of a data-mining approach for the identification of norms (Savarimuthu et al. 2010a,b). However, the number of actions that are performed by agents in their work are small (e.g. four in Savarimuthu et al. 2010a, eight in Savarimuthu et al. 2010b). We believe this area has potential for further investigation. Additionally, how much domain knowledge an agent possesses and also its prior knowledge about norms may play a role in norm identification. These aspects can be explored further.

Another limitation of current simulation-based works on norms is the lack of consideration of all three aspects of active learning on the part of an agent (i.e. learning based on doing, observing and communicating). Most studies that

investigate norm emergence using simulations employing simple games have only used learning based on doing (Sen and Airiau 2007; Walker and Wooldridge 1995). Some works have considered observation-based learning (Epstein 2001; Hoffmann 2003) and only a few have considered communication-based learning (Verhagen 2001; Walker and Wooldridge 1995). We believe there is a lot of scope for integrating these three types of learning wherever applicable. The EMIL framework (Andrighetto et al. 2007) and the framework for norm identification (Savarimuthu et al. 2010b) have considered all three aspects. However, several works have not considered combining these three aspects (Savarimuthu et al. 2011). We believe this is a good venue for future investigation. Additional problems that may arise such as the problem of lying in communication-based learning will need to be addressed.

An important aspect in the learning of norms is to endow agents with the ability to identify the presence of norms through sanctions and rewards. Thus, those actions that signal the presence of norms can be used as the starting point for learning norms. Some works have considered signalling as a starting point for norm identification (Savarimuthu et al. 2010a,b). More work in this area can be undertaken. For example, the question of *where do the motivations for these signals come from* can be investigated. The motivations could include the sanctioning agent's utility going below a certain threshold or that the sanctioning agent is altruistic and wants others to behave in a certain way. Additionally, in agent societies the action that is being sanctioned may not be known ahead of time and the sanctions/rewards may emerge dynamically and can also change during time. This can be investigated using simulation systems.

A potential area for the study of norms is to include humans in the simulation loop to seed norms where agents can learn from human agents and also investigate how software agents can recommend norms to humans who can then choose the norm that they believe to be most applicable in a given context.

11.3 Conflicting Norms in Agent Societies

The employment of norms in multi-agent societies parallels the way they are used in human societies. In both contexts, norms represent prescriptions spreading through a population, that are not rigidly encoded in law and that are not enforced by institutional authorities. Instead these normative rules evolve by mutual consent, and their use in society is encouraged in a distributed fashion by sanctioning on the part of the group members, themselves. The advantages of norms is that (a) their distributed enforcement makes them scalable with respect to the size of agent communities and (b) they can evolve according to changing social contexts.

However, the flexibility of agent norms (a strength) also leads to the likelihood of conflicting norms (a potential weakness): two norms may be invoked in a particular situation that dictate incompatible behaviour. Although conflicting norms have received considerable attention in the multi-agent research community (Oren et al. 2008; Vasconcelos et al. 2009), we believe that there are still many interesting issues to be addressed with respect to norm conflict.

A simple example is the case of a theater performance, where the norm usually prevails that the audience should be quiet. However, if someone in the audience were to become critically ill, then it would presumably be suitable for someone to cry out, "is there a doctor in the house". In this case a higher-level norm would be invoked that urges one to take extraordinary measures to save someone's life, and this norm would presumably override the conventional norm for audience silence. Similarly, we normally follow the norm of "first come, first served" while waiting in an airline ticket queue, but if there is a late-arriving passenger with an urgent need for immediate service, then that passenger might be allowed to jump the queue in order to service his or her emergency.

But norm conflict can be much more complicated than those above two examples, since norms may be conditioned by temporal, spatial, gender, cultural and social circumstances. An example of gender related norm is when normative behaviour associated with British royal society, was violated when Australian Prime Minister Paul Keating put his arm around Queen Elizabeth.

Some of the existing work has been confined to limited situations (Vasconcelos et al. 2009), where norm overlap has been considered only in the context of simple examples with two linear and measurable dimensions. There has been some interesting work using argumentation-based heuristics in order to maximize the compliancy among various norms and minimize the violations (conflicts) among them while trying to resolve conflicting norms (Giannikis and Daskalopulu 2011; Oren et al. 2008). For a complex scenario where different levels of norm may be involved, consider the case of a man who enrolls in a college class taught by a woman; here there are the overlapping norms (and potential conflicts) associated with normal classroom behavior (social circumstances) and also man-woman (gender) behaviour. But now consider further complications to this example. Suppose in this particular case that:

- The male student is a full professor and the woman is an assistant professor. Here the professor outranks the assistant professor, and there are norms associated with that hierarchical relationship.
- The male student is also the uncle or husband of the woman teacher. Now there are additional norms associated with that family relationship.
- The male student and the woman teacher come from different cultures, which may have conflicting norms associated with acceptable behavior.
- There may have been prior, special commitments made between the teacher and student, which invokes the norm that promises should be kept.

In general there may be some possible agreement concerning meta-norms associated with the resolution of norm conflict. Here are some examples:

- The principle that the least-restrictive imposition on behaviour should be chosen when norms conflict (or, alternatively, the most-restrictive interpretation may be preferred).
- The principle that the most recently installed norm of two conflicting norms is preferred, since it is presumably the most up-to-date.

- The principle that the more generally disadvantaged party in a social situation should be given the greatest consideration when it comes to norm conflict resolution.
- There may be a commonly accepted hierarchy of norms that can be invoked to resolve norm conflicts.

These meta-norms, of course, must be made public and achieve common consent for them to be effective. We believe that there would be interesting work that could be undertaken in this area that would make the use of norms in multi-agent social situation more practical and scalable in realistic open-system scenarios.

11.4 Norm Enforcement

Punishment is widely considered a viable tool for promoting and maintaining social order both in real and in virtual societies (Axelrod 1986; Fehr and Gachter 2000; Ostrom 1990). Several authors (Blanc et al. 2005; Boyd and Richerson 1992; Boyd et al. 2010; de Pinninck et al. 2007; Helbing 2010; Jaffe and Zaballa 2010) have tested the effect of punishment in regulating peer-to-peer simulated environments, showing that to solve free-riding problems a constant and stable punishment system is necessary. Other models have been designed to explain how to choose the most effective punishment to regulate (electronic) institutions (Grossi et al. 2007; Janssen et al. 2010; Rauhut and Junker 2009) (see also Chap. 15).

Although these studies have provided key insights to the understanding of punishment in artificial societies, they have largely looked at this mechanism from the classical economic perspective as a way of changing wrongdoers' conduct through the infliction of material costs (Becker 1968). This way of considering punishment is incomplete and not likely to maintain large-scale compliance at least with only a reasonable level of costs for the (artificial) social system. Instead, as suggested by Andrighetto et al. (2010b) and Villatoro et al. (2011), punishment is more effective in regulating agents' behaviour and promoting norm compliance when the economic incentive is combined with the communication of normative information about the prescribed conduct. If properly designed, punishment not only imposes a cost for the wrongdoing, but also informs violators (and the public) that the targeted behaviour is not approved of because it violates a social norm. Giardini et al. (2010) have referred to this mechanism as *sanction*, thus distinguishing it from mere *punishment*. Since sanction communicates the presence of norms and asks that they not be violated, it allows agents to learn of the existence of norms and that their violation is not condoned. As shown in Villatoro et al. (2011), sanction allows social norms to be activated and to spread more quickly in the population than if they were enforced only by mere punishment with the effect of increasing their compliance and substantially reducing the costs for achieving and maintaining social order.

Clearly, in real life situations there is often an overlap between these two mechanisms that makes it difficult to disentangle their relative effects. Therefore,

agent-based simulation seems to be the ideal tool for virtually isolating punishment and sanction. In this way, it becomes possible to (a) explore the specific contribution of each in promoting and maintaining cooperation, (b) design actions aimed to highlight and exploit such contributions, and possibly (c) to perform what-if analyses that allow us to address policy design issues. To fully operationalize the difference between punishment and sanction requires a complex cognitive agent architecture and the EMIL architecture (EMIL-A) seems a good candidate for this undertaking (for an extended description of this architecture we refer to Andrighetto et al. (2010a), Conte et al. (forthcoming) and Conte and Andrighetto (2012)). Unlike the vast majority of simulation models in which heterogeneous agents interact according to simple local rules, e.g. imitation rules, all EMIL-A agents are endowed with a normative architecture, allowing them to: recognize norms; generate new normative representations and to act on them; and finally to infer the normative information (explicitly or implicitly) conveyed by different enforcing mechanisms, such as punishment and sanction.

11.5 Benchmark Problems

A recent discussion at the COIN@AAMAS 2011 workshop (COIN 2011) identified the lack of agreement on standard benchmark problems for normative multi-agent systems as a problem for this research community.[1] While there are some scenarios that are commonly addressed in simulation-based research, particularly those based on traffic intersection scenarios or simple abstracted coordination and social dilemma games from game theory, these are generally extended to multi-agent and repeated interaction settings in different ways by different researchers. Furthermore, it can be difficult to tease apart the aspects of a simulation scenario that are essential to the research problems being addressed from those that are specific to the mechanisms used to address those problems.

It would be beneficial for the normative multi-agent systems research community to develop a culture of sharing and reusing, as exists in, for example, the machine learning community (Frank and Asuncion 2010). Adopting such a culture would help to focus the community's effort on specific challenges, and would allow a more direct comparison of the benefits of different approaches to solving those challenges. However, sharing benchmark problems is not as straightforward in the area of normative multi-agent systems as it is in machine learning, where a benchmark problem typically consists of a data set, a well defined problem, and (for supervised machine learning tasks) the "ground truth" against which the results of an applied technique can be measured. In contrast, a scenario in a multi-agent system does not

[1]This section is inspired by and elaborates on some of the views expressed at that workshop, but is not intended to be a collectively agreed report of the discussion, which had a wider focus than simulation-based studies alone.

involve interpretation of a static data set; rather it involves agents interacting with each other and possibly an environment, and the object of study (from the point of view of normative MAS, at least) is the dynamics of the society. What, then, might benchmark problems for normative MAS look like? For simulation-based research it would be beneficial to have easily reusable simulation environments with well defined interfaces for agents to be 'plugged in'. These could be scenario-specific simulations or generic MAS simulation frameworks (Neville and Pitt 2009). Another possibility for constructing benchmark simulation environments is to make use of virtual world simulators such as Second Life[2] or World of Warcraft[3] and to share tools that ease the task of connecting agents to these environments (Dignum et al. 2009; Ranathunga et al. 2011). This approach would allow investigation into the challenges of reasoning with norms in complex environments with many observable events and possible behaviours.

However, developing reusable simulation frameworks for community use is time consuming and difficult to obtain funding for. A more realistic goal, therefore, would be to develop an online forum where researchers could propose, discuss and vote on simulation scenarios with a view to establishing a set of standard benchmarks. A crucial aspect of such a forum would be to classify the proposed scenarios in terms of the research problems that they highlight and elide, to identify the benefits that the employment of norms would be expected to bring to the scenario, and to establish measures of success for any proposed implementations of the scenarios. A valuable side effect would be the emergence of a better understanding of the differing research issues that are seen as important in the community.

References

Aldewereld, H., F. Dignum, A. García-Camino, P. Noriega, J. A. Rodríguez-Aguilar, and C. Sierra. 2006. Operationalisation of norms for usage in electronic institutions. In *Proceedings of the fifth international joint conference on autonomous agents and multiagent systems (AAMAS)*, 223–225. New York: ACM.

Andrighetto, G., M. Campennì, R. Conte, and M. Paolucci. 2007. On the immergence of norms: A normative agent architecture. In *Proceedings of AAAI symposium, social and organizational aspects of intelligence*, Washington, DC.

Andrighetto, G., M. Campennì, F. Cecconi, and R. Conte. 2010a. The complex loop of norm emergence: A simulation model. In *Simulating interacting agents and social phenomena*, Agent-based social systems, vol. 7, ed. H. Deguchi et al., 19–35. Tokyo: Springer.

Andrighetto, G., D. Villatoro, and R. Conte. 2010b. Norm internalization in artificial societies. *AI Communications* 23: 325–339.

Axelrod, R. M. 1984. *The evolution of cooperation*. New York: Basic Books.

Axelrod, R. 1986. An evolutionary approach to norms. *The American Political Science Review* 80(4): 1095–1111.

[2] http://secondlife.com/

[3] http://www.battle.net/wow/

Becker, G. S. 1968. Crime and punishment: An economic approach. *The Journal of Political Economy* 76(2): 169–217.

Blanc, A., Y. K. Liu, and A. Vahdat. 2005. Designing incentives for peer-to-peer routing. In *INFOCOM*, Miami, 374–385.

Boyd, R., and P. Richerson. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology* 13(3): 171–195.

Boyd, R., H. Gintis, and S. Bowles. 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* 328(5978): 617–620.

COIN. 2011. Web site of the international workshop series on coordination, organizations, institutions and norms in agent systems. http://www.pcs.usp.br/~coin/.

Conte, R., and G. Andrighetto. 2012. Cognitive dynamics of norm compliance: From norm adoption to flexible automated conformity. *Artificial Intelligence and Law*. DOI: 10.1007/s10506-012-9135-6.

Conte, R., and M. Paolucci. 2002. *Reputation in artificial societies: Social beliefs for social order*. Berlin: Springer.

Conte, R., G. Andrighetto, and M. Campenni (eds.). (forthcoming). *Minding norms. Mechanisms and dynamics of social order in agent societies*. Oxford Series on Cognitive Models and Architectures. USA: Oxford University Press.

de Pinninck, A. P., C. Sierra, and W. M. Schorlemmer. 2007. Friends no more: Norm enforcement in multi-agent systems. In *Proceedings of AAMAS 2007*, Honolulu, ed. M. Durfee and E. H. Yokoo, 640–642.

Dignum, F., J. Westra, W. A. van Doesburg, and M. Harbers. 2009. Games and agents: Designing intelligent gameplay. *International Journal of Computer Games Technology* 1–18.

Epstein, J. M. 2001. Learning to be thoughtless: Social norms and individual computation. *Computational Economics* 18(1): 9–24.

Fehr, E., and S. Gachter. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90(4): 980–994.

Frank, A., and A. Asuncion. 2010. UCI machine learning repository. http://archive.ics.uci.edu/ml.

Giannikis, G. K., and A. Daskalopulu. 2011. Normative conflicts in electronic contracts. *Electronic Commerce Research and Applications* 10(2): 247–267.

Giardini, F., G. Andrighetto, and R. Conte. 2010. A cognitive model of punishment. In *Proceedings of the 32nd annual conference of the cognitive science society*, 11–14 Aug. 2010, 1282–1288. Austin: Cognitive Science Society.

Grossi, D., H. M. Aldewereld, and F. Dignum. 2007. Ubi lex, ibi poena: Designing norm enforcement in e-institutions. In *Coordination, organizations, institutions, and norms in agent systems II*, 101–114. Berlin: Springer.

Helbing, D., A. Szolnoki, M. Perc, and G. Szabo. 2010. Punish, but not too hard: How costly punishment spreads in the spatial public goods game. *New Journal of Physics* 12(8): 083,005.

Hoffmann, M. J. 2003. Entrepreneurs and Norm Dynamics: An Agent-Based Model of the Norm Life Cycle. Technical report, Department of Political Science and International Relations, University of Delaware.

Jaffe, K., and L. Zaballa. 2010. Co-operative punishment cements social cohesion. *Journal of Artificial Societies and Social Simulation* 13: 3. http://econpapers.repec.org/RePEc:jas:jasssj:2009-36-3.

Janssen, M. A., R. Holahan, A. Lee, and E. Ostrom. 2010. Lab experiments for the study of social-ecological systems. *Science* 328: 613–617.

Neville, B., and J. Pitt. 2009. PRESAGE: A programming environment for the simulation of agent societies. In *Programming multi-agent systems*. Lecture Notes in Artificial Intelligence, vol. 5442, ed. K. V. Hindriks, A. Pokahr, and S. Sardina, 88–103. Berlin: Springer.

Oren, N., M. Luck, S. Miles, and T. J. Norman. 2008. *An argumentation inspired heuristic for resolving normative conflict*. In *Proceedings of the fifth workshop on coordination, organizations, institutions, and norms in agent systems (COIN@AAMAS-08)*, 41–56.

Ostrom, E. 1990. *Governing the commons: The evolution of institutions for collective action*. New York: Cambridge University Press.

Ranathunga, S., S. Cranefield, and M. Purvis. 2011. Interfacing a cognitive agent platform with Second Life. Discussion Paper 2011-02, Department of Information Science, University of Otago. http://eprints.otago.ac.nz/1093/.

Rauhut, H., and M. Junker. 2009. Punishment deters crime because humans are bounded in their strategic decision-making. *Journal of Artificial Societies and Social Simulation* 12(3): 1. http://jasss.soc.surrey.ac.uk/12/3/1.html.

Savarimuthu, B. T. R., S. Cranefield, M. A. Purvis, and M. K. Purvis. 2010a. Norm identification in multi-agent societies. Discussion Paper 2010/03, Department of Information Science, University of Otago. http://eprints.otago.ac.nz/873/.

Savarimuthu, B. T. R., S. Cranefield, M. A. Purvis, and M. K. Purvis. 2010b. Obligation norm identification in agent societies. *Journal of Artificial Societies and Social Simulation* 13(4): 3. http://jasss.soc.surrey.ac.uk/13/4/3.html.

Savarimuthu, B. T. R. 2011. Norm learning in multi-agent societies. Discussion Paper 2011/05, Department of Information Science, University of Otago. http://otago.ourarchive.ac.nz/handle/10523/1690.

Sen, S., and S. Airiau. 2007. Emergence of norms through social learning. In *Proceedings of the twentieth international joint conference on artificial intelligence (IJCAI)*, 1507–1512. Menlo Park: AAAI Press.

Shoham, Y., and M. Tennenholtz. 1995. On social laws for artificial agent societies: Off-line design.*Artificial Intelligence* 73(1–2): 231–252.

Vasconcelos, W., M. Kollingbaum, and T. Norman. 2009. Normative conflict resolution in multi-agent systems. *Autonomous Agents and Multi-Agent Systems* 19: 124–152.

Verhagen, H. 2001. Simulation of the learning of norms. *Social Science Computer Review* 19(3): 296–306.

Villatoro, D., G. Andrighetto, R. Conte, and J. Sabater-Mir. 2011. *Dynamic sanctioning for robust and cost-efficient norm compliance*, 414–419. Barcelona: IJCAI/AAAI.

Walker, A., and M. Wooldridge. 1995. Understanding the emergence of conventions in multi-agent systems. In *Proceedings of the first international conference on multi-agent systems (ICMAS)*, ed. V. Lesser, 384–389. Menlo Park: AAAI.