Social Laws: Logic and Games

Thomas Ågotnes
Department of Information Science and Media Studies
University of Bergen, Norway

Social laws (or *normative systems*) have emerged as a natural and powerful paradigm for coordinating multi-agent systems. The social laws paradigm exposes the whole spectrum between fully centralised and fully decentralised coordination mechanisms. A social law is, intuitively, a constraint on the behaviour of agents, which ensures that their individual behaviours are compatible. Typically, a social law is imposed off-line, minimising the chances of on-line conflict or the need to negotiate.

The social laws paradigm is based on the use of computational logic to reason about multi-agent systems, but frequently also makes use of game theory.

I will give an overview of the state-of-the-art in the use of social laws for coordination. I discuss questions such as: how can a social law that ensures some particular global behaviour be automatically constructed? If two social laws achieve the same objective, which one should we use? How can we construct a social law that works even if some agents do not comply? Which agents are most important for a social law to achieve its objective? I will furthermore exemplify of how, e.g., formal logic, game theory, voting theory and complexity theory, can be used and combined in multi-agent systems research. Depending on the time, I will focus on the following topics:

1. *Introduction to state transition models for multi-agent systems and the social laws paradigm.* I review state-transition models of multi-agent systems. Coordination is concerned with the behaviour of a system as a whole, with the global properties of the system. To this end, I introduce the language of Computation Tree Logic (CTL), with its natural branching time semantics to model possible computations of distributed systems. I'll focus on model checking, including computational complexity issues. I'll present a by now standard framework for social laws, introduced by Shoham and Tennenholtz in the early 1990s. A key assumption is that the designer (or analyst) of the multi-agent system has an objective, a property he or she wants the global behaviour of the system to satisfy. Key problems involving social laws are discussed, including: the effectiveness problem - will a given social law ensure the objective? The feasilbility problem - given an objective, does there exist an effective social law? The synthesis problem - given an objective, construct an effective social law. I show how these problems can be framed as model checking problems,

allowing standard model checkers from multi-agent systems to be used to solve them, and we discuss the computational complexity of the problems.

2. *Dealing with non-compliance.* In many cases it might be that some agents choose not to comply with a given social law. There are many possible causes of non-compliance; it could be deliberate because the agent does not consider it to be in his best interest to comply, or it could be that a component in the system fails. I discuss how to analyse the properties of a social law under possible non-compliance. In particular, I look at how robust the social law is, and try to identify the agents that are most important for the correct functioning of the system. We say that the social law is robust if the objective is still achieved if only a small number of the agents choose not to comply. Key problems here are: which agents are necessary, in the sense that the objective does not hold unless they comply? Does there exist a social law that is robustly feasible in the sense that compliance of a given group (or number) of agents is sufficient to ensure the objective? I further analyse the relative importance of agents by employing power indices, such as the Banzhaf index, to measure the influence an agent has on satisfaction of the objective in terms of choosing to comply with the social law or not. For example, I discuss how we can ensure that power is distributed evenly amongst the agents in a system, so as to avoid bottlenecks or single points of failure, or to understand where the key risks or vulnerabilities in a social law lie.

   I look more closely at one particular type of possible non-compliance: deliberate non-compliance by rational, self-interested agents. Thus we shift from the perspective of the designer to the perspective of the agent, and assume that also each agent has his own objective. Will an agent with a given objective comply with a given social law? As satisfaction of the objective depends upon whether or not the other agents in the system comply, this is a game-theoretic scenario. Key problems here include: does there exist a social law all agents would be better off complying with (as opposed to not complying)? Does there exist a social law that is a Nash implementation, in the sense that complying forms a Nash equilibrium?

3. *Reasoning about Social Laws.* I look more closely at how we can use formal logic to reason about social laws. In particular I, first, discuss how (variants of) deontic logic can be used to reason about different social laws in the context of a multi-agent system, e.g., allowing us to say that something is permitted in one social law but forbidden in another. Second, I show how standard logics can be extended in order to be able to frame the problems discussed in part (2), including robustness properties, as model checking problems.

References

[1] Y. Shoham and M. Tennenholtz. On the synthesis of useful social laws for artificial agent societies. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*, San Diego, CA, 1992.

[2] Y. Shoham and M. Tennenholtz. On social laws for artificial agent societies: Off-line design. In P. E. Agre and S. J. Rosenschein, editors, *Computational Theories of Interaction and Agency*, pages 597–618. The MIT Press: Cambridge, MA, 1996.

[3] D. Fitoussi and M. Tennenholtz. Choosing social laws for multi-agent systems: Minimality and simplicity. *Artificial Intelligence*, 119(1-2):61–101, 2000.

[4] M. Wooldridge and W. van der Hoek. On obligations and normative ability. *Jnl Appl. Logic*, 4(3-4):396–420, 2005.

[5] W. van der Hoek, M. Roberts, and M. Wooldridge. Social laws in alternating time: Effectiveness, feasibility, and synthesis. *Synthese*, 156(1):1–19, May 2007.

[6] Thomas Ågotnes, Wiebe van der Hoek, Carles Sierra Juan A. Rodriguez-Aguilar, and Michael Wooldridge. On the logic of normative systems. In M. M. Veloso, editor, *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI 2007)*, pages 1175–1180, California, 2007. AAAI Press.

[7] Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. Normative system games. In M. Huhns and O. Shehory, editors, *Proceedings of the Sixth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007)*, pages 876–883. IFAMAAS/Research Publishing Services, May 2007.

[8] T. Ågotnes, W. van der Hoek, and M. Wooldridge. Robust normative systems. In L. Padgham, D. Parkes, J. Muller, and S. Parsons, editors, *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 747–754, Estoril, Portugal, May 2008. IFAMAAS/ACM DL.

[9] T. Ågotnes, W. van der Hoek, M. Tennenholtz, and M. Wooldridge. Power in normative systems. In Decker, Sichman, Sierra, and Castelfranchi, editors, *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 145–152, Budapest, Hungary, May 2009. IFAMAAS/ACM DL.

[10] Thomas Ågotnes, Wiebe van der Hoek, Carles Sierra Juan A. Rodriguez-Aguilar, and Michael Wooldridge. A temporal logic of normative systems. In *Towards Mathematical Philosophy*, volume 28 of *Trends in Logic*, pages 69–104. 2009.

[11] Thomas Ågotnes, Wiebe van der Hoek, Juan A. Rodriguez-Aguilar, Carles Sierra, and Michael Wooldridge. Multi-modal CTL: Completeness, complexity and an application. *Studia Logica*, 92(1):1–26, 2009.

[12] Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. Robust normative systems and a logic of norm compliance. *Logic Journal of the Interest Group in Pure and Applied Logics (IGPL)*, 18(1):4–30, 2010.

[13] T. Ågotnes and M. Wooldridge. Optimal social laws. In *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, Toronto, Canada, May 2010. IFAMAAS/ACM DL.

[14] Piotr Kazmierczak, Truls Pedersen, and Ågotnes. NORMC: a norm compliance temporal logic model checker. In Kristian Kersting and Marc Toussaint, editors, *STAIRS 2012: Proceedings of the Sixth Starting AI Researchers Symposium*, Frontiers in Artificial Intelligence and Applications, pages 168–179. IOS Press, 2012.

[15] Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. Conservative social laws. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*, pages 49–54. IOS Press, 2012.