Law Enforcement in Norm-Governed Learning Agents

(Extended Abstract)

Régis Riveret
Dpt of EEE, Imperial College
London, UK
r.riveret@imperial.ac.uk

Giuseppe Contissa
Dept of Law, EUI, Italy
giuseppe.contissa@eui.eu

Antonino Rotolo
CIRSFID, University of Bologna, Italy
antonino.rotolo@unibo.it

Jeremy V Pitt
Dpt of EEE, Imperial College
London, UK
j.pitt@imperial.ac.uk

ABSTRACT

We study law enforcement mechanisms within a population of norm-governed learning agents. We show that a traditional analysis based on expected utility can be misleading, because learning agents tend to comply even though their surveillance is stopped. This has significant implications for the design of self-organising institutions with endogenous resources, where the cost of monitoring and norm enforcement has to be taken into consideration.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

Keywords

Law enforcement; Social simulation; Reinforcement learning

1. INTRODUCTION

Many simulation frameworks have been proposed for studying norm emergence and spreading, which are based on mechanisms such as social power, leadership, sanction, reputation, imitation, and network topologies [2]. However, the attention has been so far mainly focused on *social* and non-institutionalised scenarios.

The contribution of this work is to start filling the gap by studying, in a new simulation setting, some aspects of *law enforcement*. Legal norms are assumed here to work as an *ex ante* instrument: they are aimed at setting standards for activities to reduce risks arising from such activities, so that every agent that intends to engage in a regulated activity is required to comply with the applicable standard and incur the related compliance cost. We view sanctions as incentives to compliance, and we study how norm enforcement and monitoring may induce stable compliant behaviour. The assumption is that sanctions correspond to fines, i.e., administrative measures that are set, too, *ex ante*, irrespective of the actual harm caused by violations.

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

2. THE SIMULATION

We study law enforcement in a population of norm-governed learning agents. We use the framework of [1], which combines a logic-based formalism with an equation-based counterpart. The logical layer is rooted into a probabilistic defeasible logic: rules of the logic are attached with probabilities to describe the agents' minds and behaviours as well as uncertain environments and exogenous legal norms. The equation-based model for reinforcement learning, defined over this probability distribution, allows agents to adapt to their environment and self-organise.

We consider a road traffic scenario where the law enforcement is implemented by a learning agent (police), which can fine violations and adapt the amount of surveillance to the population of learning agents; these agents in turn can adapt their behaviour to comply, or not, with norms. We work with a population of N agents having the possibility to perform an action with three levels of care: ϕ_{high} (high), ϕ_{med} (medium) and ϕ_{low} (negligent). These levels, for each agent i, are associated with a payoff out_i: for ϕ_{high}, $out_i = 5$; for ϕ_{med}, $out_i = 10$; for ϕ_{low}, $out_i = 16$. There are rules stating when accidents may occur: the higher the level of care, the lesser the probability that an accident occurs. We assume that (1) the negative utility -200 of an accident is absorbed by the environment; (2) at time 100, an obligation to act with care (ϕ_{high}) enters in force; (3) police fines violations of agents whose detection is certain when monitoring is active; (4) individual fine $out^{fine} = F = -30$ and the cost of monitoring $out^{mon} = C = -4$ for each agent.

Let us first develop an *expected utility analysis*. At each time t, when no norm exists, the possible expected utilities of any agent i equal their payoff: if $\emptyset$ denotes inaction, (a) $EU_i(\emptyset) = 0$; (b) $EU_i(\phi_{high}) = 5$; (c) $EU_i(\phi_{med}) = 10$; (d) $EU_i(\phi_{low}) = 16$. If ϕ_{high}, ϕ_{med} and ϕ_{low} have respectively a 1%, 5% and 10% of probability of causing an accident, the associated expected global wealth EW including the cost of potential accidents for a population of N agents is $EW = EW(\emptyset)+EW(\phi_{high})+EW(\phi_{med})+EW(\phi_{low}) = 3\cdot N_{\phi_{high}} - 4 \cdot N_{\phi_{low}}$, where $N_{\phi_{high}}$ and $N_{\phi_{low}}$ are the number of agents executing actions ϕ_{high} and ϕ_{low}, respectively.

Without any obligation to act with care, a rational agent shall act with negligence, whereas the expected global wealth EW would be negative ($EW(\phi_{low}) = -4 \cdot N$). When a norm is added to enforce ϕ_{high}, the lower the probability

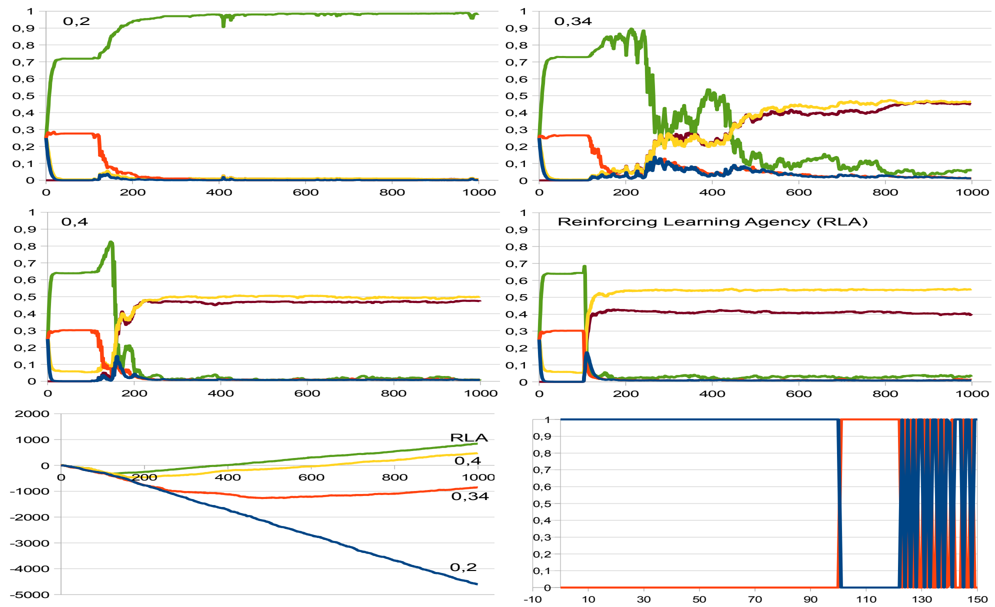

Figure 1: Top: Probability of behaviours vs. time. Green: ϕ_{low}, Orange: ϕ_{med}, Yellow: ϕ_{high}, Brown: Compliance via internalisation, Blue: Inaction. Bottom-Left: Global wealth vs. time. Bottom-Right: Probability of the behaviours of a learning enforcement agency vs. time. Blue: Inaction, Orange: Monitoring.

of monitoring, the lower the average frequency of monitoring, the lower is the cost of enforcement. However, there is a minimum frequency of surveillance to maintain agents' compliance. If we want that $EU_i^{norm}(\phi_{low}) < EU_i^{norm}(\phi_{high})$, i.e., that a rational agent acts with care, the probability of monitoring has to be greater than 11/30. In this case, $EW^{norm}(\phi_{high}) = EW(\phi_{high}) + N \cdot C = N \cdot [3 + C]$.

Consider now *learning agents*. Some typical runs are presented in Figure 1, with the evolution of behaviours and global wealth varying in function to the amount of surveillance. For a fixed probability 0.2 of monitoring, ϕ_{med} is not worthy, and thus the global wealth continues to decrease: the simulation confirms the expected utility analysis. When the probability is 0.34 ($< 11/30$), though the agents do not behave with care according to the expected utility calculus, we observe that the value 0.34 is a threshold where a majority of agents may get advantage of behaving with care in the long run. Note that the sudden loss of the probability of negligent behaviour is caused by some temporal concentration of monitoring that may randomly occur. With probability 0.4, agents comply to avoid fines (as in the expected utility analysis), and the decrease of global wealth is slowly stopped to finally increase at a steady step, but costs are high.

If police can adapt the amount of surveillance to negligent agents and the occurrence of accidents, once it enters into action the agents start behaving more carefully. When the number of negligent agents in combination with the occurrences of accidents is low enough to undermine the utility of surveillance, the enforcement is dramatically reduced. However, due to the inertia of learning, most of the agents continue to behave with care even though the surveillance has become infrequent. With learning agents, the experimental probability of surveillance is about 0.34 over the last hundreds steps. This is relevant when compared to the simulation with the fixed probability of 0.34 and the minimal sustainable frequency of 11/30 computed by the method of expected utilities. Notice that when the monitoring is fixed at 0.4, the global wealth is increasing almost as good as in the scenario with a learning enforcement agency. An advantage of a learning agency is clear at the introduction of the norm: police initiates a high frequency surveillance and thus agents quickly act with care. This advantage has to be compared to the case where the frequency of surveillance is fixed at 0.4 and less violations appear in the long term.

These results suggest that if normative behaviour is based on adaptive learning, the agents tend to comply even when surveillance is reduced or stopped (so that violation would be more convenient on the basis of a calculus of expected utilities). Thus, the enforcement (and the costs of surveillance) can be inferior to what would be necessary if agents were acting on the basis of a rational calculation of their expected utilities.

Acknowledgments

This work has been partially supported by the EC co-funded project SMART (FP7-287583) and a Marie Curie Intra-European Fellowship (PIEF-GA-2010-274057).

3. REFERENCES

[1] R. Riveret, A. Rotolo, and G. Sartor. Probabilistic rule-based argumentation for norm-governed learning agents. *Artificial Intelligence and Law*, 20:383–420, 2012.

[2] B. T. R. Savarimuthu and S. Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54, 2011.