Compliance Games
(extended abstract)

Piotr Kaźmierczak

Dept. of Computing, Mathematics and Physics
Bergen University College, Norway
phk@hib.no

Normative systems or *social laws*[1] have been studied in the multi-agent systems community as a framework for coordinating players [7,3,2,4,8]. The idea is that given a Kripke structure (which is simply a directed graph, sometimes labelled with extra elements, such as players), we list the edges that are *blacklisted*, i.e. deemed illegal by the system designer. The set of these edges is called a *normative system* or *social law*.

One of the key problems in the literature is that of *compliance* with a given social law. Ågotnes et al. [3] presented a logical approach to the problem by designing a logic of norm compliance which allows to reason about agents' goal achieving capabilities depending on whether they comply with given laws or not.

Here we present a new approach where agents are presented with a cooperative game in which successful coalitions are those that comply with the norm, and those not complying are punished (by means of null payoffs). We adapt a specific type of non-transferable utility game called Qualitative Coalitional Game to model compliance game scenarios, and we show how known decision problems for this class of games can be used to reason about compliance.

We begin by concisely presenting all the formal background for our work. First we describe the logical framework for Social Laws, which is based on Kripke semantics for modal logics. Following Ågotnes et al. [3], we define our models as agent-labelled Kripke structures in the following way:

Definition 0.1 (Agent-labelled Kripke Structure). *An* agent-labelled Kripke structure *(henceforth referred to simply as* structure*) K is a tuple $\langle S, R, V, \Phi, A, \alpha \rangle$ where:*

- S *is the non-empty, finite set of* states,
- $R \subseteq S \times S$ *is the total*[2] *relation between elements of S that captures* transitions *between states,*
- Φ *is a non-empty, finite set of propositional symbols,*
- $V : S \to 2^{\Phi}$ *is a labelling function which assigns propositions to states in which they are satisfied,*

[1] In the multi-agent systems literature normative systems and social laws stand for the same. However, in this paper we will always use the notion of a social law, since calling our restrictions "normative systems" can perhaps be confusing for readers with a background in deontic logic.

[2] That is, $\forall s \exists t \ (s,t) \in R$. This kind of relation is also sometimes called serial.

– *A is a non-empty finite set of agents, and*
– *$\alpha : R \to A$ a function that labels edges with agents.*

A *path* π over a relation R is an infinite sequence of states $s_0, s_1, s_2, \ldots$ such that $\forall u \in \mathbb{N} : (s_u, s_{u+1}) \in R$. $\pi[0]$ denotes the first element of the sequence, $\pi[1]$ the second, and so on. An *s-path* is a path π such that $\pi[0] = s$. $\Pi_R(s)$ is the set of s-paths over R, and we write $\Pi(s)$, if R is clear from the context.

Objectives are specified using the language of *Computation Tree Logic* (CTL), a popular branching-time temporal logic. We use an adequate fragment of the language defined by the following grammar:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathsf{E}\bigcirc \varphi \mid \mathsf{E}(\varphi \mathcal{U} \varphi) \mid \mathsf{A}(\varphi \mathcal{U} \varphi)$$

where p is a propositional symbol. The standard derived propositional connectives are used, in addition to standard derived CTL connectives such as $\mathsf{A}\bigcirc \varphi$ for $\neg\mathsf{E}\bigcirc \neg\varphi$ (see [6] for details). Satisfaction of a formula φ in a state s of a structure K, $K, s \models \varphi$, is defined as follows:

$$
\begin{aligned}
&K, s \models \top;\\
&K, s \models p \text{ iff } p \in V(s);\\
&K, s \models \neg\varphi \text{ iff not } K, s \models \varphi;\\
&K, s \models \varphi \vee \psi \text{ iff } K, s \models \varphi \text{ or } K, s \models \psi;\\
&K, s \models \mathsf{E}\bigcirc \varphi \text{ iff } \exists \pi \in \Pi(s) : K, \pi[1] \models \varphi;\\
&K, s \models \mathsf{E}(\varphi \mathcal{U} \psi) \text{ iff } \exists \pi \in \Pi(s), \exists u \in \mathbb{N}, \text{s.t. } K, \pi[u] \models \psi\\
&\qquad\qquad\qquad\qquad \text{and } \forall v, (0 \leq v < u) : K, \pi[v] \models \varphi;\\
&K, s \models \mathsf{A}(\varphi \mathcal{U} \psi) \text{ iff } \forall \pi \in \Pi(s), \exists u \in \mathbb{N}, \text{s.t. } K, \pi[u] \models \psi\\
&\qquad\qquad\qquad\qquad \text{and } \forall v, (0 \leq v < u) : K, \pi[v] \models \varphi.
\end{aligned}
$$

A *social law* $\eta \subseteq R$ is a set of black-listed ("illegal") transitions, such that $R \setminus \eta$ remains total.[3] A set of all social laws over R is denoted as $N(R)$. We say that $K \dagger \eta$ is a structure with a social law η *implemented* on it, i.e. for $K = \langle S, R, \Phi, V, A, \alpha \rangle$ and η, $K \dagger \eta = K'$ iff $K' = \langle S, R', \Phi, V, A, \alpha' \rangle$ with $R' = R \setminus \eta$ and:

$$\alpha'(s, s') = \begin{cases} \alpha'(s, s') & \text{if } (s, s') \in R' \\ \text{undefined} & \text{otherwise} \end{cases}$$

Also, $\eta \restriction C = \{(v, v') | (v, v') \in \eta \ \&\ \alpha(v, v') \in C\}$ for any $C \subseteq A$ – that is to account for agents that do not necessarily comply with the social law (i.e. we can consider situation in which only those edges that are "owned" by members of C are blacklisted).

Example 0.1. We introduce a running example that illustrates modeling a very simple Kripke structure.

Figure 1 presents and example Kripke structure with:

[3] This is a so-called "reasonableness" constraint – we do not want social laws implementation of which results in systems with dead-end states.

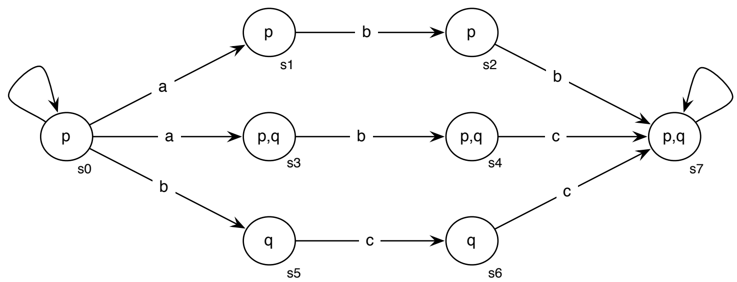

Fig. 1. Simple Kripke structure example.

- $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6, s_7\}$,
- $R = \{(s_0, s_0), (s_0, s_1), (s_1, s_2), (s_2, s_7), \ldots\}$,
- $\Phi = \{p, q\}$,
- $V(s_0) = \{p\}$, $V(s_1) = \{p\}$, $\ldots$, $V(s_7) = \{p, q\}$,
- $A = \{a, b, c\}$,
- $\alpha(s_0, s_1) = \alpha(s_0, s_3) = \{a\}$,
  $\alpha(s_1, s_2) = \alpha(s_3, s_4) = \alpha(s_0, s_5) = \alpha(s_2, s_7) = \{b\}$,
  $\alpha(s_4, s_7) = \alpha(s_5, s_6) = \alpha(s_6, s_7) = \{c\}$.

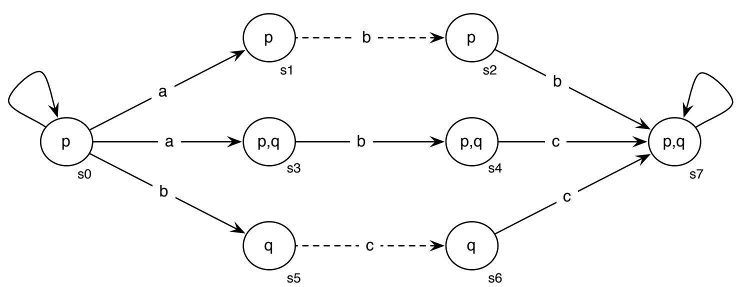

Fig. 2. Kripke structure with a social law implemented on it.

Figure 2 presents the same structure, but with a social law $\eta = \{(s_1, s_2), (s_5, s_6)\}$ implemented on it.

For the purpose of modelling compliance games we employ a formalism known as *Qualitative Coalitional Games* [9]. These are non-transferable utility cooperative games, which are particularly suitable for modelling goal-based scenarios. They are called "qualitative" because in contrast to most cooperative games, where a characteristic function assigns a numeric value (usually a real number) to each coalition, these games' characteristic functions assign a "good" or "bad" value to coalitions. Formal definitions follow below:

Definition 0.2 (Qualitative Coalitional Game). *A Qualitative Coalitional Game (abbreviated* QCG*) Γ is given by a tuple $\Gamma = \langle A, \Theta, \Theta_1, \dots, \Theta_n, \nu\rangle$, where A is a finite set of players, $\Theta = \{\theta_1, \dots, \theta_m\}$ is a set of* goals*, $\Theta_i \subseteq \Theta$ is a set of goals* per player*, and $\nu : 2^A \to 2^{2^{\Theta}}$ is a* characteristic function *of the game, assigning to each coalition $C \subseteq A$ a set of goals this coalition can achieve.*

We now use the above definition to formulate a particular kind of cooperative game, where agents are "rewarded" for complying to a social law (the characteristic function of the game pays coalitions whose members achieve at least one of their respective goals).

Definition 0.3 (Compliance Game). *A* compliance game *(abbreviated* CG*) is a* QCG *$\Gamma = \langle A, \Theta, \Theta_1, \dots, \Theta_n, \nu\rangle$ induced by a Kripke structure $K = \langle S, R, \Phi, V, A, \alpha\rangle$ and defined as follows:*

- *A is the set of players from K,*
- *$\Theta = \{\varphi_1, \dots, \varphi_m\}$ is a set of* goal formulas *expressed in the language of* CTL*,*
- $\nu(C) = \left\{ \bigcup_{i \in C} \left\{ \varphi_i \in \Theta_i \middle| K \dagger (\eta \restriction C) \models \varphi_i \right\} \right\}$

Example 0.2. Take a look at Figure 1 first. It is implicitly a Kripke structure with an empty social law implemented on it (all transitions are legal), and then we could formulate the following CG based on it:

- $\Theta = \{\varphi_1, \varphi_2\}$,
- $\Theta_a = \Theta_b = \{\varphi_1\}$, $\Theta_c = \{\varphi_2\}$.

where $\varphi_1 = \mathsf{E}\Diamond p$ and $\varphi_2 = \mathsf{E}\Diamond q$.[4] In the structure from Figure 1, the following coalition are successful: $\{a, b\}$, $\{b, c\}$, $\{a, b, c\}$. However, in a structure presented in Figure 2, only the grand coalition is successful.

We now make some observations about what the power of Qualitative Coalitional Games can bring to the area of social laws. This is mostly based on the study of fourteen intuitive decision problems for those games studied by Wooldridge & Dunne [9], and the natural problems when reasoning about social laws are taken from the work of Ågotnes et al., especially [3].

The first problem is that of *C-sufficiency* – namely checking whether given a Kripke structure K, a social law η and a goal formula φ, a coalition C of agents in K are *sufficient* for achieving φ under η. Deciding *C-sufficiency* is proved to be co-NP-complete [3]. We can make an observation that given the framework defined as above, C-sufficiency can be represented as an instance of a game-theoretic decision problem known as *successful coalition* (SC) or *selfish successful coalition* (SSC). SC asks whether given a particular QCG and some coalition, is there a feasible choice available such that it will satisfy all members

[4] $\mathsf{E}\Diamond \equiv \mathsf{E}(\top \mathcal{U} \varphi)$, a standard abbreviation in CTL to denote the "at some point in the future" modality.

of said coalition. The problem of SSC is then to answer the question of whether it is the case that a coalition in question has a feasible choice that will satisfy a given agent only if he is part of this coalition. These two problems answer more general questions than C-sufficiency, but capture similar intuitions. Also, these two problems are shown to be NP-complete [9].

The second important problem is C-*necessity*, which, similarly to C-*sufficiency*, asks whether a given coalition is necessary to achieve a given goal under certain restrictions implemented upon a structure. This problem is also co-NP-complete [3], and it corresponds directly to the game-theoretic problem of finding a *minimal coalition* (MC), which happens to be co-NP-complete [9] as well.

We do not address problems of C-*sufficient feasibility* (checking whether there is a social law such that a given coalition C will be sufficient when this system is implemented) and k-*robustness* (checking whether a given social law is effective as long as number k of agents complies with it), but we can exploit game-theoretic nature of our formulation of social law compliance problems. Given the framework of QCGs we can answer many decision problems that are very interesting from the point of view of a social law designer. One of such problems is *core membership* and *core non-emptiness* – checking whether there are such outcomes of the game in which agents have no incentive to abandon a certain coalition structure. Knowing one has a non-empty core can be very useful, because then the system designer knows that a set of stable coalitions will emerge. And even if it does not, game theory literature offers some solutions to ensure a non-empty core, such as *cost of stability* [5] (paying subsidies to agents such that they remain in a coalition). Furthermore, with a game-theoretic framework defined as above, we can formulate questions about the nature of the game itself: e.g. is it *trivial* (every coalition is successful) or *empty* (every coalition fails)?, and about the nature of goals: are they realizable, are some of them necessary?

Finally, we are able to address one other problem mentioned in the literature, namely finding "influential" players and measuring their power by means of computing the Banzhaf index and Banzhaf measure, as it is done in [1].

In conclusion, we claim that the framework presented above is more powerful, more flexible, and more appropriate for the study of social laws than the work that has been presented in the literature until now. We plan to further study it, with an emphasis on computational properties of a number of decision problems sketched in the paragraph above.

References

1. T. Ågotnes, W. van der Hoek, M. Tennenholtz, and M. Wooldridge. Power in normative systems. In *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 145–152, 2009.
2. T. Ågotnes, W. van der Hoek, and M. Wooldridge. Normative system games. In *Proc. of 6th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2007)*, pages 1–8, 2007.

3. T. Ågotnes, W. van der Hoek, and M. Wooldridge. Robust normative systems and a logic of norm compliance. *Logic Journal of the IGPL*, 18(1):4–30, 2009.
4. T. Ågotnes, W. van der Hoek, and M. Wooldridge. Conservative social laws. In L. D. R. et al., editor, *Proceedings of ECAI 2012*, 2012.
5. Y. Bachrach, E. Elkind, R. Meir, D. Pasechnik, M. Zuckerman, J. Rothe, and J. S. Rosenschein. The cost of stability in coalitional games. In *Proc. of the 2nd Int. Symposium on Algorithmic Game Theory (SAGT '09)*, pages 122–134, Berlin, Heidelberg, 2009. Springer-Verlag.
6. E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science Volume B: Formal Models and Semantics*. Elsevier Science Publishers B.V.: Amsterdam, The Netherlands, 1990.
7. Y. Shoham and M. Tennenholtz. On social laws for artificial agent societies: off-line design. *Artificial Intelligence*, 73:231–252, 1995.
8. W. van der Hoek, M. Roberts, and M. Wooldridge. Social laws in alternating time: effectiveness, feasibility, and synthesis. *Synthese*, 156(1):1–19, Sept. 2006.
9. M. Wooldridge and P. E. Dunne. On the Computational Complexity of Qualitative Coalitional Games. *Artificial Intelligence*, 158, 2004.