Social Laws for Multi-Agent Systems: Logic and Games

Lecture 3: Dealing with Non-Compliance

Thomas Ågotnes[1]

[1]Department of Information Science and Media Studies
University of Bergen, Norway

NII Tokyo 27 December 2011

Contents

What Happens if Agents Don't Comply?

- An implicit assumption above is that agents will *comply* with the rules we present them with.
- In the real world, this is not the case: people murder, commit adultery, steal, and bear false witness...
- Why should multi-agent systems be any different?

Why Wouldn't Agent's Comply?

- Deliberately:
  - because they personally benefit
  - because they enjoy causing trouble
- Accidentally:
  - bugs in the program, system crash
  - failed communication, misunderstanding

Three Ways of Dealing with Non-Compliance

1. Try to design the social law to be *robust* against failure
2. Find out who the important agents are, and devote your attention to them
3. Try to design the social law so that compliance is in everybody's interest (next lecture)

Compliance:notation

$$\eta \restriction C$$

is the social law that is the same as η except that it only contains the arcs of η that correspond to the actions of agents in C.

Contents

Design for Robustness

- With this approach, we try to design the social law so that *it does not matter if some agents do not comply*.
- We make the social law *robust* against non-compliance.
- Typical approach: include *redundancy*.

Goals and Social Laws

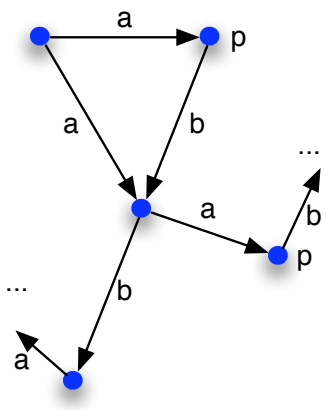
- We assume that there is a global goal in the form of a logical formula

$$\varphi$$

  expressing how the designer wants the system to behave
- The goal is typically not satisfied
- Design a social law: identify a set of "illegal" transitions such that the goal formula will be true if those transitions never are followed

Example

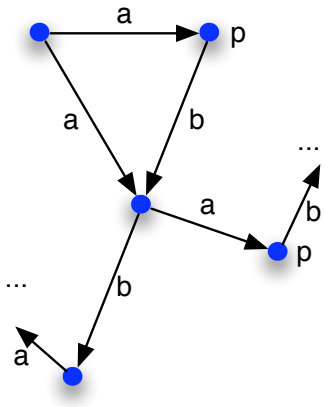
$$\varphi = \mathsf{A}\Diamond p$$



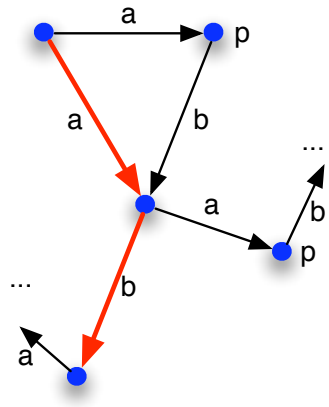
$$K \not\models \mathsf{A}\Diamond p$$

Example

$$\varphi = \mathsf{A}\Diamond p$$



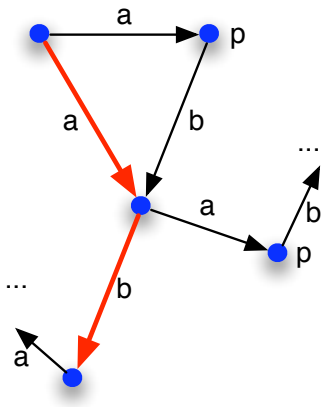
$K \not\models \mathsf{A}\Diamond p$



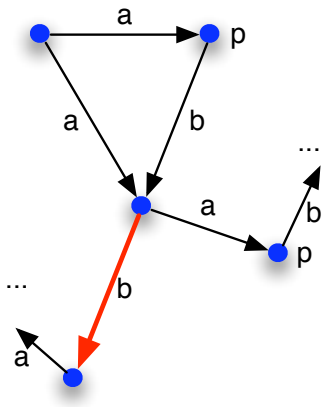
$K \dagger \eta \models \mathsf{A}\Diamond p$

Not all norms are created equal

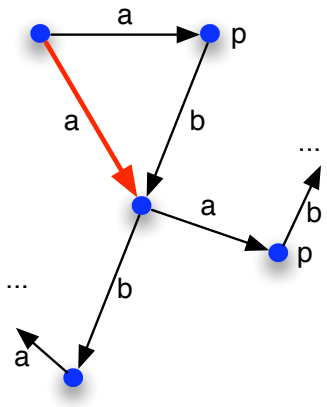
- There might be several effective social laws
- Which one is better?



$K \dagger \eta \models \mathsf{A}\Diamond p$



$K \dagger \eta' \models \mathsf{A}\Diamond p$



$K \dagger \eta'' \models \mathsf{A}\Diamond p$

Robustness

- A social law is *robust* to the extent to which *it remains effective (i.e., the system goal is still satisfied) in the event of non-compliance by some agents*

Example

Example

The system will not overheat as long as at least one sensor works as it should and either one of the relief valves is working as it should or the automatic shutdown is working as it should

Robustness

We formalise three approaches to characterising robustness:

1. Identify coalitions whose compliance is *necessary* or *sufficient*
2. Find the *number* of agents that we can tolerate non-compliance from
3. Logical characterisations (later)

Sufficiency

Let a model K, a norm η and an objective φ be given.
We say that $C \subseteq Ag$ are *sufficient* for η if the compliance of C with η is *effective*, i.e., iff:

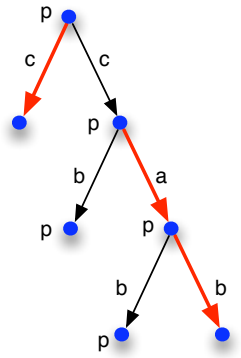
$$\forall C' \subseteq Ag : (C \subseteq C') \quad \Rightarrow \quad [K \dagger (\eta \restriction C') \models \varphi].$$

Sufficiency

Let a model K, a norm η and an objective φ be given.
We say that $C \subseteq Ag$ are *sufficient* for η if the compliance of C with η is *effective*, i.e., iff:

$$\forall C' \subseteq Ag : (C \subseteq C') \quad \Rightarrow \quad [K \dagger (\eta \restriction C') \models \varphi].$$

$$\varphi = \mathsf{A} \Box p$$



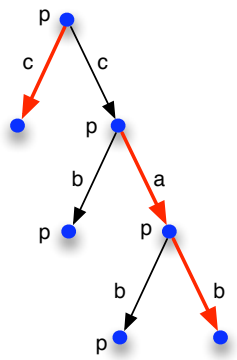
Compliance Robustness Power References

Sufficiency

Let a model K, a norm η and an objective φ be given.
We say that $C \subseteq Ag$ are *sufficient* for η if the compliance of C with η is *effective*, i.e., iff:

$$\forall C' \subseteq Ag : (C \subseteq C') \quad \Rightarrow \quad [K \dagger (\eta \restriction C') \models \varphi].$$

$$\varphi = \mathsf{A} \Box p$$



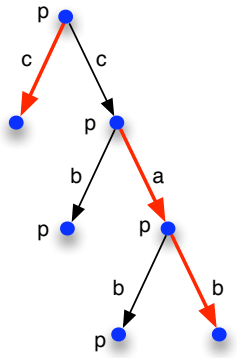
Sufficient: $\{c, a\}$

Sufficiency

Let a model K, a norm η and an objective φ be given.
We say that $C \subseteq Ag$ are *sufficient* for η if the compliance of C with η is *effective*, i.e., iff:

$$\forall C' \subseteq Ag : (C \subseteq C') \quad \Rightarrow \quad [K \dagger (\eta \restriction C') \models \varphi].$$

$$\varphi = \mathsf{A} \Box p$$



Sufficient: $\{c, a\}$ $\{c, b\}$

Compliance Robustness Power References

Sufficiency

Let a context K, a norm η and an objective φ be given.
We say that $C \subseteq Ag$ are *sufficient* for η if the compliance of C with η is *effective*, i.e., iff:

$$\forall C' \subseteq Ag : (C \subseteq C') \quad \Rightarrow \quad [K \dagger (\eta \restriction C') \models \varphi].$$

Sufficiency

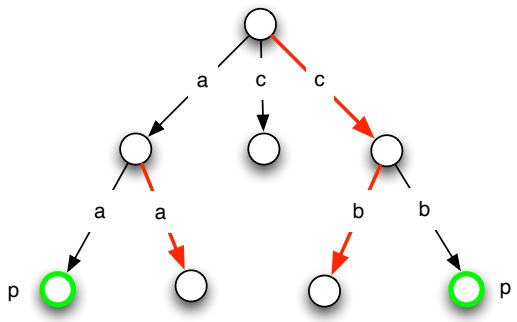
Let a context K, a norm η and an objective φ be given.
We say that $C \subseteq Ag$ are *sufficient* for η if the compliance of C with η is *effective*, i.e., iff:

$$\forall C' \subseteq Ag : (C \subseteq C') \quad \Rightarrow \quad [K \dagger (\eta \restriction C') \models \varphi].$$

Note: we can have that:

$$K \dagger (\eta \restriction C) \models \mathsf{E}\Diamond happy(d) \;\&\; K \dagger (\eta \restriction C \cup \{d\}) \not\models \mathsf{E}\Diamond happy(d)$$

Example



Take the system above, with $\varphi = \mathsf{E}\bigcirc\mathsf{A}\bigcirc p$.

1. $\{a\}$ is sufficient
2. $K \dagger (\eta \restriction \{b\}) \models \varphi$;
3. none of $\{b\}$, $\{c\}$ or $\{b, c\}$ is sufficient

Necessity

We say that C are *necessary* for η iff C *must* comply with η in order for it to be effective, i.e., iff:

$$\forall C' \subseteq A : [K \dagger (\eta \restriction C') \models \varphi] \quad \Rightarrow \quad (C \subseteq C').$$

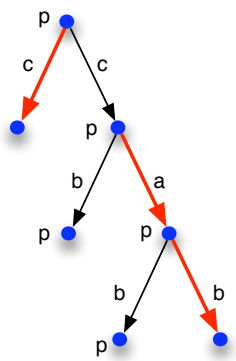
Compliance Robustness Power References

Necessity

We say that C are *necessary* for η iff C *must* comply with η in order for it to be effective, i.e., iff:

$$\forall C' \subseteq A : [K \dagger (\eta \restriction C') \models \varphi] \quad \Rightarrow \quad (C \subseteq C').$$

$$\varphi = \mathsf{A}\Box p$$

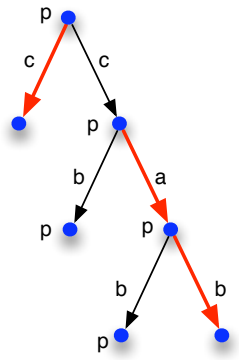


Necessity

We say that C are *necessary* for η iff C *must* comply with η in order for it to be effective, i.e., iff:

$$\forall C' \subseteq A : [K \dagger (\eta \restriction C') \models \varphi] \quad \Rightarrow \quad (C \subseteq C').$$

$$\varphi = \mathsf{A} \Box p$$



Necessary: $\{c\}$

Some Results

General C-sufficiency

Deciding C-sufficiency is co-NP-complete

Universal C-sufficiency

Deciding C-sufficiency for *universal* objectives is polynomial time decidable

Some Results

General *C*-sufficiency

Deciding *C*-sufficiency is co-NP-complete

Universal *C*-sufficiency

Deciding *C*-sufficiency for *universal* objectives is polynomial time decidable

Universal and Existential Goals

Universal and existential fragment of CTL, respectively:

$$\mu ::= \top \mid p \mid \neg p \mid \mu \vee \mu \mid \mathsf{A}\mu \mid \mathsf{A}\Box\mu \mid \mathsf{A}(\mu\,\mathcal{U}\,\mu)$$
$$\varepsilon ::= \top \mid p \mid \neg p \mid \varepsilon \vee \varepsilon \mid \mathsf{E}\varepsilon \mid \mathsf{E}\Box\varepsilon \mid \mathsf{E}(\varepsilon\,\mathcal{U}\,\varepsilon)$$

Some properties

Some Properties

- There might be no sufficient coalitions.
- There is always a necessary coalition: the empty coalition.
- There might be disjoint sufficient coalitions.
- There might be no non-empty necessary coalitions.
- If C is necessary and C' sufficient, then $C \subseteq C'$.
- ...

Feasibility of Robust Systems

Given a goal φ, and a 'reliable' coalition C:

C-sufficient feasibility

$$\exists \eta : (K \dagger \eta \models \varphi) \wedge \\ \forall C' \subseteq Ag : (C \subseteq C') \Rightarrow [K \dagger (\eta \restriction C') \models \varphi].$$

Theorem

Deciding C-sufficient feasibility is Σ_2^p-complete.

Feasibility of Robust Systems

Given a goal φ, and a 'reliable' coalition C:

C-sufficient feasibility

$$\exists \eta : (K \dagger \eta \models \varphi) \wedge \forall C' \subseteq Ag : (C \subseteq C') \Rightarrow [K \dagger (\eta \restriction C') \models \varphi].$$

Theorem

Deciding C-sufficient feasibility is Σ_2^p-complete.

k-sufficiency

Let K and φ be given.

Definition

Where $k \geq 1$, we say a social law η is *k-sufficient* if the compliance of *any arbitrary k agents* is sufficient to ensure that the social law is effective with respect to φ. Formally, this involves checking that:

$$\forall C \subseteq A : (|C| \geq k) \quad \Rightarrow \quad (K \dagger (\eta \restriction C)) \models \varphi.$$

Example

Example (thanks to Dov Gabbay)

A senate with n members. Social law: follow the party line. The social law is robust in the sense that we can tolerate k rebels and still function towards our goals.

k-necessity

Let K and φ be given.

Definition

η is k-necessary (w.r.t. K, φ) iff:

$$\forall C \subseteq A : (K \dagger (\eta \restriction C)) \models \varphi \quad \Rightarrow \quad (|C| \geq k).$$

Resilience

We define the *resilience* of a social law η (w.r.t. K, φ) as the largest number of non-compliant agents the system can tolerate.

Definition

the resilience is the largest number k, $k \leq n$, such that

$$\forall C \subseteq A : (|C| \leq k) \quad \Rightarrow \quad (K \dagger (\eta \restriction A \setminus C)) \models \varphi.$$

where n is the number of agents.

Results

Theorems

Deciding k-sufficiency, k-necessity and resilience is co-NP-complete.

Contents

Focus on the Important Agents

- Both sufficient and neccessary agents are important. However, agents who (for example) are neither sufficient or necessary might have a very different degree of importance
  - For example, it might be that one agent ensures the goal when he joins *almost all* coalitions, but not *all* (hence, she is not sufficient)
- It makes sense to consider in more detail *how important* agents are to success/failure of the social law; how likely it is that their compliance or otherwise will affect the objective.
- We can then devote our attention to the *most important* agents.
- Idea: use *power indices* developed in coalitional game theory/voting theory for this purpose.

Coalitional Games

- A *cooperative game* is a pair

$$G = \langle A, \nu \rangle$$

where

  - $A = \{1, \ldots, n\}$ is a set of *players*, and
  - $\nu : 2^A \to \mathbb{R}$ is the *characteristic function* of the game, which assigns to every set of agents a numeric value, intuitively corresponding to the utility that this group of agents could obtain if they chose to cooperate.

- A game is *simple* if it gives 0,1 values only: if $\nu(C) = 0$ then C is losing, if $\nu(C) = 1$ then C are *winning*.

Power Indices

- *Power indices* characterise the *influence* that an agent has, by measuring how effective the agent is at turning a losing coalition into a winning coalition.
- Agent i is said to be a *swing player* for $C \subseteq A$ if C is not winning but $C \cup \{i\}$ is.
  Define a function $swing(C, i)$ (where $i \notin C$) so that this function returns 1 if i is a swing player for C, and 0 otherwise, i.e.,

$$swing(C, i) = \begin{cases} 1 & \text{if } \nu(C) = 0 \text{ and } \nu(C \cup \{i\}) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The Banzhaf Score

- The *Banzhaf score* for i, σ_i, is number of coalitions for which i is a swing player:

$$\sigma_i = \sum_{C \subseteq A \setminus \{i\}} swing(C, i). \quad (1)$$

The Banzhaf Measure

- The *Banzhaf measure*, denoted μ_i, is the probability that i would be a swing player for a coalition chosen at random from $2^{A \setminus \{i\}}$:

$$\mu_i = \frac{\sigma_i}{2^{n-1}} \quad (2)$$

The Banzhaf Index

- *Banzhaf index* for $i \in A$, denoted by η_i, is the proportion of coalitions for which i is a swing to the total number of swings in the game:

$$\eta_i = \frac{\sigma_i}{\sum_{j \in A} \sigma_j} \tag{3}$$

Power in Social Laws

- Idea: measure the power of agents by complying/not complying with a norm.
- Given K, η, φ, we obtain a simple coalitional game:

$$\nu_S(C) = \begin{cases} 1 & \text{if } K \dagger (\eta \restriction C) \models \varphi \\ 0 & \text{otherwise.} \end{cases}$$

In other words, a coalition "wins" if their compliance to η (and the others not complying) will make φ hold.

Complexity of the Banzhaf Score

Theorem

Given a social system $S = \langle K, \varphi, \eta \rangle$ and agent i in K, computing the Banzhaf score σ_i for i in the corresponding coalitional game $G(S)$ is #P-complete.

This is a very negative result: *worse than NP hardness*.

Complexity of Computing Power

Theorem

Given a social system $S = \langle K, \varphi, \eta \rangle$ and agent i in K, the following problems are #P-equivalent: computing the Banzhaf index η_i; and computing the Banzhaf measure μ_i.

Dummies and Dictators

We say that a player i is a *dictator* in a social system if $\mu_i = 1$, and a *dummy* if $\mu_i = 0$.

Theorem

Given a social system $S = \langle K, \varphi, \eta \rangle$ and agent i in K, the following problems are co-NP-complete: checking whether $\sigma_i = 0$; checking whether $\mu_i = 0$; checking whether $\mu_i = 1$; checking whether $\eta_i = 0$; checking whether $\eta_i = 1$; checking whether $\varsigma_i = 0$; and checking whether $\varsigma_i = 1$.

Measuring Relative Power

Given two agents $i, j \in A$ and a power index $M \in \{\sigma, \mu, \beta, \varsigma\}$, we write $i \succ_M j$ to mean $M_i > M_j$.

Theorem

Given a social system $S = \langle K, \varphi, \eta \rangle$, agents i, j in K, and power measure $M \in \{\sigma, \mu, \beta, \varsigma\}$, it is NP-hard to decide whether $i \succ_M j$.

Tractable Instances: Minimal Social Laws

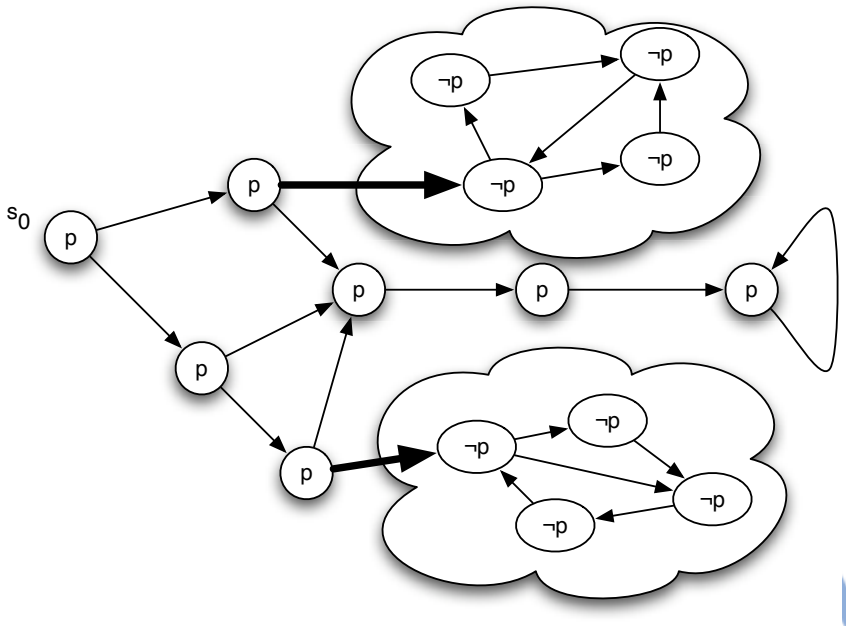
- A social law is *minimal* if no transitions can be eliminated without the norm failing.

Theorem

If $S = \langle K, \varphi, \eta \rangle$ is a minimal social system, then for each $i \in A(\eta)$, the values σ_i, μ_i, η_i, and ς_i are polynomial time computable.

Tractable Instances: Bridge Social Laws

Suppose your objective is to keep *p* true.



Tractable Instances: Bridge Social Laws

- A *bridge* social law is an easily identified type of minimal system: where we have a single transition (the bridge) leading to a “bad region” in which the objective is never satisfied.
- Bridge norms are minimal, and can be easily identified.
- Certain *tree-like* systems can also be seen to have minimal social laws, and easily computable power indices.

Contents

Some references I

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.

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.

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.