A Deontic Logic of Knowingly Complying

Carlos Areces, Valentin Cassano, Pablo F. Castro, Raul Fervari & Andrés R. Saravia

UNC, UNRC & CONICET, Argentina; GTIIT, China

AAMAS 2023 LONDON

Abstract

We introduce a logic for representing the deontic notion of *knowingly complying* –associated to an agent's conciousness of taking a normative course of action for achieving a certain goal. Our logic features operators:

- for describing normative courses of actions, and
- for describing what each agent knowingly complies with.

We provide:

- a sound and complete axiom system,
- the computational complexity of its satisfiability problem, and
- an extension with an additional operator for capturing abilities (with a sound and complete axiom system).

A Motivating Example

EMERGENCY PROCEDURE

– FIRE – SMOKE – EXPLOSION	KEEP CALM PULL FIRE ALARM, FROM A SAFE LOCATION CALL 999 (FIRE BRIGADE)

Evacuate: close doors behind, use only stairs or ramps.

If unsafe to evacuate: shut door, block cracks, stay low near window.

Figure 1: Fire Emergency Evacuation Plan

The main norms are:

- In the event of a fire/smoke/explosion, sound the nearest fire alarm, move to a safe location, and call the Fire Brigade.
- When evacuating close doors behind, use only stairs or ramps (never the elevator)
- Remain calm in any possible situation.

The DLKc logic

Definition 0.1. *Let* Prop *be a set for proposition symbols, and* Agt *a non-empty finite set of agent names. The* language *of* DLKc *is:*

$$\varphi := p \mid \neg\varphi \mid \varphi \vee \psi \mid \mathsf{N}(\psi,\varphi) \mid \mathsf{Kc}_i(\psi,\varphi),$$

where: $p \in \mathsf{Prop}$ *and* $i \in \mathsf{Agt}$. $\mathsf{A}\varphi = \mathsf{N}(\neg\varphi,\bot)$ *and* $\mathsf{E}\varphi = \neg\mathsf{A}\neg\varphi$. *Intuitively,* $\mathsf{N}(\psi,\varphi)$*:* "there is a normative course of action that brings about φ given ψ"; $\mathsf{Kc}_i(\psi,\varphi)$*:* "agent i knowingly complies with φ given ψ".

Definition 0.2 (LTS)**.** *A* LTS *is a tuple* $\mathfrak{L} = \langle \mathrm{S},\mathrm{R},\mathrm{V}\rangle$ *where:*

- S *is a non-empty set of states;*
- $\mathrm{R} = \{\mathrm{R}_a \subseteq \mathrm{S}^2 \mid a \in \mathsf{Act}\}$*; and*
- $\mathrm{V} : \mathrm{S} \to 2^{\mathsf{Prop}}$ *is an assignment function.*

Definition 0.3 (Plans)**.** *Let* Act *a set of basic actions,* $\pi \in \mathsf{Act}^*$. *For* $0 \leq k \leq |\pi|$, π_k *is the initial segment of* π *of length* k *and* $\pi[k]$ *is the* k^{th} *element of* π.

Definition 0.4 (Strong Executability)**.** *Let* $\mathfrak{L} = \langle \mathrm{S},\mathrm{R},\mathrm{V}\rangle$ *be an* LTS, $s \in \mathrm{S}$, $\pi \in \mathsf{Act}^*$. π *is* SE *at* s *iff for all* $k \in [0,|\pi|-1]$ *and all* $s' \in \mathrm{R}_{\pi_k}(s)$*, implies* $\mathrm{R}_{\pi[k+1]}(s') \neq \emptyset$. $\mathrm{SE}(\pi) = \{s \in \mathrm{S} \mid \pi$ *is SE at* $s\}$. $\Pi \subseteq \mathsf{Act}^*$, $\mathrm{SE}(\Pi) = \bigcap_{\pi\in\Pi}\mathrm{SE}(\pi)$.

Definition 0.5 (U-NLTS)**.** *An* uncertainty-based normative LTSs *(*U-NLTS*) is a tuple* $\mathfrak{M} = \langle \mathrm{S},\mathrm{R},\mathrm{V},\mathrm{U},\mathrm{N}\rangle$ *where:*

- $\mathfrak{L} = \langle \mathrm{S},\mathrm{R},\mathrm{V}\rangle$ *is an* LTS*;*
- $\mathrm{N} \subseteq \mathsf{Act}^*$ *s.t. there is* $\pi \in \mathrm{N}$ *with* $\mathrm{SE}(\pi) = \mathrm{S}$*;*
- $\mathrm{U} : \mathsf{Agt} \to 2^{2^{\mathsf{Act}^*}}$ *satisfies:*
  - $\emptyset \in \mathrm{U}(i)$*, and*
  - *for all* $\{\Pi,\Pi'\} \subseteq \mathrm{U}(i)$, $\Pi \neq \Pi'$ *implies* $\Pi \cap \Pi' = \emptyset$.

Definition 0.6 (Semantics)**.** *Let* $\mathfrak{M} = \langle \mathrm{S},\mathrm{R},\mathrm{V},\mathrm{U},\mathrm{N}\rangle$ *be a* U-NLTS, $s \in \mathrm{S}$, φ *and* ψ *formulas:*

$\mathfrak{M}, s \Vdash p$ *iff* $p \in \mathrm{V}(s)$,
$\mathfrak{M}, s \Vdash \neg\varphi$ *iff* $\mathfrak{M}, s \nVdash \varphi$,
$\mathfrak{M}, s \Vdash \varphi \vee \psi$ *iff* $\mathfrak{M}, s \Vdash \varphi$ *or* $\mathfrak{M}, s \Vdash \psi$,
$\mathfrak{M}, s \Vdash \mathsf{N}(\psi,\varphi)$ *iff exists* $\pi \in \mathrm{N}$ *such that*
(i) $[\![\psi]\!]^{\mathfrak{M}} \subseteq \mathrm{SE}(\pi)$ *and*
(ii) $\mathrm{R}_\pi([\![\psi]\!]^{\mathfrak{M}}) \subseteq [\![\varphi]\!]^{\mathfrak{M}}$,

$\mathfrak{M}, s \Vdash \mathsf{Kc}_i(\psi,\varphi)$ *iff exists* $\Pi \in \mathrm{U}(i)$ *such that*
(i) $\Pi \subseteq \mathrm{N}$,
(ii) $[\![\psi]\!]^{\mathfrak{M}} \subseteq \mathrm{SE}(\Pi)$*, and*
(iii) $\mathrm{R}_\Pi([\![\psi]\!]^{\mathfrak{M}}) \subseteq [\![\varphi]\!]^{\mathfrak{M}}$,

where $[\![\chi]\!]^{\mathfrak{M}} = \{s \in \mathrm{S} \mid \mathfrak{M}, s \Vdash \chi\}$.

A Motivating Example (cont.)

Fig. 1 can be represented as an U-NLTS $\mathfrak{M} = \langle \mathrm{S},\mathrm{R},\mathrm{V},\mathrm{U},\mathrm{N}\rangle$ s.t. the LTS $\mathfrak{L}$ part is modelled as in Fig. 2:

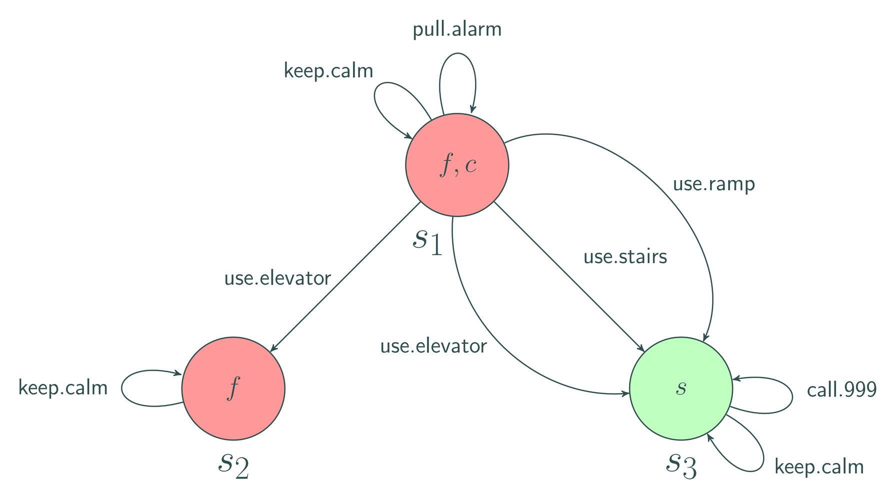

Figure 2: An LTS for the FEEP.

Where the basic actions are:

keep.calm, pull.alarm, call.999,

use.stairs, use.ramp, use.elevator.

The plans considered are:

$$\pi_0 = \mathsf{keep.calm}$$
$$\pi_r = \mathsf{pull.alarm}; \mathsf{use.ramp}; \mathsf{call.999}$$
$$\pi_s = \mathsf{pull.alarm}; \mathsf{use.stairs}; \mathsf{call.999}$$
$$\pi_e = \mathsf{pull.alarm}; \mathsf{use.elevator}; \mathsf{call.999}$$

And each of the states in S represents a different situation:

- s_1: a fire ocurrs (f), there is the capacity to follow the FEEP (c).
- s_2: a fire ocurrs (f), there is no capacity to follow the FEEP ($\neg c$).
- s_3: a safe location has been reached (s), there is no fire ($\neg f$).

Thus, $\mathrm{SE}(\pi_0) = \mathrm{S}$, $\mathrm{SE}(\pi_r) = \mathrm{SE}(\pi_s) = \{s_1\}$, and $\mathrm{SE}(\pi_e) = \emptyset$ and the set of normative plans is $\mathrm{N} = \{\pi_0, \pi_r, \pi_s\}$.

Suppose we have two agents:

- i: has taken an occupational safety course and knows the difference between using stairs/ramps and using the elevator.
  $$\mathrm{U}(i) = \{\emptyset, \{\pi_s, \pi_r\}, \{\pi_e\}\}$$
- j: has not taken the course and considers all possible ways of exiting the building might be equally good.
  $$\mathrm{U}(j) = \{\emptyset, \{\pi_s, \pi_e, \pi_r\}\}.$$

The following properties hold in $\mathfrak{M}$

(1) $\mathfrak{M}, s_1 \Vdash \mathsf{A}(s \to \neg f)$
(2) $\mathfrak{M}, s_1 \Vdash \mathsf{E}f$
(3) $\mathfrak{M}, s_1 \Vdash \mathsf{N}(f \wedge c, s)$
(4) $\mathfrak{M}, s_1 \Vdash \mathsf{Kc}_i(f \wedge c, s)$
(5) $\mathfrak{M}, s_1 \nVdash \mathsf{Kc}_j(f \wedge c, s)$

(1) and (2) are immediate. As a witness for (3) we can take the plan π_s. As a witness for (4) we can take the set $\{\pi_s, \pi_r\} \in \mathrm{U}(i)$. Failure of (5) obtains from the fact that $\{\pi_s, \pi_e, \pi_r\} \nsubseteq \mathrm{N}$ and $[\![f \wedge c]\!]^{\mathfrak{M}} \nsubseteq \mathrm{SE}(\emptyset) = \emptyset$.

Axiomatization & complexity

Axioms:	
Taut	$\vdash \varphi$ for φ a propositional tautology
DistA	$\vdash \mathsf{A}(\psi \to \varphi) \to (\mathsf{A}\psi \to \mathsf{A}\varphi)$
TA	$\vdash \mathsf{A}\varphi \to \varphi$
4KcA	$\vdash \mathsf{Kc}_i(\psi,\varphi) \to \mathsf{A}\mathsf{Kc}_i(\psi,\varphi)$
5KcA	$\vdash \neg\mathsf{Kc}_i(\psi,\varphi) \to \mathsf{A}\neg\mathsf{Kc}_i(\psi,\varphi)$
4NA	$\vdash \mathsf{N}(\psi,\varphi) \to \mathsf{A}\,\mathsf{N}(\psi,\varphi)$
5NA	$\vdash \neg\,\mathsf{N}(\psi,\varphi) \to \mathsf{A}\neg\,\mathsf{N}(\psi,\varphi)$
KcN	$\vdash \mathsf{Kc}_i(\psi,\varphi) \to \mathsf{N}(\psi,\varphi)$
DN	$\vdash \mathsf{N}(\varphi,\top)$
KcA	$\vdash (\mathsf{A}(\psi \to \chi) \wedge \mathsf{Kc}_i(\chi,\rho) \wedge \mathsf{A}(\rho \to \varphi)) \to \mathsf{Kc}_i(\psi,\varphi)$
NA	$\vdash (\mathsf{A}(\psi \to \chi) \wedge \mathsf{N}(\chi,\rho) \wedge \mathsf{A}(\rho \to \varphi)) \to \mathsf{N}(\psi,\varphi)$
Kc⊥	$\vdash \mathsf{Kc}_i(\bot,\bot)$
Rules:	
$\dfrac{\vdash \psi \quad \vdash (\psi \to \varphi)}{\vdash \varphi}$ (MP)	$\dfrac{\vdash \varphi}{\vdash \mathsf{A}\varphi}$ (Nec)

Table 1: Axiom system $\mathcal{DLKc}$ for DLKc over U-NLTSs.

Theorem 1. *The axiom system* $\mathcal{DLKc}$ *in Tab. 1 is sound and strongly complete for* DLKc *over the class of all* U-NLTS*s.*

Proposition 0.1. *The model checking problem for* DLKc *is in* P.

Theorem 2. *The satisfiability problem for* DLKc *is* NP*-complete.*

Reasoning About Abilities

Definition 0.7. *The language of* DLKc^+ *is defined by:*

$$\varphi := p \mid \neg\varphi \mid \varphi \vee \psi \mid \mathsf{S}(\psi,\varphi) \mid \mathsf{N}(\psi,\varphi) \mid \mathsf{Kc}_i(\psi,\varphi),$$

where: $p \in \mathsf{Prop}$ *and* $i \in \mathsf{Agt}$. *Intuitively,* $\mathsf{S}(\psi,\varphi)$*:* "there is a course of action that brings about φ given ψ".

Definition 0.8. *Let* $\mathfrak{M} = \langle \mathrm{S},\mathrm{R},\mathrm{V},\mathrm{U},\mathrm{N}\rangle$ *be a* U-NLTS, $s \in \mathrm{S}$, ψ *and* φ *formulas:*

$\mathfrak{M}, s \Vdash \mathsf{S}(\psi,\varphi)$ *iff exists* $\pi \in \mathsf{Act}^*$ *such that*
(i) $[\![\psi]\!]^{\mathfrak{M}} \subseteq \mathrm{SE}(\pi)$ *and*
(ii) $\mathrm{R}_\pi([\![\psi]\!]^{\mathfrak{M}}) \subseteq [\![\varphi]\!]^{\mathfrak{M}}$.

4SA	$\vdash \mathsf{S}(\psi,\varphi) \to \mathsf{A}\mathsf{S}(\psi,\varphi)$	NS	$\vdash \mathsf{N}(\psi,\varphi) \to \mathsf{S}(\psi,\varphi)$
5SA	$\vdash \neg\,\mathsf{S}(\psi,\varphi) \to \mathsf{A}\neg\,\mathsf{S}(\psi,\varphi)$	EmpS	$\vdash \mathsf{A}(\psi \to \varphi) \to \mathsf{S}(\psi,\varphi)$
		CompS	$\vdash (\mathsf{S}(\psi,\chi) \wedge \mathsf{S}(\chi,\varphi)) \to \mathsf{S}(\psi,\varphi)$

Table 2: Additional axioms for DLKc^+.

Theorem 3. *The axioms and rules in Tabs. 1 and 2 yield a sound and strongly complete axiom system for* DLKc^+ *over the class of all* U-NLTS*s.*

Future work

- Characterize the exact complexity of the satisfiability poblem of the extended logic DLKc^+.
- Establish different levels of responsibility for the agents using the relation between the set of plans $\mathrm{U}(i)$ of each agent and the set of norms N.
- Impose new restrictions on the different components of the model (or weakening them), and obtain new logics.

Acknowledgements

We thank the reviewers and the PC member for their constructive comments and suggestions. This work is partially supported by ANPCyT-PICT-2020-3780, CONICET PIP 11220200100812CO, the EU Grant Agreement 101008233 (MISSION), and by the Laboratoire International Associé SINFIN.