Neighborhood semantics for deontic and agency logics

Olga Pacheco

FAST Group
DI/CCTC, University of Minho

CIC'07
October, 2007

Non-ideal systems

- In complex sytems we may not have full control over the behaviour of all its components:
  - incomplete information,
  - "black box" components,
  - it´s too expensive or complex to do so,
  - humans are involved,...
- Thus, *failure* may occur and the system must be prepared to react to that.
- Non-ideality has to be taken as a natural ingredient, from first stages of development.
- Instead of describing how the system *will behave* we can only say how the system *should behave*:
  - **it is necessary** is replaced by **it is obligatory**,
  - **it is possible** is replaced by **it is permitted**.

Non-ideal systems

- Contract-based (normative) specification :
  - specify what is the obligatory and permitted behaviour (norms),
  - assume that components may deviate from that ideal behaviour (violate norms),
  - define what to do when violations to expected behaviour occur (sanctions, recovery procedures)
- Norms: represented by the set of obligations and permissions that result from them.
- Our aim: contribute with a high-level model and a logic to reason about it.

Non-ideal systems

Relevant concepts:

- We want to be able to speak about *obligations* and *permissions*.
- We are interested in: *obligation (and permission) to do* (as opposed to *obligation to be*).
- Obligations are fulfilled by agents through actions: "*Agent x is obliged to pay the debt*" meaning "*It is obligatory that agent x pays the debt*".
- So, we need an *agency* concept.
- We also need to relate *obligations* with *actions of agents*.

Non-ideal systems

As failure may occur, it is important to confront *expected behavior* (obligations, permissions, ...) with *actual behaviour* (actions of agents), detect *violations* of obligations (forbidden actions or not permitted actions) and identify *agents responsible for them*.

We will use deontic and agency logics.

Deontic Logic

Deontic modal language $\mathcal{L}_D(At)$ (At set of atomic propositions)

$\psi ::= p|\neg\psi|\psi \to \psi|O\psi \quad p \in At$

$\wedge, \vee, \leftrightarrow$ defined as usual.

$P\psi \stackrel{def}{=} \neg O \neg \psi.$

$O\ \phi$: "it is obligatory that ϕ"

- O: states what is obligatory to do, what ought to be done.
- P: states what is permitted.

SDL-Standard Deontic Logic

Axiomatics

PC Any axiomatization of proposition logic.

(K) $O(\psi \rightarrow \phi) \rightarrow (O\psi \rightarrow O\phi)$

(D) $O\psi \rightarrow \neg O\neg\psi$

(MP) $\frac{\psi \quad \psi\rightarrow\phi}{\phi}$

(Nec) $\frac{\psi}{O\psi}$

Axiom (D) tells that "what is obligatory is permitted" or, equivalently, that "there cannot exist conflicts of obligations": (D) $\neg(O\psi \wedge O\neg\psi)$.

SDL is a *KD* normal modal logic.

SDL: Paradoxes

SDL leads to well known paradoxes :Ross paradox, Chisholm paradox, gentle murder paradox,...

Questions rised by the "paradox of gentle murder" are relevant to our context.

SDL: Paradox of gentle murder

Statements:

- (1) Jones murders Smith.
- (2) Jones ought not to murder Smith.
- (3) If Jones murders Smith, then Jones ought to murder Smith gently.

Another fact:

- (4) If Jones murders Smith gently, then Jones murders Smith.

From (4) and (RM) rule we can infer:

- (5) If Jones ought to murder Smith gently, then Jones ought to murder Smith.

Fom (1) and (3) we have:

- (6) Jones ought to murder Smith gently.

And from (5) and (6) we infer

- (7) Jones ought to murder Smith.

which contradicts (2).

SDL: Paradoxes

Monotonicity

is the main cause for this paradox. We will need *weaker logics than K* in order to avoid undesirable inferences of this kind.

Other paradoxes are related with different problems: the representation of contrary to duties or conditional obligations, for instance.

Deontic logic

The deontic logic we use:

Axiomatics

(PC) Any axiomatization of proposition logic.

(D) $O\psi \rightarrow \neg O \neg \psi$

(MP) $$\frac{\psi \quad \psi \rightarrow \phi}{\phi}$$

(RE) $$\frac{\psi \leftrightarrow \phi}{O\psi \leftrightarrow O\phi}$$

This is a non-normal ED modal logic.

Deontic Logic

The semantic we adopt:

Semantics: neighbourhood deontic models

A neighborhood deontic frame F is a pair $F =< W, N_o >$ where W is a non-empty set of worlds and N_o is a neighborhood deontic function $N_o : W \longrightarrow \mathcal{P}(\mathcal{P}(W))$. A model based on F is a tuple $< W, N_o, V >$ where V is a valuation function $V : W \longrightarrow \mathcal{P}(At)$.

$N_o(w)$ assigns to each world the set of propositions obligatory in it.
Propositions are represented by its truth set:
$\parallel \psi \parallel_M = \{w | M, w \Vdash \psi\}$

Deontic Logic

Validity of formulas in a model:

- $M, w \Vdash p$ iff $p \in V(w)$
- $M, w \Vdash \neg\psi$ iff $M, w \nVdash \psi$
- $M, w \Vdash \psi \rightarrow \phi$ iff $M, w \nVdash \psi$ or $M, w \Vdash \phi$
- $M, w \Vdash O\psi$ iff $\| \psi \|_M \in N_o(w)$

$F \Vdash \psi$

A frame F validates a formula ψ if all models based on F validate ψ.

Deontic Logic

Some known results:

Properties of neighborhood deontic frames

Let $F =< W, N_o >$ be a neighborhood deontic frame. The axiom (D) defines a *proper frame*, i.e., $F \Vdash O\psi \rightarrow \neg O \neg \psi$ iff for all w, if $X \in N_o(w)$ then $(W - X) \notin N_o(w)$.

Agency Logic

Agency modal language $\mathcal{L}_A(At)$ (At set of atomic propositions)

$\psi ::= p|\neg\psi|\psi \rightarrow \psi|\{E_a\psi\}_{a\in Ag}$ where Ag is a set of agents and $p \in At$

$\wedge, \vee, \leftrightarrow$ defined as usual.

$E_i\ \phi$: "agent i brings about ϕ"

$E_i\ \phi$ relates the **agent** (actor, component, ...) i with the **state of affairs** ϕ he brings about, abstracting from the *concrete actions* done to obtain that *state of affairs* and putting aside temporal issues.

Agency logic

Axiomatics

PC Any axiomatization of proposition logic.

(T) $E_i\psi \rightarrow \psi$

(C) $E_i\psi \wedge E_i\phi \rightarrow E_i(\psi \wedge \phi)$

(MP) $\dfrac{\psi \quad \psi \rightarrow \phi}{\phi}$

(RE) $\dfrac{\psi \leftrightarrow \phi}{E_i\psi \leftrightarrow E_i\phi}$

This is a non-normal ETC modal logic.

Agency Logic

Semantics: neighbourhood agency models

A neighborhood agency frame F is a pair $F =< W, \{N_{e_i}\}_{i \in Ag} >$ where W is a non-empty set of worlds and N_{e_i} is a neighborhood agency function $N_{e_i} : W \longrightarrow \mathcal{P}(\mathcal{P}(W))$. A model based on F is a tuple $< W, \{N_{e_i}\}_{i \in Ag}, V >$ where V is a valuation function $V : W \longrightarrow \mathcal{P}(At)$.

$N_{e_i}(w)$ assigns to the world w the set of propositions the agent i brings about in w.

Validity of formulas in a neighborhood agency model:

- $M, w \Vdash p$ iff $p \in V(w)$
- $M, w \Vdash \neg\psi$ iff $M, w \nVdash \psi$
- $M, w \Vdash \psi \rightarrow \phi$ iff $M, w \nVdash \psi$ or $M, w \Vdash \phi$
- $M, w \Vdash E_i\psi$ iff $\parallel \psi \parallel_M \in N_{e_i}(w)$

Agency Logic

Some known results:

Properties of neighborhood agency frames

Let $F =< W, N_{e_i} >$ be a neighborhood agency frame.

- $F \Vdash E_i\psi \wedge E_i\phi \rightarrow E_i(\psi \wedge \phi)$ iff F is *closed under finite intersections* (i.e., if for any collection of sets $\{X_i\}_{i \in I}$ (I finite), for each $i \in I$, $X_i \in N_{e_i}(w)$, then $(\bigcap_{i \in I} X_i) \in N_{e_i}(w)$.
- $F \Vdash E_i\psi \rightarrow \psi$ iff for each $w \in W$, $N_{e_i}(w) \neq \emptyset$ and $w \in \bigcap N_{e_i}(w)$

Deontic and Agency Logic

Deontic and agency modal language $\mathcal{L}_{DA}(At)$ (At set of atomic propositions)

$\psi ::= p|\neg\psi|\psi \rightarrow \psi|O\psi|\{E_a\psi\}_{a\in Ag}$ where Ag is a set of agents and $p \in At$

$\wedge, \vee, \leftrightarrow$ defined as usual, P defined as above.

Deontic and Agency Logic

Logical properties:

PC Any axiomatization of proposition logic.

(MP) $\dfrac{\psi \quad \psi \rightarrow \phi}{\phi}$

(Te) $E_i\psi \rightarrow \neg\psi$

(Ce) $E_i\psi \wedge E_i\phi \rightarrow E_i(\psi \wedge \phi)$

(REe) $\dfrac{\psi \leftrightarrow \phi}{E_i\psi \leftrightarrow E_i\phi}$

(Do) $O\psi \rightarrow \neg O\neg\psi$

(REo) $\dfrac{\psi \leftrightarrow \phi}{O\psi \leftrightarrow O\phi}$

(Coe) $OE_i\psi \wedge OE_i\phi \rightarrow OE_i(\psi \wedge \phi)$

(Cop) $OE_i\psi \wedge PE_i\phi \rightarrow PE_i(\psi \wedge \phi)$

(RMep) $\dfrac{E_i\psi \rightarrow E_k\phi}{PE_i\psi \rightarrow PE_k\phi}$

Deontic and Agency Logic

Neighborhood deontic and agency models:

$M =< W, N_o, \{N_{e_i}\}_{i \in Ag}, V >$ where:

- $N_o : W \longrightarrow \mathcal{P}(\mathcal{P}(W))$
- $N_{e_i} : W \longrightarrow \mathcal{P}(\mathcal{P}(W))$
- $V : W \longrightarrow \mathcal{P}(At)$

Deontic and Agency Logic

We can reformulate a neighborhood function as follows:
$f_{\Box} : \mathcal{P}(W) \longrightarrow \mathcal{P}(W)$
$w \in f_{\Box}(X)$ iff $X \in N_{\Box}(w)$
$f_{\Box}(X)$ gives the set of worlds where X is necessary.

Thus:

- $f_{e_i}(X)$ gives the set of worlds where the agent i brings about (the proposition) X.
- $f_o(X)$ gives the set of worlds where (the proposition) X is obligatory.

Now we have: $\| \Box\psi \| = f_{\Box}(\| \psi \|)$ which facilitates the expression of the semantics of iterated modal operators (as composition of neighborhood functions).

- $\| \top \| = W$
- $\|\bot\| = \emptyset$
- $\| \neg\psi \| = W - \| \psi \|$
- $\| \psi \wedge \phi \| = \| \psi \| \cap \| \phi \|$
- $\| \psi \vee \phi \| = \| \psi \| \cup \| \phi \|$
- $\| \psi \rightarrow \phi \| = \| \psi \| \subseteq \| \phi \|$
- $\| E_i\psi \| = f_{e_i}(\| \psi \|)$
- $\| O\psi \| = f_o(\| \psi \|)$

Using this function, the semantic characterization of formulas is "closest" to the syntactic form of formulas.

Logical Formulas vs. Semantic Properties:

$$
\begin{array}{rl}
\text{Te} & f_{ei}(X) \subseteq X \\
\text{Ce} & f_{ei}(X) \cap f_{ei}(Y) \subseteq f_{ei}(X \cap Y) \\
\text{Do} & f_o(X) \cap f_o(W - X) = \emptyset \\
\text{Coe} & f_o(f_{ei}(X)) \cap f_o(f_{ei}(Y)) \subseteq f_o(f_{ei}(X \cap Y)) \\
\text{Cop} & (f_o(f_{ei}(X)) - f_o(W - f_{ei}(Y))) \cap f_o(W - f_{ei}(X \cap Y)) = \emptyset \\
\text{RMep} & \text{if } f_{ei}(X) \subseteq f_{ek}(Y) \text{ then} \\
 & f_o(W - f_{ek}(Y)) \subseteq f_o(W - f_{ei}(X))
\end{array}
$$

Analysis supported

Expressivity

- $OE_i\psi$ (obligatory actions)
- $PE_i\psi$ (permitted actions)
- $E_iE_k\psi$ (control)
- $E_iOE_k\psi$ (command)
- $E_iPE_k\psi$ (authorisation)
- $\dots$

We will restrict our attention here to the first two formula schemas.

Personal deontic operators

- $O_i\phi \stackrel{abv}{=} OE_i\phi$
- $P_i\phi \stackrel{abv}{=} PE_i\phi$

Analysis supported

- Verify if an action is permitted: $E_i\psi \wedge P_i\psi$
- Detect norm violations:
  - $O_i\psi \wedge E_i\neg\psi$
  - $\neg P_i\psi \wedge E_i\psi$
- Detect the fulfillment of some obligation: $O_i\psi \wedge E_i\psi$
- Recovery or sanctioning of agents involved (effects of actions):
  - $(O_i\psi \wedge E_i\neg\psi) \rightarrow O_i\phi$
  - $(O_i\psi \wedge E_i\neg\psi) \rightarrow \neg P_i\phi$
- Other effects:
  - representation: $E_i\psi \rightarrow E_k\psi$
  - conventional acts (count as): $E_i\psi \rightarrow E_i\phi$

Adding Context

Effects of an action depend on action context

- The same action done by the same agent may have different effects depending on the context where the action was done.
- Roles may capture context of action.
- Permissions and obligations depend on roles. An agent may have permission to do ψ when acting in a role and not have permission to do the same action when acting in a different role.

Adding Context

Effects of an action depend on action context

- The same action done by the same agent may have different effects depending on the context where the action was done.
- Roles may capture context of action.
- Permissions and obligations depend on roles. An agent may have permission to do ψ when acting in a role and not have permission to do the same action when acting in a different role.
- Action in a role: $E_{i:r}\psi$: "agent i playing role r brings about ψ".
- Distinction between roles and agents.

Additional expressivity

- $O_{i:r}\psi \stackrel{abv}{=} OE_{i:r}\psi$
- $P_{i:r}\psi \stackrel{abv}{=} PE_{i:r}\psi$
- $P_{i:r1}\psi \wedge \neg P_{i:r2}\psi$ or (contradictory permissions)
- $O_{i:r1}\psi \wedge O_{i:r2}\neg\psi$ (conflicting obligations)

Questions

- What about dynamics?
- Effects of actions are not instantaneous.
- What is the meaning of worlds and neighborhoods in specification?
- How to combine the logics?
- ...

Future work

- Add dynamics.
- Explore the fact that a neighborhood frame is a coalgebra for the contravariant powerset functor composed with itself 2^2. (c.f. work of Y. Venema, H. Hansen, C. Kupke, E. Pacuit)

Bibliography

- H. Arlo-Costa and E. Pacuit. First-order classical modal logic. *Studia Logica* 84, pp. 171-210, 2006.
- J. Carmo and O. Pacheco. Deontic and action logics for organized collective agency, modeled through institutionalized agents and roles. *Fundamenta Informaticae*, Vol.48 (No. 2,3), pp. 129-163, IOS Press, November, 2001.
- B. J. Chellas. *Modal Logic - an Introduction*, Cambridge University Press, 1980.
- H. Hansen. Monotonic modal logic (Master's thesis). Research Report PP-2003-24, ILLC, University of Amsterdam, 2003.
- O. Pacheco and J. Carmo: "A Role Based Model for the Normative Specification of Organized Collective Agency and Agents Interaction", *Journal of Autonomous Agents and Multi-Agent Systems*, Vol. 6, Issue 2, pp.145-184, Kluwer, 2003.