Defeasible Modalities

Ivan Varzinczak

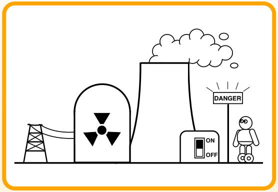
Centre for Artificial Intelligence Research
CSIR Meraka Institute and UKZN, South Africa

ijv@acm.org

(Joint work with Arina Britz)

TARK-13, Chennai

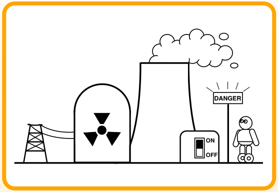
Motivation



Nuclear Power Plant

- Atomic pile, cooling system: On/Off
- Agent controls pile and cooler
- Detect hazardous situations

Motivation

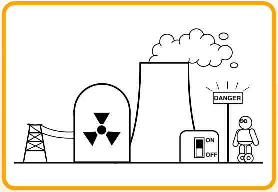


Nuclear Power Plant

- Atomic pile, cooling system: On/Off
- Agent controls pile and cooler
- Detect hazardous situations

- If *pile on*, then usually *not hazardous*
- If *pile on* and *cooling off*, then usually *hazardous*
- If *hazardous*, then usually after switching *not hazardous*
- If *pile on* and *cooling off*, then usually agent knows *hazardous*
- If *pile on*, then usually ought to be *not hazardous*
- An *atomic pile* usually has as component *some piece of Uranium*

Motivation



Nuclear Power Plant

- Atomic pile, cooling system: On/Off
- Agent controls pile and cooler
- Detect hazardous situations

- If *pile on*, then usually *not hazardous*
- If *pile on* and *cooling off*, then usually *hazardous*
- If *hazardous*, then usually after switching *not hazardous*
- If *pile on* and *cooling off*, then usually agent knows *hazardous*
- If *pile on*, then usually ought to be *not hazardous*
- An *atomic pile* usually has as component *some piece of Uranium*

Central notions: *Normality*, *typicality*, *plausibility*, *preferences*, etc.

Motivation

Accounts of Normality

$$\alpha \Rightarrow_{C} \beta, \;\; \alpha \Rightarrow_{B} \beta, \;\; \alpha \mid\!\sim \beta, \;\; \mathsf{N}(\alpha \rightarrow \beta), \;\; \ldots$$

Motivation

Accounts of Normality

$$\alpha \Rightarrow_C \beta,\ \ \alpha \Rightarrow_B \beta,\ \ \alpha \mathrel{|\!\sim} \beta,\ \ \mathsf{N}(\alpha \rightarrow \beta),\ \ \ldots$$

"The most normal α-worlds are β-worlds"

"If a world is a normal α-world, then it is a β-world"

Motivation

Accounts of Normality

$$\alpha \Rightarrow_C \beta, \ \ \alpha \Rightarrow_B \beta, \ \ \alpha \mid\!\sim \beta, \ \ \mathsf{N}(\alpha \rightarrow \beta), \ \ \ldots$$

"The most normal α-worlds are β-worlds"

"If a world is a normal α-world, then it is a β-world"

Other Aspects of Defeasibility

- "A *normal toggling* of the switch turns the light on"
- "We *sort of know* that the speed of light is the fastest"
- "In soccer, it is my *normal duty* to be fair play"
- "Cells that *naturally constitute* my body are human"

Motivation

Accounts of Normality

$$\alpha \Rightarrow_C \beta, \ \ \alpha \Rightarrow_B \beta, \ \ \alpha \mid\!\sim \beta, \ \ \mathsf{N}(\alpha \rightarrow \beta), \ \ \ldots$$

"The most normal α-worlds are β-worlds"

"If a world is a normal α-world, then it is a β-world"

Other Aspects of Defeasibility

- "A *normal toggling* of the switch turns the light on"
- "We *sort of know* that the speed of light is the fastest"
- "In soccer, it is my *normal duty* to be fair play"
- "Cells that *naturally constitute* my body are human"

Need to formalize these notions. Importantly, in a general and simple way!

Outline

Outline

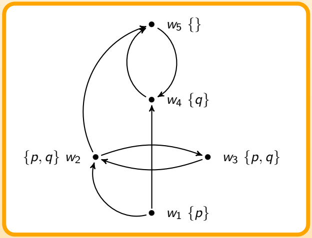
Modal Logic

Formulas

- $\alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \mid \Box_i\alpha \mid \Diamond_i\alpha$

Kripke models: tuples $\mathscr{M} = \langle W, R, V \rangle$ where

- $W \neq \emptyset$ is a set of possible *worlds*
- $R = \langle R_1, \ldots, R_n \rangle$, where $R_i \subseteq W \times W$ is an *accessibility relation*
- $V : W \longrightarrow 2^{\mathcal{P}}$ is a *valuation* function



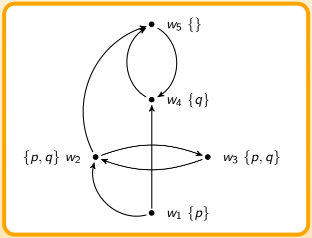
Modal Logic

Formulas

- $\alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \to \alpha \mid \Box_i\alpha \mid \Diamond_i\alpha$

Kripke models: tuples $\mathscr{M} = \langle W, R, V \rangle$ where

- $W \neq \emptyset$ is a set of possible *worlds*
- $R = \langle R_1, \ldots, R_n \rangle$, where $R_i \subseteq W \times W$ is an *accessibility relation*
- $V : W \longrightarrow 2^{\mathcal{P}}$ is a *valuation* function



Satisfaction:

- $\Box_i\alpha$: "α holds in all accessible worlds"
- $\Diamond_i\alpha$: "α holds in some accessible world"
- Boolean formulas: as usual

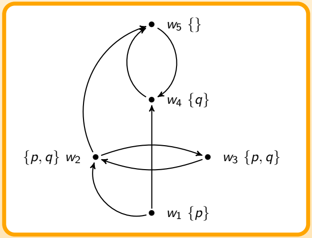
Modal Logic

Formulas

- $\alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \mid \Box_i\alpha \mid \Diamond_i\alpha$

Kripke models: tuples $\mathscr{M} = \langle W, R, V \rangle$ where

- $W \neq \emptyset$ is a set of possible *worlds*
- $R = \langle R_1, \ldots, R_n \rangle$, where $R_i \subseteq W \times W$ is an *accessibility relation*
- $V : W \longrightarrow 2^{\mathcal{P}}$ is a *valuation* function



Truth, validity, entailment:

- α is *true* in $\mathscr{M}$ iff every world satisfies α
- α is *valid* iff α is true in every $\mathscr{M}$
- $\alpha \models \beta$ iff every α-world is a β-world

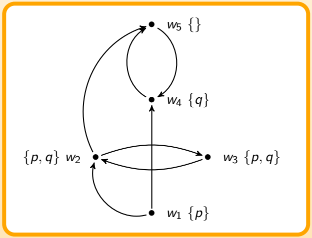
Modal Logic

Formulas

- $\alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \mid \Box_i\alpha \mid \Diamond_i\alpha$

Kripke models: tuples $\mathscr{M} = \langle W, R, V \rangle$ where

- $W \neq \emptyset$ is a set of possible *worlds*
- $R = \langle R_1, \ldots, R_n \rangle$, where $R_i \subseteq W \times W$ is an *accessibility relation*
- $V : W \longrightarrow 2^{\mathcal{P}}$ is a *valuation* function



Properties:

- $\models \Diamond\alpha \leftrightarrow \neg\Box\neg\alpha$
- $\models \Box(\alpha \rightarrow \beta) \rightarrow (\Box\alpha \rightarrow \Box\beta)$
- More depending on the application

Outline

Defeasible Modalities

Defeasible versions of modalities: $\boxdot$ ('flag') and $\Diamonddot$ ('flame')

Extended language

- $\alpha ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \mid \Box_i\alpha \mid \Diamond_i\alpha \mid \boxdot_i\alpha \mid \Diamonddot_i\alpha$

Intuition

- $\boxdot_i\alpha$: "all most normal i-successors are α" (Normal Necessity)
- $\Diamonddot_i\alpha$: "some most normal i-successors are α" (Distinct Possibility)

Example

- $\boxdot_s pile$, $\boxdot_A(\neg cooler \rightarrow hazardous)$
- $\boxdot_A(pile \rightarrow cooler)$, $\Diamonddot_A(hazardous \leftrightarrow (pile \wedge \neg cooler))$
- $(pile \wedge \neg cooler) \rightarrow (\boxdot_A hazardous \wedge \Diamonddot_m \top)$

Defeasible Modalities: Which Semantics?

Stalnaker's system of conditional logic

- *Selection function* picking out the closest (most plausible) world to w

$$f : \mathcal{L} \times W \longrightarrow W$$

- However, it assumes uniqueness of $f(\alpha, w)$ (we want more than one)

Lewis's systems of conditional logic

- Uniqueness is dropped
- Some systems require MP (unwanted in default reasoning)

$$\frac{\alpha,\ \alpha \Rightarrow \beta}{\beta}$$

- As many preferences as there are possible worlds

Defeasible Modalities: Which Semantics?

Boutilier's (1994) conditional logics of normality

- Modalities remain classical
- $\alpha \Rightarrow \beta$ holds in w iff it holds everywhere (OK for conditionals)

Baltag & Smets's (2008) approach

- *Plausibility models* $\mathscr{M} = \langle W, \sim, V, \leq \rangle$ in epistemic / doxastic context
- New modalities defined based on the preference $\leq$
- Actions, obligations, ontologies, ...?

Defeasible Modalities: Which Semantics?

We want a framework

- That is general yet elegant
- That is simple
- That accounts for defeasible argument forms like $\alpha \mid\!\sim \beta$

KLM Approach: Kraus, Lehmann & Magidor (1990,1992)

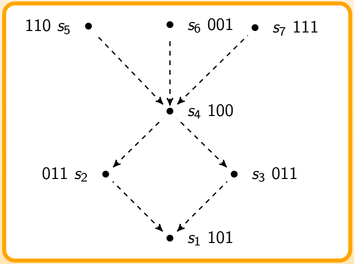
- Successful in the propositional case
- Provides a general proof-theoretic characterization of $\mid\!\sim$
- Basis for the important notion of *rational closure*
- Recently extended to modal and description logics [BMV11,BMV12]

Outline

Preferential Semantics [Shoham88; KLM90]

Preferential models: tuples $\mathscr{P} = \langle S, \ell, \prec \rangle$ where

- S is a set of states
- ℓ labels states with valuations
- $\prec \subseteq S \times S$ is a smooth ($\approx$ well-founded) partial order over S
- For all α, $[\![\alpha]\!] := \{s \in S \mid \ell(s) \Vdash \alpha\}$ has a minimal element



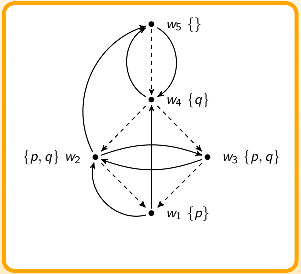
Given $\mathscr{P} = \langle S, \ell, \prec \rangle$

- α is *satisfiable* iff $[\![\alpha]\!] \neq \emptyset$
- α is *true* in $\mathscr{P}$ ($\mathscr{P} \Vdash \alpha$) iff $[\![\alpha]\!] = S$
- $\mathscr{P} \Vdash \alpha \mid\!\sim \beta$ iff $\min_{\prec}[\![\alpha]\!] \subseteq [\![\beta]\!]$
- $p \mid\!\sim \neg h, \quad p \land \neg c \mid\!\sim h$

Preferential Semantics for Defeasible Modalities

Preferential Kripke models: tuples $\mathscr{P} = \langle W, R, V, \prec \rangle$ where

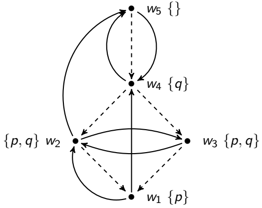
- W, R, V as before
- $\prec \subseteq W \times W$ is a smooth partial order on W



Preferential Semantics for Defeasible Modalities

Preferential Kripke models: tuples $\mathscr{P} = \langle W, R, V, \prec \rangle$ where

- W, R, V as before
- $\prec \subseteq W \times W$ is a smooth partial order on W



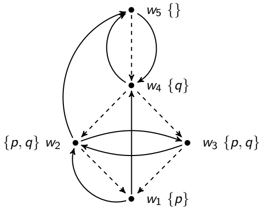
Satisfaction

- Modal sentences: as before
- $[\![\widetilde{\Box}_i \alpha]\!] := \{ w \in W \mid \min_{\prec} R_i(w) \subseteq [\![\alpha]\!] \}$
- $[\![\widetilde{\Diamond}_i \alpha]\!] := \{ w \in W \mid \min_{\prec} R_i(w) \cap [\![\alpha]\!] \neq \emptyset \}$

Preferential Semantics for Defeasible Modalities

Preferential Kripke models: tuples $\mathscr{P} = \langle W, R, V, \prec \rangle$ where

- W, R, V as before
- $\prec \subseteq W \times W$ is a smooth partial order on W



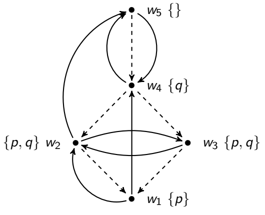
Truth, validity, etc.

- α is *satisfiable* iff $[\![\alpha]\!] \neq \emptyset$
- α is *true* in $\mathscr{P}$ ($\mathscr{P} \Vdash \alpha$) iff $[\![\alpha]\!] = W$
- α is *valid* iff $\mathscr{P} \Vdash \alpha$ for every $\mathscr{P}$

Preferential Semantics for Defeasible Modalities

Preferential Kripke models: tuples $\mathscr{P} = \langle W, R, V, \prec \rangle$ where

- W, R, V as before
- $\prec \subseteq W \times W$ is a smooth partial order on W



Truth, validity, etc.

- α is *satisfiable* iff $[\![\alpha]\!] \neq \emptyset$
- α is *true* in $\mathscr{P}$ ($\mathscr{P} \Vdash \alpha$) iff $[\![\alpha]\!] = W$
- α is *valid* iff $\mathscr{P} \Vdash \alpha$ for every $\mathscr{P}$

Lemma

All modal validities and rules of inference are preserved

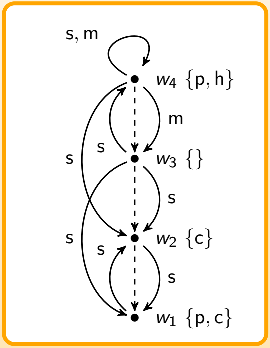
Preferential Semantics for Defeasible Modalities

$[\![\Box\!\!\!\sim_i \alpha]\!] := \{w \in W \mid \min_{\prec} R_i(w) \subseteq [\![\alpha]\!]\}$

$[\![\Diamond\!\!\!\sim_i \alpha]\!] := \{w \in W \mid \min_{\prec} R_i(w) \cap [\![\alpha]\!] \neq \emptyset\}$

Example in reasoning about actions

- Let $\mathcal{P} = \{\mathsf{p}, \mathsf{c}, \mathsf{h}\}$ and $\mathcal{A} = \{\mathsf{s}, \mathsf{m}\}$



In this model

- $\mathscr{P} \Vdash (\mathsf{p} \land \neg \mathsf{c}) \leftrightarrow \mathsf{h}$
- $w_4 \in [\![\mathsf{h} \land \Diamond\!\!\!\sim_{\mathsf{s}} \neg \mathsf{h}]\!]$
- $w_1 \in [\![\Box\!\!\!\sim_{\mathsf{m}} \bot]\!]$
- $\mathscr{P} \Vdash \neg \mathsf{p} \to \Box\!\!\!\sim_{\mathsf{s}} \mathsf{p}$ but $\mathscr{P} \not\Vdash \neg \mathsf{p} \to \Box_{\mathsf{s}} \mathsf{p}$
- $\mathscr{P} \Vdash \mathsf{c} \to \Box\!\!\!\sim_{\mathsf{s}} \neg \mathsf{h}$

Preferential Semantics for Defeasible Modalities

A few validities

- $\models \widetilde{\Box}_i\alpha \leftrightarrow \neg\widetilde{\Diamond}_i\neg\alpha$
- $\models \widetilde{\Box}_i\bot \leftrightarrow \Box_i\bot$
- $\models \widetilde{\Box}_i\top \leftrightarrow \top$
- $\models \widetilde{\Box}_i(\alpha \rightarrow \beta) \rightarrow (\widetilde{\Box}_i\alpha \rightarrow \widetilde{\Box}_i\beta)$ $\quad (NK)$
- $\models \widetilde{\Box}_i(\alpha \wedge \beta) \leftrightarrow (\widetilde{\Box}_i\alpha \wedge \widetilde{\Box}_i\beta)$ $\quad (NR)$
- $\models \Box_i\alpha \rightarrow \widetilde{\Box}_i\alpha$ $\quad (N)$

A few rules

$$(RNN)\quad \frac{\alpha}{\widetilde{\Box}_i\alpha} \qquad\qquad (NRK)\quad \frac{(\alpha_1 \wedge \ldots \wedge \alpha_k) \rightarrow \beta}{(\widetilde{\Box}_i\alpha_1 \wedge \ldots \wedge \widetilde{\Box}_i\alpha_k) \rightarrow \widetilde{\Box}_i\beta}$$

Outline

Entailment

Knowledge bases $\mathcal{K}$

- Arbitrary sets of $\mathcal{L}$-formulas
- $\mathscr{P} \Vdash \mathcal{K}$ iff $\mathscr{P} \Vdash \alpha$ for every $\alpha \in \mathcal{K}$

Definition

$\mathcal{K}$ *entails* α (denoted $\mathcal{K} \models \alpha$) iff for every $\mathscr{P}$, if $\mathscr{P} \Vdash \mathcal{K}$, then $\mathscr{P} \Vdash \alpha$

Theorem

Let $Cn(\mathcal{K}) \equiv_{\text{def}} \{\alpha \mid \mathcal{K} \models \alpha\}$. Then

- $\mathcal{K} \subseteq Cn(\mathcal{K})$ (Inclusion)
- $Cn(\mathcal{K}) = Cn(Cn(\mathcal{K}))$ (Idempotency)
- If $\mathcal{K}_1 \subseteq \mathcal{K}_2$, then $Cn(\mathcal{K}_1) \subseteq Cn(\mathcal{K}_2)$ (Monotonicity)

Tableaux for Defeasible Modalities

$$(\bot)\ \frac{n :: \alpha,\ n :: \neg\alpha}{n :: \bot} \qquad (\neg)\ \frac{n :: \neg\neg\alpha}{n :: \alpha} \qquad (\wedge)\ \frac{n :: \alpha \wedge \beta}{n :: \alpha,\ n :: \beta} \qquad (\vee)\ \frac{n :: \neg(\alpha \wedge \beta)}{n :: \neg\alpha \mid n :: \neg\beta}$$

$$(\Box_i)\ \frac{n :: \Box_i\alpha\ ;\ n \xrightarrow{i} n'}{n' :: \alpha} \qquad (\Diamond_i)\ \frac{n :: \neg\Box_i\alpha}{n'^{\star} :: \neg\alpha\ ;\ \Gamma'_1 \mid n'^{\star} :: \neg\alpha\ ;\ \Gamma'_2},\ \text{where:}$$

$$\Gamma'_1 = \{n \xrightarrow{i} n'^{\star},\ n'^{\star} \in \min_{\prec} \Sigma_i(n)\}$$
$$\Gamma'_2 = \{n \xrightarrow{i} n'^{\star},\ n \xrightarrow{i} n''^{\star},\ n''^{\star} \prec n'^{\star},\ n''^{\star} \in \min_{\prec} \Sigma_i(n)\}$$

$$(\widetilde{\Box}_i)\ \frac{n :: \widetilde{\Box}_i\alpha\ ;\ n \xrightarrow{i} n',\ n' \in \min_{\prec} \Sigma_i(n)}{n' :: \alpha} \qquad (\widetilde{\Diamond}_i)\ \frac{n :: \neg\widetilde{\Box}_i\alpha}{n'^{\star} :: \neg\alpha\ ;\ n \xrightarrow{i} n'^{\star},\ n'^{\star} \in \min_{\prec} \Sigma_i(n)}$$

Theorem

The tableau calculus for defeasible modalities is sound and complete with respect to the modal preferential semantics

Outline

Discussion and Related Work

Comparison with Baltag & Smets's approach

- *Safe belief* operator $\Box_{\leq}$
- $w \in [\![\Box_{\leq}\alpha]\!]$ iff for every w' s.t. $w' \leq w$, $w' \in [\![\alpha]\!]$
- $\Box_{\leq}\alpha$ is true in w iff α is true in all better accessible worlds
- $\mathrel{\approx\!\!\!\!\Box}\alpha$ is true in w iff α is true in all best accessible worlds

The $\mathrel{\approx\!\!\!\!\Box}$-logic as a conditional logic

- For each w, let f pick out the minimal accessible worlds
- $\mathrel{\approx\!\!\!\!\Box}\alpha$ is then $\top \Rightarrow \alpha$
- For each $\alpha \in \mathcal{L}$, let $R_{\alpha} := \{(w, w') \mid w \Vdash \alpha \text{ and } w' \Vdash \alpha\}$
- $\alpha \Rightarrow \beta$ is then $\mathrel{\approx\!\!\!\!\Box}_{\alpha}\beta$

Outline

Conclusion

What we have done

- Moved beyond defeasible argument forms
- Definition of defeasible versions of classical modalities
- Moved beyond the (propositional) KLM approach
- Provision of a core formalism for further extensions
- Tableau method for defeasible modalities
- Results transfer to other similarly structured logics (DLs)

To do list

- Generalization to a multi-preference setting
- Further restrictions on the semantics
- Study of specific modal systems and respective properties

References

- K. Britz, T. Meyer & I. Varzinczak. Preferential Reasoning for Modal Logic. In *Proc. of Methods for Modalities*. Osuna, Spain, 2011.
- R. Booth, T. Meyer & I. Varzinczak. PTL: A Propositional Typicality Logic. In *Proc. of JELIA*. Toulouse, France, 2012.
- K. Britz, T. Meyer & I. Varzinczak. Normal Modal Preferential Consequence. In *Proc. of Australasian AI Conf.* Sydney, Australia, 2012.

For more

`http://cair.meraka.org.za`

Thank you!