Conditional Obligations in Justification Logic

F. Faroldi, A. Rohani and T. Studer

Logic and Theory Group
University of Bern, Switzerland

WoLLIC 2023

Presention outline

What is deontic logic?

- the word "deontic" is given from Greek expression "*deon*", means what is binding or proper
- logic of normative concepts, norm systems and normative reasoning
- normative concepts: obligation, permission, prohibition

Standard Deontic Logic (SDL)

Axiom schemas and rule schemas

$\vdash A$ where A is a propositional tautology	(CL)
$\bigcirc(A \to B) \to \bigcirc A \to \bigcirc B$	($\bigcirc$-K)
$\bigcirc A \to \neg \bigcirc \neg A$	($\bigcirc$-D)
if $\vdash A$ and $\vdash A \to B$ then $\vdash B$	(MP)
if $\vdash A$ then $\vdash \bigcirc A$	($\bigcirc$-Nec)

Semantics

Relational model $\mathscr{M} = (W, R, V)$, where R is serial, i.e.,

$$(\forall w \in W)(\exists v \in W)(wRv)$$

Chisholm's Puzzle

1. Thomas should take the math exam. (primary obligation)
2. If he takes the math exam, he should register for it. (ATD:university rule)
3. If he does not take the math exam, he should not register for the math exam. (CTD)
4. He does not take the math exam.

(1.1) $\bigcirc E$	(2.1) $\bigcirc E$
(1.2) $E \to \bigcirc R$	(2.2) $\bigcirc (E \to R)$
(1.3) $\bigcirc (\neg E \to \neg R)$	(2.3) $\neg E \to \bigcirc \neg R$
(1.4) $\neg E$	(2.4) $\neg E$

(3.1) $\bigcirc E$	(4.1) $\bigcirc E$
(3.2) $\bigcirc (E \to R)$	(4.2) $E \to \bigcirc R$
(3.3) $\bigcirc (\neg E \to \neg R)$	(4.3) $\neg E \to \bigcirc \neg R$
(3.4) $\neg E$	(4.4) $\neg E$

Presention outline

1. Standard Deontic Logic
2. Dyadic Deontic Logic
3. Justification Logic
4. Dyadic deontic system in Justification Logic
5. Soundness and Completeness of $\mathsf{JE}_{\mathsf{CS}}$

Dyadic Deontic Logic (DDL)

Dyadic conditional

$$\bigcirc(B/A) \quad \underleftrightarrow{\text{weaker than}} \quad A \to \bigcirc B$$

$$\bigcirc(B/A) \quad \underleftrightarrow{\text{is read as}} \quad B \text{ (consequent) is obligatory, given } A \text{ (antecedent)}$$

Formulas

$$F ::= P_i \mid \neg F \mid F \to F \mid \Box F \mid \bigcirc(F/F).$$

$\Box F$	F is setteled as true	
$\bigcirc(G/F)$	G is obligatory, given F	
$P(G/F)$	G is permitted, given F	$\neg\bigcirc(\neg G/F)$
$\bigcirc F$	F is unconditionally obligatory	$\bigcirc(F/\top)$
PF	F is unconditionally permitted	$P(F/\top)$
$\Diamond F$	short for	$\neg\Box\neg F$

Proof Systems for Alethic-Deontic Logic

System E

Axioms of classical propositional logic	CL
S5-scheme axioms for $\Box$	
$\bigcirc(B/A) \to \Box\bigcirc(B/A)$	(Abs)
$\Box A \to \bigcirc(A/B)$	(Nec)
$\Box(A \leftrightarrow B) \to (\bigcirc(C/A) \leftrightarrow \bigcirc(C/B))$	(Ext)
$\bigcirc(A/A)$	(Id)
$\bigcirc(C/A \wedge B) \to \bigcirc(B \to C/A)$	(Sh)
$\bigcirc(B \to C/A) \to (\bigcirc(B/A) \to \bigcirc(C/A))$	(COK)

$$\frac{A \quad A \to B}{B}\ \text{(MP)} \qquad\qquad \frac{A}{\Box A}\ \text{(Nec)}$$

S5 axioms:

- (K): $\Box(A \to B) \to (\Box A \to \Box B)$
- (T): $\Box A \to A$
- (5): $\Diamond A \to \Box\Diamond A$

Semantics

Preference Model

A *preference model* is a triple $\mathcal{M} = \langle W, \preceq, v \rangle$ where:

- W is a non-empty set of worlds;
- $\preceq \subseteq W \times W$, such that: $w_1 \preceq w_2$: world w_2 is at least as good as world w_1.
- $v : P_i \rightarrow \mathscr{P}(W)$

Truth under Preference Model

Given $\mathcal{M} = \langle W, \preceq, v \rangle$, and $w \in W$

- $\mathcal{M}, w \Vdash P_i$ iff $w \in v(P_i)$
- $\mathcal{M}, w \Vdash \neg A$ iff $\mathcal{M}, w \nVdash A$
- $\mathcal{M}, w \Vdash A \rightarrow B$ iff $\mathcal{M}, w \Vdash \neg A$ or $\mathcal{M}, w \Vdash B$
- $\mathcal{M}, w \Vdash \Box A$ iff $\forall v \in W, \mathcal{M}, v \Vdash A$
- $\mathcal{M}, w \Vdash \bigcirc(B/A)$ iff $\text{best}_{\preceq} \|A\|^{\mathcal{M}} \subseteq \|B\|^{\mathcal{M}}$

Truth Set

Let $\mathscr{M} = \langle W, \preceq, v \rangle$ be a preference model. The *truth set* of $F \in \mathsf{Fm}$ is :

$$\|F\| = \{w \in W \mid \mathscr{M}, w \Vdash F\}$$

$\text{best}_{\preceq}\|F\|$: best worlds in which F is true, according to $\preceq$.

Two notions of "best"

There are two ways to formalize the notion of "best world" respecting optimality and maximality:

- $\text{best}\|A\|$ under opt rule:

$$opt_{\preceq}(\|A\|) = \{w \in \|A\| \ s.t. \ \forall v(\mathscr{M}, v \Vdash A \rightarrow v \preceq w)\}$$

- $\text{best}\|A\|$ under max rule:

$$max_{\preceq}(\|A\|) = \{w \in \|A\| \ s.t. \ \forall v((\mathscr{M}, w \Vdash A \wedge w \preceq v) \rightarrow v \preceq w)\}$$

Properties of $\preceq$

We consider folowing properties on the relation $\preceq$, which can hold in a model:

- reflexivity: for all $w \in W, w \preceq w$;
- limitedness: if $\|A\| \neq \emptyset$ then $best(\|A\|) \neq \emptyset$
- transitivity: for all $w, v, u \in W$, if $v \preceq w$ and $u \preceq v$, then $u \preceq w$.
- totalness: for all $w, v \in W, v \preceq w$ or $w \preceq v$.

Lemma

$max_{\preceq}(\|A\|) = opt_{\preceq}(\|A\|)$ iff $\preceq$ is total.

Chisholm's Set Revisited

1. Thomas should take the math exam. (Primary obligation)
2. If he takes the math exam, he should register for it. (ATD)
3. If he does not take the math exam, he should not register for it. (CTD)
4. He does not take the math exam.

$$\Gamma := \{\bigcirc E, \bigcirc(R/E), \bigcirc(\neg R/\neg E), \neg E\}$$

best	$\bullet w_1, R, E$	
2nd best	$\bullet w_2, E$	$\bullet w_3$
worst	$\bullet w_4, R$	

Factual detachment (FD) and strong factual detachment (SFD)

Factual Detachment (FD)

$$(\bigcirc(A/B) \wedge B) \rightarrow \bigcirc A$$

is not valid in DDL.
strong factual detachment (SFD)

$$(\bigcirc(A/B) \wedge \Box B) \rightarrow \bigcirc A$$

is valid in DDL.

Example

1. It's obligatory to pay fine in case someone doesn't pay the tax. ($\bigcirc(F/\neg T)$)
2. The deadline for paying taxes is over and someone didn't pay the tax. ($\Box\neg T$)
3. from (1) and (2) and SFD we conclude that it's obligatory for this person to pay the fine. ($\bigcirc F$)

Presention outline

1. Standard Deontic Logic
2. Dyadic Deontic Logic
3. Justification Logic
4. Dyadic deontic system in Justification Logic
5. Soundness and Completeness of $\mathsf{JE}_{\mathsf{CS}}$

Justification logic

$\Box A$	*justification* 	$t : A$
A is known		*t justifies the agent's knowledge of A*
A is obligatory		*A is obligatory for reason t*

- Logic of Proofs: Artemov: a classical provability semantics for S4 (and thus also for intuitionistic logic).
- combining justifications with traditional possible world models: epistemic and deontic contexts.
- Why justification logic is a proper candidate for deontic context? Hyperintensional by nature, consistency of obligations,...

Axioms schema for Logic of Proofs

Definition

Axioms of Classical Propositional Logic	CL
$\lambda : (F \to G) \to (\kappa : F \to \lambda \cdot \kappa : G)$	j
$(\lambda : F \vee \kappa : F) \to (\lambda + \kappa) : F$	j+
$\lambda : F \to F$	jt
$\lambda : F \to !\lambda : \lambda : F$	j4

The set of *proof terms*:

$$\lambda ::= \alpha_i \mid \xi_i \mid (\lambda + \lambda) \mid (\lambda \cdot \lambda) \mid !\lambda$$

Formulas are inductively defined as follows:

$$F ::= P_i \mid (F \to F) \mid \lambda : F \ ,$$

Justification logics are parametrized by a constant specification, which is a set

Constant specification

$$\mathsf{CS} \subseteq \{(c,A) \mid c \text{ is a constant justification term and } A \text{ is an axiom of justification logic}\}.$$

Rule of necessitation $\xrightarrow{\text{axiom necessitation}}$ $c : A$ if $(c,A) \in \mathsf{CS}$.

A constant specification CS is called *axiomatically appropriate* if for each axiom A there is a constant c such that $(c,A) \in \mathsf{CS}$.
In epistemic settings, we can calibrate the reasoning power of the agents by adapting the constant specification.

Ross' Paradox

You ought to mail the letter. $(\Box A)$ (1)

implies

You ought to mail the letter or burn it. $(\Box(A \vee B))$ (2)

It is a classical validity that

you mail the letter implies you mail the letter or burn it. $(A \rightarrow A \vee B)$ (3)

By the monotonicity rule we find that (1) implies (2).

Fardoli and Protopopescu avoid this paradox by restricting the constant specification such that although (3) is a logical validity, there will no justification term for it. Thus the rule of monotonicity cannot be derived and there is no paradox.

Hyperintensionality

Faroldi claims that deontic modalities are *hyperintensional*, i.e. they can distinguish between logically equivalent formulas.

Example

Consider the following sentences:

$$\text{You ought to drive. } (\Box A) \tag{4}$$

$$\text{You ought to drive or to drive and drink.} (\Box(A \vee (A \wedge B))) \tag{5}$$

Intuitively sentences (4) and (5) are not equivalent, yet their formalizations in modal logic are so.

$A \leftrightarrow A \vee (A \wedge B)$ by propositional reasoning and

$\Box A \leftrightarrow \Box(A \vee (A \wedge B))$ by the rule of equivalence we infer .

However, hyperintensionlity is one of the distinguishing features of justification logics:
they are hyperintensional by design.

Justification logic

if $A \leftrightarrow B$ then $t : A \nrightarrow t : B$

Think of the Logic of Proofs, where the terms represent proofs in a formal system (like Peano arithmetic).
Let A and B be logically equivalent formulas.
In general, a proof of A will not also be a proof of B.
In order to obtain a proof of B we have to extend the proof of A with a proof of $A \rightarrow B$ and an application of modus ponens.
Thus in justification logic, terms do distinguish between equivalent formulas

Presention outline

Justification Version of System E

Dyadic Deontic Logic

$\Box$-operator $\xrightarrow{\text{replaced by}}$ proof term

$\bigcirc$-operator $\xrightarrow{\text{replaced by}}$ justification term

proof terms (PTm) and justification terms (JTm)

$$\lambda ::= \alpha_i \mid \xi_i \mid \Delta t \mid (\lambda + \lambda) \mid (\lambda \cdot \lambda) \mid !\lambda \mid ?\lambda$$

$$t ::= \mathsf{i} \mid x_i \mid t \cdot t \mid \nabla t \mid \mathsf{e}(t, \lambda) \mid \mathsf{n}(\lambda)$$

Formulas (Fm)

$$F ::= P_i \mid \neg F \mid (F \to F) \mid \lambda : F \mid [t](F/F) \ ,$$

where $P_i \in \mathsf{Prop}$, $\lambda \in \mathsf{PTm}$, and $t \in \mathsf{JTm}$. $[t]F$ is an abbreviation for $[t](F/\top)$.

Axiom Schemas of JE

Axioms of Classical Propositional Logic	CL
$\lambda : (F \to G) \to (\kappa : F \to \lambda \cdot \kappa : G)$	j
$(\lambda : F \vee \kappa : F) \to (\lambda + \kappa) : F$	j+
$\lambda : F \to F$	jt
$\lambda : F \to !\lambda : \lambda : F$	j4
$\neg\lambda : A \to ?\lambda : (\neg\lambda : A)$	j5
$[t](B/A) \to \Delta t : [t](B/A)$	*(Abs)*
$\lambda : B \to [\mathsf{n}(\lambda)](B/A)$	*(Nec)*
$\lambda : (A \leftrightarrow B) \to ([t](C/A) \to [\mathsf{e}(t,\lambda)](C/B)$	*(Ext)*
$[\mathsf{i}](A/A)$	*(Id)*
$[t](C/A \wedge B) \to [\nabla t](B \to C/A)$	*(Sh)*
$[t](B \to C/A) \to ([s](B/A) \to [t \cdot s](C/A))$	*(COK)*

Constant Specification

A *constant specification* CS is any subset:

$$\mathsf{CS} \subseteq \{(\alpha, A) \mid \alpha \text{ is a proof constant and } A \text{ is an axiom of } \mathsf{JE}\} .$$

A constant specification CS is called *axiomatically appropriate* if for each axiom A of JE, there is a constant α with $(\alpha, A) \in \mathsf{CS}$.

System $\mathsf{JE}_{\mathsf{CS}}$

For a constant specification CS, the system $\mathsf{JE}_{\mathsf{CS}}$ is defined by a Hilbert-style system with the axioms of JE and the following inference rules:

$$\frac{A \quad A \to B}{B}\ (\mathrm{MP}) \qquad \frac{}{\alpha : A}\ \mathrm{AN}_{\mathsf{CS}} \text{ where } (\alpha : A) \in \mathsf{CS}$$

Internalization for proof terms

Let CS be an axiomatically appropriate constant specification. For any formula A with $\mathsf{JE}_{\mathsf{CS}} \vdash A$, there exists a proof term λ such that $\mathsf{JE}_{\mathsf{CS}} \vdash \lambda : A$.

Example

The explicit version of

$$(\bigcirc(A/B) \wedge \Box B) \rightarrow \bigcirc A \qquad \text{(SFD)}$$

strong factual detachment is derivable in $\mathsf{JE}_{\mathsf{CS}}$ as follows for an axiomatically appropriate CS and a suitable term γ:

$[t](A/B) \wedge \lambda : B$	
$\gamma : ((B \wedge \top) \leftrightarrow B)$	Tautology and internalization
$[t](A/B) \rightarrow [\mathsf{e}(t,\gamma)](A/B \wedge \top)$	(Ext)
$[\mathsf{e}(t,\gamma)](A/B \wedge \top)$	(MP)
$[\nabla \mathsf{e}(t,\gamma)](B \rightarrow A/\top)$	(Sh)
$[\mathsf{n}(\lambda)](B/\top)$	(Nec)
$[\nabla \mathsf{e}(t,\gamma) \cdot \mathsf{n}(\lambda)](A/\top)$	(COK)

Semantics

Let X, Y be sets of formulas, U, V be sets of pairs of formulas, and λ be a proof term. We define the following operations:

$$\lambda : X := \{\lambda : F \mid F \in X\};$$

$$X \cdot Y := \{F \mid G \to F \in X \text{ for some } G \in Y\};$$

$$U \ominus V := \{(F, G) \mid (H \to F, G) \in U \text{ for some } (H, G) \in V\};$$

$$X \odot V := \{(F, G) \mid (G \leftrightarrow H) \in X \text{ for some } (F, H) \in V\};$$

$$\mathsf{n}(X) := \{(F, G) \mid F \in X, G \in \mathsf{Fm}\};$$

$$\nabla X := \{(F \to G, H) \mid (G, (H \wedge F)) \in X\}.$$

Basic Evaluation

A *basic evaluation* for $\mathsf{JE_{CS}}$ is a function ε that

- maps atomic propositions to 0 and 1:

$$\varepsilon(P_i) \in \{0,1\}, \ \text{ for } P_i \in \mathsf{Prop}$$

- maps proof terms to sets of formulas:

$$\varepsilon : \mathsf{PTm} \to \mathscr{P}(\mathsf{Fm})$$

such that for arbitrary $\lambda, \kappa \in \mathsf{PTm}$:

(i) $\varepsilon(\lambda) \cdot \varepsilon(\kappa) \subseteq \varepsilon(\lambda \cdot \kappa)$
(ii) $\varepsilon(\lambda) \cup \varepsilon(\kappa) \subseteq \varepsilon(\lambda + \kappa)$
(iii) $F \in \varepsilon(\alpha)$ if $(\alpha, F) \in \mathsf{CS}$
(iv) $\lambda : \varepsilon(\lambda) \subseteq \varepsilon(!\lambda)$
(v) $F \notin \varepsilon(\lambda)$ implies $\neg\lambda : F \in \varepsilon(?\lambda)$

- maps justification terms to sets of pairs of formulas:

$$\varepsilon\colon \mathsf{JTm} \to \{(A,B) \mid A,B \in \mathsf{Fm}\}$$

such that for any proof term λ and justification terms t, s:

1. $\varepsilon(t) \ominus \varepsilon(s) \subseteq \varepsilon(t \cdot s)$
2. $\varepsilon(\lambda) \odot \varepsilon(t) \subseteq \varepsilon(\mathsf{e}(t,\lambda))$
3. $\mathsf{n}(\varepsilon(\lambda)) \subseteq \varepsilon(\mathsf{n}(\lambda))$
4. $\nabla\varepsilon(t) \subseteq \varepsilon(\nabla t)$
5. $\varepsilon(\Delta t) = \{[t](A/B) \mid (A,B) \in \varepsilon(t)\}$
6. $\varepsilon(\mathsf{i}) = \{(A,A) \mid A \in \mathsf{Fm}\}$.

Truth Under a Basic Evaluation

We define truth of a formula F under a basic evaluation ε inductively as follows:

1. $\varepsilon \Vdash P$ iff $\varepsilon(P) = 1$ for $P \in \mathsf{Prop}$;
2. $\varepsilon \Vdash F \rightarrow G$ iff $\varepsilon \nVdash F$ or $\varepsilon \Vdash G$;
3. $\varepsilon \Vdash \neg F$ iff $\varepsilon \nVdash F$;
4. $\varepsilon \Vdash \lambda : F$ iff $F \in \varepsilon(\lambda)$;
5. $\varepsilon \Vdash [t](F/G)$ iff $(F, G) \in \varepsilon(t)$.

Definition

Factive Basic Evaluation A basic evaluation ε is called *factive* if for any formula $\lambda : F$ we have $\varepsilon \Vdash \lambda : F$ implies $\varepsilon \Vdash F$.

Definition

Basic Model Given an arbitrary CS, a *basic model* for $\mathsf{JE}_{\mathsf{CS}}$ is a basic evaluation that is *factive*.

Preference Models for $\mathsf{JE_{CS}}$

Quasi-model

A quasi-model for $\mathsf{JE_{CS}}$ is a triple $\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$ where:

- $W \neq \emptyset$;
- $\preceq \subseteq W \times W$
- ε is an evaluation function that asigns a basic evaluation ε_w to each world w.

Truth in Quasi-model

1. $\mathcal{M}, w \Vdash P$ iff $\varepsilon_w(P) = 1$, for $P \in \mathsf{Prop}$
2. $\mathcal{M}, w \Vdash F \rightarrow G$ iff $\mathcal{M}, w \nVdash F$ or $\mathcal{M}, w \Vdash G$
3. $\mathcal{M}, w \Vdash \neg F$ iff $\mathcal{M}, w \nVdash F$
4. $\mathcal{M}, w \Vdash \lambda : F$ iff $F \in \varepsilon_w(\lambda)$
5. $\mathcal{M}, w \Vdash [t](F/G)$ iff $(F, G) \in \varepsilon_w(t)$.

We will write $\mathcal{M} \Vdash F$ if $\mathcal{M}, w \Vdash F$ for all $w \in W$.

Locality of Truth in Quasi-models

For a quasi-model $\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$ and $w \in W$, we have for any $F \in \mathsf{Fm}$:

$$\mathcal{M}, w \Vdash F \text{ iff } \varepsilon_w \Vdash F.$$

Preference Model

A *preference model* is a quasi-model where ε_w is factive and satisfies the following condition:

$$\text{for any } t \in \mathsf{Tm} \text{ and } w \in W,$$

$$(A, B) \in \varepsilon(t) \text{ implies best}\|B\| \subseteq \|A\| \tag{JYB}$$

in other words, all best B-worlds are A-worlds. This condition is called *justification yields belief*.

Presention outline

Theorem

System $\mathsf{JE}_{\mathsf{CS}}$ is sound and complete with respect to the class of all basic models.

Soundness and Completeness w.r.t. Preference Models

System $\mathsf{JE}_{\mathsf{CS}}$ is sound and complete with respect to the class of all preference models under opt rule.

Theorem

For every preference model $\mathscr{M} = \langle W, \preceq, \varepsilon \rangle$ under opt rule, there is an equivqlent preference model $\mathscr{M}' = \langle W', \preceq', \varepsilon' \rangle$, such that $\preceq'$ is total.

Corollary

System $\mathsf{JE_{CS}}$ is sound and complete with respect to preference models with a total betterness relation.

Corollary

System $\mathsf{JE_{CS}}$ is sound and complete with respect to preference models under max rule.

Future Work

- justification logic for preference models where the betterness relation satisfies the limitedness condition. The modal axiom that corresponds to this is $\Diamond A \to (\bigcirc(B/A) \to \mathsf{P}(B/A))$;

Thanks for your attention