Conditional Obligations in Justification Logic *

Federico L.G. Faroldi, Atefeh Rohani, and Thomas Studer

Institute of Computer Science, University of Bern, Switzerland
{federico.faroldi, atefeh.rohani, thomas.studer}@unibe.ch

Abstract. This paper presents a justification counterpart for dyadic deontic logic, which is often argued to be better than Standard Deontic Logic at representing conditional and contrary-to-duty obligations, such as those exemplified by the notorious Chisholm's puzzle. We consider the alethic-deontic system (E) and present the explicit version of this system (JE) by replacing the alethic Box-modality with proof terms and the dyadic deontic Circ-modality with justification terms. The explicit representation of strong factual detachment (SFD) is given and finally soundness and completeness of the system (JE) with respect to basic models and preference models is established.

Keywords: dyadic deontic logic · justification logic · preference models.

1 Introduction

Dyadic Deontic Logic (**DDL**) is an extension of Monadic Deontic Logic (**MDL**) that employs a dyadic conditional represented by $\bigcirc(B/A)$, which is weaker than the expression $A \to \bigcirc B$ from **MDL**. The conditional $\bigcirc(B/A)$ is read as "B is obligatory, given A" so that A is the antecedent and B is the consequent [7]. In contrast to Monadic Deontic Logic, which relies on Kripke-style possible world models, Dyadic Deontic Logic works with preference-based semantics, in which the possible worlds are related according to their betterness or relative goodness. Under this semantics, $\bigcirc(B/A)$ is true when all best A-worlds are B-worlds [17]. One of the puzzles that is solved by preference models is the so-called *Chisholm's set*.

1.1 Chisholm's Set

Chisholm [6] was the initiator of the so-called "contrary-to-duty" problem, which deals with the question of what to do when primary obligations are violated. The main goal of **DDL** was to deal with these obligations, which works with setting an order on the set of worlds [15,23,24]. Here is an example of Chisholm's set. Consider the following sentences:

* F. Faroldi was supported by the Ambizione grant PZ00P1_201906, A. Rohani and T. Studer were supported by the Swiss National Science Foundation grant 200020_184625.

1. Thomas should take the math exam.
2. If he takes the math exam, he should register for it.
3. If he does not take the math exam, he should not register for it.
4. He does not take the math exam.

(1) is a primary obligation. (2) is an according-to-duty (ATD) obligation, which says what is obligatory when the primary obligation is satisfied. (3) is a contrary-to-duty obligation (CTD), which says what is obligatory when the primary obligation is violated. (4) is a descriptive premise, saying that the primary obligation is violated. Now we consider how these sentences are formalized in **MDL** and in **DDL** [25].

The paradox raises from formulating the set of formulas:

$$\Gamma = \{(1), (2), (3), (4)\}$$

in monadic deontic logic, where this set is either inconsistent or one sentence is derivable from another sentence in this set. However, Chisholm's set seems intuitively consistent and they also seem to be logically independent sentences. There are four ways to formalize this set in **MDL** as follows:

(1.1) $\bigcirc E$	(2.1) $\bigcirc E$	(3.1) $\bigcirc E$	(4.1) $\bigcirc E$
(1.2)$E \to \bigcirc R$	(2.2) $\bigcirc (E \to R)$	(3.2) $\bigcirc (E \to R)$	(4.2)$E \to \bigcirc R$
(1.3) $\bigcirc (\neg E \to \neg R)$	(2.3)$\neg E \to \bigcirc \neg R$	(3.3) $\bigcirc (\neg E \to \neg R)$	(4.3)$\neg E \to \bigcirc \neg R$
(1.4)$\neg E$	(2.4)$\neg E$	(3.4)$\neg E$	(4.4)$\neg E$

We use Γ_i to denote the set $\{(i.1), (i.2), (i.3), (i.4)\}$. Observe that

$$P \to (\neg P \to Q) \tag{1}$$

is a propositional tautology. Using (1) we find that (1.4) implies (1.2). The set Γ_2 is inconsistent: from (2.1) and (2.2) we get $\bigcirc R$ whereas from (2.3) and (2.4) we get $\bigcirc \neg R$; but in **MDL** obligations must not contradict each other. For Γ_3, note that applying necessitation to (1) and then using distributivity of $\bigcirc$ over $\to$ yields

$$\bigcirc P \to \bigcirc(\neg P \to Q).$$

Therefore, (3.1) implies (3.3). For Γ_4 we again obtain by (1) that (4.4) implies (4.2).

In **DDL**, where there is a ranking on the set of worlds according to their betterness, Chisholm's set does not yield an inconsistency because of the layers of betterness. This ranking can be defined based on the number of obligations violated in each state. Where more obligations are violated, the distance to the ideal state is bigger. The set Γ that models Chisholm's set is given by

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

The following diagram shows a model for Γ. Both R and E are true in w_1, so w_1 is the best world since no obligation of Γ is violated there. E is true in w_2 and neither E nor R is true in w_3. So w_2, w_3 are second best because one obligation is violated there. R is true in w_4 and w_4 is the worst world where two obligations of Γ are violated there.

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

1.2 Factual Detachment (FD) and Strong Factual Detachment (SFD)

In **DDL**, we do not have the validity of *Factual Detachment* (FD), which is sometimes called "deontic modus-ponens" [16]:

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

However, a restricted form of factual detachment, namely *strong factual detachment* (SFD),

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

is valid in DDL. One can interpret SFD as *if A is obligatory given B, and B is settled or proved, then A is obligatory.* An example is as follows:

1. It is obligatory to pay a fine in case someone doesn't pay taxes. ($\bigcirc(F/\neg T)$)
2. The deadline for paying taxes is over and it is proved that 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. ($\bigcirc F$)

One can consider $\Box A$ as A is proved, which guarantees that from now on we can believe that the person has not paid the taxes. Another principle, which is not valid in **DDL**, is the law of Strengtheninng of the Antecedent (SA):

$$\bigcirc(A/B) \to \bigcirc(A/B \wedge C)$$

However, the restricted form of strengthening the antecedent is valid in some systems of **DDL**, which is called "Rational Monotony", where $P(A/B)$ is read as A is permissible, given B:

$$P(A/B) \wedge \bigcirc(C/B) \to \bigcirc(C/B \wedge A)$$

Replacing a modal operator with explicit justifications first appeared in the Logic of Proofs [1], the first justification logic, which was developed by Artemov in order to introduce an explicit counterpart of the modal logic $\mathsf{S4}$ by using classical provability semantics. Various interpretations of justification logic combing justifications with traditional possible worlds models were presented after Fitting [14]. The combination of justification logic and traditional possible world models leads to various interpretations of justification logics [3,21,22]. They make it possible to apply justification logic in many different epistemic and deontic contexts [2,5,18,29,32,33].

Using justification logic for resolving deontic puzzles is already discussed by Faroldi in [9,10,12,13] where the advantages of using explicit reasons are thoroughly explained. In particular, the fact that deontic modalities are hyperintensional, i.e., they can distinguish between logically equivalent formulas, is a good motivation to use justification logic. By replacing the modal operator with a justification term, hyperintensionality is guaranteed by design in justification logic, because two logically equivalent formulas can be justified by different terms. Moreover, the problem of conflicting obligations can be handled well in justification logic [8,11].

This article aims to present an explicit version of **DDL**, where the $\Box$-operator is replaced with *proof terms* satisfying an $\mathsf{S5}$-type axioms and the $\bigcirc$-operator is replaced with suitable *justification terms*. The idea of using two types of terms is already used in [20] and also in our work on explicit non-normal modal logic [30,31]. We are going to extend the latter framework so that justification terms represent conditional obligations. One of the main motivations for developing justification counterpart of **DDL** is to find explicit reasons for countrary-to-duty and according-to-duty conditional obligations.

The problem with explicit non-normal logics is that the logic is too weak and hardly derives a formula. In the present paper, we remedy this by introducing an explicit version of dyadic deontic logic. This is much stronger than non-normal modal logic and we have appropriate formulations of according-to-duty and contrary-to-duty obligations.

In this article an axiomatization of the justification counterpart of the minimal **DDL** system JE is presented and based on this axiomatization, we provide examples that show the explicit derivation of some well-known formulas such as strong factual detachment (SFD) in our new system. For semantics, basic models are defined, and based on this, preference models are adopted for this system. Soundness and completeness of system $\mathsf{JE}_{\mathsf{CS}}$ with respect to basic models and then preference models are established.

2 Proof Systems for Alethic-Deontic Logic

We consider the proof system for alethic-deontic logic as a basis for our work. In this system, which is denoted by E, two types of modal operators are used: the alethic $\Box$-operator and dyadic deontic $\bigcirc$-operator.

2.1 Modal System

Let Prop be a countable set of atomic propositions. The set of formulas of the language of Dyadic Deontic Logic is constructed inductively as follows: [26]

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

such that $P_i \in \mathsf{Prop}$, $\Box F$ is read as "F is settled true" and $\bigcirc(F/G)$ as "F is obligatory, given G". $P(F/G)$ is a short form for $\neg \bigcirc (\neg F/G)$, $\Diamond F$ is a short form of $\neg\Box\neg F$, and $\bigcirc F$ is an abbreviation for $\bigcirc(F/\top)$ which is read as "F is unconditionally obligatory". Formulas with iterated modalities, such as $\bigcirc(p/(\bigcirc(p/q) \wedge q))$, are well-formed formulas. System E with the two operators $\Box$ and $\bigcirc$ is axiomatized as follows:

Axioms of classical propositional logic	CL
$\mathsf{S5}$-scheme axioms for $\Box$	$\mathsf{S5}$
$\bigcirc(B/A) \to \Box \bigcirc (B/A)$	(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)

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

As we see, these axioms can be categorized as follows:

- The axioms containing one operator $\Box$. These are axiom schemas of $\mathsf{S5}$, namely K, T, and 5.

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

- The axioms containing one operator $\bigcirc$. (COK) is a deontic version of the K-axiom, (Id) is the principle of identity, and (Sh), named after Shoham, is a deontic analogue of the deduction theorem.
- Finally, the axioms containing two operators $\Box$ and $\bigcirc$. (Abs), which is Lewis' principle of absoluteness, shows that the betterness relation is not world-relative. (Nec) is a deontic version of necessitation. (Ext), extensionality, makes it possible to replace necessarily equivalent sentences in the antecedent of deontic conditionals.

The following principles are derived in system E:

if $A \leftrightarrow B$ then $\bigcirc(C/A) \leftrightarrow \bigcirc(C/B)$	(LLE)
if $A \to B$ then $\bigcirc(A/C) \to \bigcirc(B/C)$	(RW)
$\bigcirc(B/A) \wedge \bigcirc(C/A) \to \bigcirc(B \wedge C/A)$	(AND)
$\bigcirc(C/A) \wedge \bigcirc(C/B) \to \bigcirc(C/A \vee B)$	(OR)
$\bigcirc(C/A) \wedge \bigcirc(D/B) \to \bigcirc(C \vee D/A \vee B)$	(OR')

2.2 Preference Models

Now we review the preference model semantics for system E as follows:

Definition 1 (Preference model). *A preference model is a tuple*

$$\mathcal{M} = (W, \preceq, V),$$

where:

- *W is a non-empty set of worlds;*
- *$\preceq$ is a binary relation on W, called betterness relation, which orders the set of worlds according to their relative goodness. So for $w, v \in W$ we read $w \preceq v$ as "state v is at least as good as state w;*
- *V is a valuation function assigning a set $V(p) \subseteq W$ to each atomic formula p.*

Definition 2 (Truth under preference model). *Given a preference model $\mathcal{M} = (W, \preceq, V)$, for $w, v \in W$ and $A, B \in \mathsf{Fm}$, the truth for formulas under $\mathcal{M}$ is defined as follows:*

- *for propositional formulas is in standard way;*
- *$\mathcal{M}, w \Vdash \Box A$ iff, for all $v \in W$, $\mathcal{M}, v \Vdash A$;*
- *$\mathcal{M}, w \Vdash \bigcirc(A/B)$ iff $\mathrm{best}\|B\| \subseteq \|A\|$;*

where $\|A\|$ is truth set of A, i.e., the set of all worlds in which A is true. $\mathrm{best}\|B\|$ *is the subset of $\|B\|$ which is best according to $\preceq$.*

2.3 Justification Version of System E

Now we present the explicit version of E denoted by JE. We first define the set of terms and formulas as follows.

Definition 3. *The set of* proof terms*, shown by* PTm*, and* justification terms*, shown by* JTm*, are defined as follows:*

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

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

where α_i are proof constants, ξ_i are proof variables, i *is a justification constant and x_i are justification variables.*

Formulas are inductively defined as follows:

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

where $P_i \in \mathsf{Prop}$, $\lambda \in \mathsf{PTm}$, and $t \in \mathsf{JTm}$. $[t]F$ is an abbreviation for $[t](F/\top)$. We use Fm for the set of formulas.

Definition 4 (Axiom Schemas of JE).

Axioms of Classical Propositional Logic	CL
$\lambda : (F \to G) \to (\kappa : F \to \lambda \cdot \kappa : G)$	j
$(\lambda : F \vee \kappa : F) \to (\lambda + \kappa) : F$	$\mathsf{j+}$
$\lambda : F \to F$	jt
$\lambda : F \to !\lambda : \lambda : F$	$\mathsf{j4}$
$\neg\lambda : A \to ?\lambda : (\neg\lambda : A)$	$\mathsf{j5}$
$[t](B/A) \to \Delta t : [t](B/A)$	(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)

Definition 5 (Constant Specification). *A* constant specification CS *is any subset:*

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

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

Definition 6 (System $\mathsf{JE_{CS}}$). *For a constant specification* CS*, the system* $\mathsf{JE_{CS}}$ *is defined by a Hilbert-style system with the axioms of* JE *and the following inference rules:*

$$\frac{A \quad A \to B}{B}\ (MP) \qquad \frac{}{\alpha : A}\ AN_{\mathsf{CS}} \textit{ where } (\alpha : A) \in \mathsf{CS}$$

As usual in justification logic [1,4,19], $\mathsf{JE_{CS}}$ internalizes its own notion of proof.

Lemma 1 (Internalization). *Let* CS *be an axiomatically appropriate constant specification. For any formula A with $\mathsf{JE_{CS}} \vdash A$, there exists a proof term λ such that $\mathsf{JE_{CS}} \vdash \lambda : A$.*

To have a better understanding of the axiomatic system of JE, we provide Hilbert-style proofs of some typical formulas in the following examples. It is notable how terms are constructed as a justification for obligations.

Example 1. The explicit version of

$$\text{if} \quad A \to B \quad \text{then} \quad \bigcirc(A/C) \to \bigcirc(B/C) \tag{RW}$$

is derivable in $\mathsf{JE_{CS}}$ as follows for an axiomatically appropriate CS and a suitable term λ:

$A \to B$	
$\lambda : (A \to B)$	(Internalization)
$[\mathsf{n}(\lambda)](A \to B/C)$	(Nec)
$[s](A/C) \to [\mathsf{n}(\lambda) \cdot s](B/C)$	(COK)

Example 2. The explicit version of

$$\bigcirc(B/A) \land \bigcirc(C/A) \to \bigcirc(B \land C/A) \tag{AND}$$

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

$[t](B/A) \land [s](C/A)$	
$B \to (C \to B \land C)$	(Tautology)
$[t](B/A) \to [\mathsf{n}(\lambda) \cdot t](C \to B \land C/A)$	(RW)
$[\mathsf{n}(\lambda) \cdot t](C \to B \land C/A)$	(MP)
$[s](C/A) \to [\mathsf{n}(\lambda) \cdot t \cdot s](B \land C/A)$	(COK)
$[\mathsf{n}(\lambda) \cdot t \cdot s](B \land C/A)$	(MP)

Example 3. The explicit version of

$$(\bigcirc(A/B) \land \Box B) \to \bigcirc A \tag{SFD}$$

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

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

3 Semantics

We first consider the following operations on the sets of formulas and sets of pairs of formulas in order to define basic evaluations.

Definition 7. *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:*

$\lambda : X := \{\lambda : F \mid F \in X\}$;
$X \cdot Y := \{F \mid G \to F \in X \text{ for some } G \in Y\}$;
$U \ominus V := \{(F, G) \mid (H \to F, G) \in U \text{ for some } (H, G) \in V\}$;
$X \odot V := \{(F, G) \mid (G \leftrightarrow H) \in X \text{ for some } (F, H) \in V\}$;
$\mathsf{n}(X) := \{(F, G) \mid F \in X, G \in \mathsf{Fm}\}$;
$\nabla U := \{(F \to G, H) \mid (G, (H \land F)) \in U\}$.

Definition 8 (Basic Evaluation). *A* basic evaluation *for $\mathsf{JE}_{\mathsf{CS}}$ is a function ε that*

- *maps atomic propositions to 0 and 1:*

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

– *maps proof terms to sets of formulas:*

$$\varepsilon(\lambda) \in \mathcal{P}(\mathsf{Fm}) \textit{ for } \lambda \in \mathsf{PTm}$$

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:*

$$\varepsilon(t) := \{(A, B) \mid A, B \in \mathsf{Fm}\}, \textit{ for } t \in \mathsf{JTm}$$

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. $\varepsilon(\mathsf{i}) = \{(A, A) \mid A \in \mathsf{Fm}\}$.

Definition 9 (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. $\varepsilon \Vdash F \to G$ *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 10 (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 11 (Basic Model). *Given an arbitrary* CS*, a* basic model *for* $\mathsf{JE}_{\mathsf{CS}}$ *is a basic evaluation that is factive.*

The following theorem gives us the expected soundness and completeness with respect to basic models which is proved in Appendix A.

Theorem 1 (Soundness and Completeness w.r.t. Basic Models). *Let* CS *be an arbitrary constant specification. System* $\mathsf{JE}_{\mathsf{CS}}$ *is sound and complete with respect to the class of all basic models. For any formula* F*,*

$$\mathsf{JE}_{\mathsf{CS}} \vdash F \quad \textit{iff} \quad \varepsilon \Vdash F \textit{ for all basic models } \varepsilon \textit{ for } \mathsf{JE}_{\mathsf{CS}} \ .$$

4 Preference Models

In this section, we introduce preference models for $\mathsf{JE_{CS}}$, which feature a set of possible worlds together with a *betterness* or *comparative goodness* relation on them.

Definition 12 (Quasi-model). *A quasi-model for* $\mathsf{JE_{CS}}$ *is a triple*

$$\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$$

where:

- W *is a non-empty set of worlds;*
- $\preceq \subseteq W \times W$ *is a binary relation on the set of worlds where* $w_1 \preceq w_2$ *is read as* world w_2 is at least as good as world w_1.
- ε *is an evaluation function that asigns a basic evaluation* ε_w *to each world* w.

Definition 13 (Truth in Quasi-model). *Let* $\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$ *be a quasi-model.* Truth of a formula at a world w *in a quasi-model is defined inductively as follows:*

1. $\mathcal{M}, w \Vdash P$ *iff* $\varepsilon_w(P) = 1$, *for* $P \in \mathsf{Prop}$
2. $\mathcal{M}, w \Vdash F \to G$ *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* $w \in W$.

Remark 1. As usual for quasi-models for justification logic [3,19,21], truth is local, i.e., for a quasi-model $\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$ and $w \in W$, we have for any $F \in \mathsf{Fm}$:

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

Remark 2. Let $\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$ be a quasi-model. The *truth set* of $F \in \mathsf{Fm}$ is the set of all worlds in which F is true (denoted by $\|F\|^{\mathcal{M}}$),

$$\|F\|^{\mathcal{M}} := \{w \in W \mid \mathcal{M}, w \Vdash F\}.$$

Moreover, the best worlds in which F is true, according to $\preceq$, are called *best F-worlds* and are denoted by $\mathrm{best}_{\preceq}\|F\|^{\mathcal{M}}$. For simplicity we often write $\|F\|$ for $\|F\|^{\mathcal{M}}$ and $\mathrm{best}\|F\|$ for $\mathrm{best}_{\preceq}\|F\|^{\mathcal{M}}$ when the model is clear from the context.

Remark 3 (Two Notions of "Best"). There are two ways to formalize the notion of "best world" respecting optimality and maximality [27]:

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

$$opt_{\preceq}(\|A\|) = \{w \in \|A\|^{\mathcal{M}} \mid \forall v(\mathcal{M}, v \Vdash A \to v \preceq w)\}$$

– best$\|A\|$ under "max rule":

$$max_{\preceq}(\|A\|) = \{w \in \|A\|^{\mathcal{M}} \mid \forall v((\mathcal{M}, v \Vdash A \wedge w \preceq v) \to v \preceq w)\}$$

Definition 14 (Preference Model). *A* preference model *is a quasi-model where ε_w is factive and satisfies the following condition:*

$$\textit{for any} \quad t \in \mathsf{JTm} \textit{ and } w \in W,$$

$$(A, B) \in \varepsilon_w(t) \textit{ implies } \mathrm{best}\|B\| \subseteq \|A\| \tag{JYB}$$

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

Definition 15 (Properties of $\preceq$). *We can require additional properties for the relation $\preceq$ such as:*

- *reflexivity: for all $w \in W, w \preceq w$*
- *totalness: for all $w, v \in W, w \preceq v$ or $v \preceq w$*
- *limitedness: if $\|A\| \neq \emptyset$ then* best$\|A\| \neq \emptyset$.

Limitedness avoids the case of not having a best state, i.e., of having infinitely many strictly better states. Morover, totalness yields reflexivity.

Lemma 2. *$max_{\preceq}(\|A\|) = opt_{\preceq}(\|A\|)$ if $\preceq$ is total.*

Proof. If $\preceq$ is total, then clearly from the definition $opt_{\preceq}(\|A\|) \subseteq max_{\preceq}(\|A\|)$. For the converse inclusion, suppose $w \in max_{\preceq}(\|A\|)$. By totalness, for any $v \in W$ with $\mathcal{M}, v \Vdash A$, either $v \preceq w$ or $w \preceq v$. In first case $w \in opt_{\preceq}(\|A\|)$ and in latter case, by definition of $max_{\preceq}, v \preceq w$ and $w \in opt_{\preceq}(\|A\|)$.

4.1 Soundness and Completeness w.r.t. Preference Models

Theorem 2. *System $\mathsf{JE}_{\mathsf{CS}}$ is sound and complete with respect to the class of all preference models under opt rule.*

Proof. To prove soundness, suppose $\mathcal{M} = \langle W, \preceq, \varepsilon\rangle$ is a preference model and $\mathsf{JE} \vdash A$. We show that A is true in every world $w \in W$. By soundness of JE with respect to basic models, we get $\varepsilon_w \Vdash A$ for all ε_w and by locality of truth in quasi-models, we conclude $\mathcal{M}, w \Vdash A$.

To prove completeness, suppose that $\mathsf{JE} \nvdash A$. By completeness of JE with respect to basic models, there is a basic model ε such that $\varepsilon \nVdash A$. Now construct a preference model $\mathcal{M} := \langle \{w_1\}, \preceq, \varepsilon'\rangle$ with $\varepsilon'_{w_1} := \varepsilon$ and $\preceq := \emptyset$. Then by locality of truth, we have $\mathcal{M}, w_1 \nVdash A$. It is easy to see that $\mathcal{M}$ is a preference model, i.e., to show (JYB). For any $t \in \mathsf{Tm}$ if $(B, C) \in \varepsilon(t)$, we have best$\|C\| \subseteq \|B\|$ since best$\|C\| = \emptyset$.

Remark 4. The above proof does not give us completeness under the max rule. The problem is that for the max rule, we cannot define the relation $\preceq$ such that best$\|C\| = \emptyset$.

However, by proving the following theorem we get desired results analogous to result in [28].

Theorem 3. *For every preference model $\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$ under opt rule, there is an equivalent preference model $\mathcal{M}' = \langle W', \preceq', \varepsilon' \rangle$, such that $\preceq'$ is total (and hence reflexive).*

Proof. Let $\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$. We define $\mathcal{M}' = \langle W', \preceq', \varepsilon' \rangle$ as follows:

- $W' = \{\langle w, n \rangle \mid w \in W, n \in \omega\}$;
- $\langle w, n \rangle \preceq' \langle v, m \rangle$ iff $w \preceq v$ or $n \leq m$;
- $\varepsilon'(p) = \{\langle w, n \rangle \mid w \in \varepsilon(p)\}$, for $p \in \mathsf{Prop}$;
- $\varepsilon'_{\langle w,n \rangle}(\lambda) = \varepsilon_w(\lambda)$;
- $\varepsilon'_{\langle w,n \rangle}(t) = \varepsilon_w(t)$;

where ω is the set of natural numbers. One can easily see that $\preceq'$ is total, since for any $\langle w, n \rangle$ and $\langle v, m \rangle$ in W', we have either $\langle w, n \rangle \preceq' \langle v, m \rangle$ or $\langle v, m \rangle \preceq' \langle w, n \rangle$, by totality of $\leq$ on the set of natural numbers. By locality of truth, for any formula $F \in \mathsf{Fm}$, we have $\mathcal{M}, w \Vdash F$ iff $\mathcal{M}', \langle w, n \rangle \Vdash F$ for all $n \in \omega$.

In order to show (JYB) in $\mathcal{M}'$, suppose $\mathcal{M}', \langle w, n \rangle \Vdash [t](A/B)$. By definition of $\mathcal{M}'$ we get $(A, B) \in \varepsilon'_{\langle w,n \rangle}(t)$ and so $(A, B) \in \varepsilon_w(t)$.

By applying (JYB) in $\mathcal{M}$, we get $\text{best}\|B\|^{\mathcal{M}} \subseteq \|A\|^{\mathcal{M}}$. We need to show that $\text{best}\|B\|^{\mathcal{M}'} \subseteq \|A\|^{\mathcal{M}'}$. Suppose $\langle v, k \rangle \in \text{best}\|B\|^{\mathcal{M}'}$, which means $\mathcal{M}', \langle v, k \rangle \Vdash B$. Then by definition of $\mathcal{M}'$ we have $\mathcal{M}, v \Vdash B$. We will show that $v \in \text{best}\|B\|^{\mathcal{M}}$. Suppose towards contradiction that $v \notin \text{best}\|B\|^{\mathcal{M}}$. Based on this, there is a world $u \in W$ such that $u \not\preceq v$ and $\mathcal{M}, u \Vdash B$. From this we get $\langle u, k \rangle \in W'$ and $\langle u, k+1 \rangle \in W'$ as well. By definition of $\mathcal{M}'$ we have $\mathcal{M}', \langle u, k+1 \rangle \Vdash B$, where $\langle v, k \rangle \preceq' \langle u, k+1 \rangle$. This is a contradiction with the assumption that $\langle v, k \rangle \in \text{best}\|B\|^{\mathcal{M}'}$. As a result $v \in \text{best}\|B\|^{\mathcal{M}}$ and by (JYB) in $\mathcal{M}$ we get $v \in \|A\|^{\mathcal{M}}$, which means $\mathcal{M}, v \Vdash A$. As a result $\mathcal{M}', \langle v, k \rangle \Vdash A$, which means $\langle v, k \rangle \in \|A\|^{\mathcal{M}'}$.

We conclude that the following strengthening of Theorem 2 holds.

Corollary 1. *System $\mathsf{JE_{CS}}$ is sound and complete with respect to preference models with a total betterness relation.*

By Lemma 2 this implies completeness of $\mathsf{JE_{CS}}$ with respect to preference models under max rule.

Corollary 2. *System $\mathsf{JE_{CS}}$ is sound and complete with respect to preference models under max rule.*

5 Conclusion and Future Work

Having explicit counterparts of modalities is valuable not only in epistemic but also in deontic contexts, where justification terms can be interpreted as reasons for obligations. Explicit non-normal modal logic [30] avoids the usual deontic paradoxes at the cost of being very (too) weak with respect to deductive power [31]. In the present paper, we introduced an explicit version $\mathsf{JE_{CS}}$ of the alethic-deontic system E, which features dyadic modalities to capture deontic conditionals. Semantics for E is given in terms of preference models, where the set of worlds is ordered according to a betterness relation. The language of $\mathsf{JE_{CS}}$ includes proof terms for the alethic modality and justification terms for the deontic modality.

We established soundness and completeness of $\mathsf{JE_{CS}}$ with respect to basic models and preference models. In preference models, the property "justification yields belief" (JYB) holds, which means justified formulas act like obligatory formulas.

The converse direction, however, only holds in fully explanatory models. A preference model is *fully explanatory* if the converse of (JYB) holds, that is for any world w and any formulas A, B:

$$\mathrm{best}\|B\| \subseteq \|A\| \text{ implies } (A, B) \in \varepsilon_w(t) \text{ for some } t \in \mathsf{JTm}.$$

To prove the completeness for $\mathsf{JE_{CS}}$ with respect to fully explanatory preference models, one would have to follow the strategy of the completeness proof for the modal system E [28]. That is define so-called selection function models for $\mathsf{JE_{CS}}$, establish completeness with respect to the selection function models, and show that for each selection function model, there is an equivalent preference model.

Another line of future work is to study justification logic for preference models where the betterness relation satisfies the limitedness condition. The modal axiom that corresponds to this is $\Diamond A \to (\bigcirc(B/A) \to \mathsf{P}(B/A))$, where $\Diamond A$ and $\mathsf{P}(B/A))$ stand for $\neg\Box\neg A$ and $\neg \bigcirc (\neg B/A)$, respectively. The problem of finding a justification logic version for this axiom is that terms in justification logic usually stand for $\Box$-type modalities. A notable exception is the work on justified constructive modal logic [20].

A Soundness and completeness with respect to basic models

Theorem 4. *System* $\mathsf{JE}_{\mathsf{CS}}$ *is sound with respect to the class of all basic models.*

Proof. The proof is by induction on the length of derivations in $\mathsf{JE}_{\mathsf{CS}}$. For an arbitrary basic model ε, soundness of the propositional axioms is trivial and soundness of $\mathsf{S5}$ axioms $\mathsf{j}, \mathsf{jt}, \mathsf{j4}, \mathsf{j5}, \mathsf{j}+$ immediately follows from the definition of basic evaluation and factivity. We just check the cases for the axioms containing justification terms. Suppose $\mathsf{JE}_{\mathsf{CS}} \vdash F$ and F is an instance of:

- (COK): Suppose $\varepsilon \Vdash [t](B \to C/A)$ and $\varepsilon \Vdash [s](B/A)$. Thus we have

$$(B \to C, A) \in \varepsilon(t) \quad \text{and} \quad (B, A) \in \varepsilon(s).$$

  By the definition of basic model, we have $\varepsilon(t) \ominus \varepsilon(s) \subseteq \varepsilon(t \cdot s)$ and as a result $(C, A) \in \varepsilon(t \cdot s)$, which means $\varepsilon \Vdash [t \cdot s](C/A)$.
- (Nec): Suppose $\varepsilon \Vdash (\lambda : A)$. Thus $A \in \varepsilon(\lambda)$. By the definition of $\mathsf{n}(\varepsilon(\lambda))$ we have $(A, B) \in \mathsf{n}(\varepsilon(\lambda))$ for any $B \in \mathsf{Fm}$ and by the definition of basic evaluation $\mathsf{n}(\varepsilon(\lambda)) \subseteq \varepsilon(\mathsf{n}(\lambda))$, so $(A, B) \in \varepsilon(\mathsf{n}(\lambda))$, which means $\varepsilon \Vdash [\mathsf{n}(\lambda)](A/B)$.
- (Ext): Suppose $\varepsilon \Vdash \lambda : (A \leftrightarrow B)$, so $(A \leftrightarrow B) \in \varepsilon(\lambda)$. Since $\varepsilon(\lambda) \odot \varepsilon(t) \subseteq \varepsilon(\mathsf{e}(t, \lambda))$, we have $(C, B) \in \varepsilon(\mathsf{e}(t, \lambda))$ if $(C, A) \in \varepsilon(t)$. Hence $\varepsilon \Vdash ([t](C/A) \to [\mathsf{e}(t, \lambda)](C/B))$.
- (Sh): Suppose $\varepsilon \Vdash [t](C/A \wedge B)$, then $(C, (A \wedge B)) \in \varepsilon(t)$. By definition of $\nabla(\varepsilon(t))$ we have $(B \to C, A) \in \nabla(\varepsilon(t))$ and by definition of basic models, $\nabla\varepsilon(t) \subseteq \varepsilon(\nabla t)$. As a result, $((B \to C), A) \in \varepsilon(\nabla t)$ which means $\varepsilon \Vdash [\nabla t](B \to C/A)$.

For the axioms (Abs) and (Id) soundness is immediate from the definition of basic evaluation. □

Theorem 5. *System* $\mathsf{JE}_{\mathsf{CS}}$ *is complete with respect to the class of all basic models.*

Proof. Given a maximal consistent Γ, we define the canonical model ε^c induced by Γ as follows:

- $\varepsilon^c_\Gamma(P) := 1$, if $P \in \Gamma$ and $\varepsilon^c := 0$, if $P \notin \Gamma$;
- $\varepsilon^c_\Gamma(\lambda) := \{F \mid \lambda : F \in \Gamma\}$;
- $\varepsilon^c_\Gamma(t) := \{(F, G) \mid [t](F/G) \in \Gamma\}$.

We first show that ε^c is a basic evaluation. Conditions (i)–(v) follow immediately from the maximal consistency of Γ and axioms of $\mathsf{j} - \mathsf{j5}$. Conditions (1)–(6) are obtained from the axioms (Abs), (COK), (Nec), (Id), (Ext), and (Sh). Let us only show (1) and (3).

To check condition (1), suppose $(C, B) \in \varepsilon^c(t) \ominus \varepsilon^c(s)$. Then there is an $A \in \mathsf{Fm}$ such that $(A \to C, B) \in \varepsilon^c(t)$ and $(A, B) \in \varepsilon^c(s)$. By the definition of canonical model $[t](A \to C/B) \in \Gamma$ and $[s](A/B) \in \Gamma$, by maximal consistency of Γ and axiom (COK) we have $[t \cdot s](C/B) \in \Gamma$, which gives $(C/B) \in \varepsilon^c(t \cdot s)$.

To check condition (3), suppose $(A, B) \in \mathsf{n}(\varepsilon^c(\lambda))$. Then $A \in \varepsilon^c(\lambda)$, which means $\lambda : A \in \Gamma$. By maximal consistency of Γ and axiom (Nec) we get $[\mathsf{n}(\lambda)](A/B) \in \Gamma$. By the definition of canonical model we conclude $(A, B) \in \varepsilon^c(\mathsf{n}(\lambda))$. Thus ε^c is a basic evaluation.

The truth lemma states:

$$F \in \Gamma \quad \text{iff} \quad \varepsilon^c \Vdash F \ ,$$

which is established as usual by induction on the structure of F. In case $F = [t](A/B)$, we have $[t](A/B) \in \Gamma$ iff $(A, B) \in \varepsilon^c(t)$ iff $\varepsilon^c \Vdash [t](A/B)$.

Due to axiom jt, ε^c is factive by the following reasoning: if $\varepsilon^c \Vdash \lambda : F$, we get by the truth lemma that $\lambda : F \in \Gamma$. By the maximal consistency of Γ we have $F \in \Gamma$ which means $\varepsilon^c \Vdash F$ by the truth lemma. $\square$

References

1. Artemov, S.: Explicit provability and constructive semantics. BSL **7**(1), 1–36 (Mar 2001)
2. Artemov, S.: Justified common knowledge. TCS **357**(1–3), 4–22 (2006). https://doi.org/10.1016/j.tcs.2006.03.009
3. Artemov, S.: The ontology of justifications in the logical setting. Studia Logica **100**(1–2), 17–30 (Apr 2012). https://doi.org/10.1007/s11225-012-9387-x
4. Artemov, S., Fitting, M.: Justification Logic: Reasoning with Reasons. Cambridge Tracts in Mathematics, Cambridge University Press (2019). https://doi.org/10.1017/9781108348034
5. Bucheli, S., Kuznets, R., Studer, T.: Justifications for common knowledge. Applied Non-Classical Logics **21**(1), 35–60 (2011). https://doi.org/10.3166/JANCL.21.35-60
6. Chisholm, R.M.: Contrary-To-Duty Imperatives and Deontic Logic. Analysis **24**(2), 33–36 (12 1963). https://doi.org/10.1093/analys/24.2.33
7. Danielsson, S.: Preference and obligation. : Studies in the logic of ethics. Filosofiska föreningen (1968)
8. Faroldi, F., Ghari, M., Lehmann, E., Studer, T.: Impossible and conflicting obligations in justification logic. In: Marra, A., Liu, F., Portner, P., Van De Putte, F. (eds.) Proceedings of DEON 2020 (2020)
9. Faroldi, F.L.G.: Deontic modals and hyperintensionality. Logic Journal of the IGPL **27**, 387–410 (2019). https://doi.org/10.1093/jigpal/jzz011
10. Faroldi, F.L.G.: Hyperintensionality and Normativity. Springer (2019)
11. Faroldi, F.L.G., Ghari, M., Lehmann, E., Studer, T.: Consistency and permission in deontic justification logic. Journal of Logic and Computation (2022). https://doi.org/10.1093/logcom/exac045
12. Faroldi, F.L.G., Protopopescu, T.: All-things-considered ought via reasons in justification logic (2018), preprint
13. Faroldi, F.L.G., Protopopescu, T.: A hyperintensional logical framework for deontic reasons. Logic Journal of the IGPL **27**, 411–433 (2019). https://doi.org/10.1093/jigpal/jzz013
14. Fitting, M.: The logic of proofs, semantically. APAL **132**(1), 1–25 (2005). https://doi.org/10.1016/j.apal.2004.04.009

15. van Fraassen, B.C.: The logic of conditional obligation. Journal of Philosophical Logic **1**(3/4), 417–438 (1972)
16. Gabbay, D., Horty, J., Parent, X., van der Meyden, R., van der Torre, L.: Handbook of Deontic Logic and Normative System, Volume 2. College Publications (2021)
17. Hansson, B.: An analysis of some deontic logics. Noûs **3**(4), 373–398 (1969)
18. Kokkinis, I., Maksimović, P., Ognjanović, Z., Studer, T.: First steps towards probabilistic justification logic. Logic Journal of IGPL **23**(4), 662–687 (2015). https://doi.org/10.1093/jigpal/jzv025
19. Kuznets, R., Studer, T.: Logics of Proofs and Justifications. Studies in logic, College Publications (2019)
20. Kuznets, R., Marin, S., Straßburger, L.: Justification logic for constructive modal logic. Journal of Applied Logic **8**, 2313–2332 (2021)
21. Kuznets, R., Studer, T.: Justifications, ontology, and conservativity. In: Bolander, T., Braüner, T., Ghilardi, S., Moss, L. (eds.) Advances in Modal Logic, Volume 9, pp. 437–458. College Publications (2012)
22. Lehmann, E., Studer, T.: Subset models for justification logic. In: Iemhoff, R., Moortgat, M., de Queiroz, R. (eds.) Logic, Language, Information, and Computation - WoLLIC 2019. pp. 433–449. Springer (2019)
23. Lewis, D.K.: Counterfactuals. Cambridge, MA, USA: Blackwell (1973)
24. Lewis, D.K.: Semantic analyses for dyadic deontic logic. In: Stenlund, S. (ed.) Logical Theory and Semantic Analysis: Essays Dedicated to Stig Kanger on His Fiftieth Birthday, pp. 1–14. Reidel (1974)
25. Parent, X., van der Torre, L.: Introduction to Deontic Logic and Normative Systems. Texts in logic and reasoning, College Publications (2018)
26. Parent, X.: A complete axiom set for Hansson's deontic logic DSDL2. Logic Journal of the IGPL **18**(3), 422–429 (2010). https://doi.org/10.1093/jigpal/jzp050
27. Parent, X.: Maximality vs. optimality in dyadic deontic logic. J. Philos. Log. **43**(6), 1101–1128 (2014). https://doi.org/10.1007/s10992-013-9308-0, https://doi.org/10.1007/s10992-013-9308-0
28. Parent, X.: Completeness of Åqvist's systems E and F. Rev. Symb. Log. **8**(1), 164–177 (2015). https://doi.org/10.1017/S1755020314000367
29. Renne, B.: Dynamic Epistemic Logic with Justification. Ph.D. thesis, City University of New York (May 2008)
30. Rohani, A., Studer, T.: Explicit non-normal modal logic. In: Silva, A., Wassermann, R., de Queiroz, R. (eds.) Logic, Language, Information, and Computation. pp. 64–81. Springer International Publishing, Cham (2021)
31. Rohani, A., Studer, T.: Explicit non-normal modal logic. Journal of Logic and Computation (in print)
32. Studer, T.: Decidability for some justification logics with negative introspection. JSL **78**(2), 388–402 (Jun 2013). https://doi.org/10.2178/jsl.7802030
33. Xu, C., Wang, Y., Studer, T.: A logic of knowing why. Synthese **198**, 1259–1285 (2021)