Deontic Obligations in Justification Logic

Inaugural dissertation
of the Faculty of Science
University of Bern

presented by

Atefeh Rohani

from Iran

Supervisor of the doctoral thesis:
Prof. Dr. Thomas Studer

Institute of Computer Science

Deontic Obligations in Justification Logic

Inaugural dissertation
of the Faculty of Science
University of Bern

presented by

Atefeh Rohani

from Iran

Supervisor of the doctoral thesis:
Prof. Dr. Thomas Studer

Accepted by the Faculty of Science.

Bern, 23rd of August 2023 | The Dean
Prof. Dr. Marco Herwegh

Dedication

To women of Iran,
those who are standing in a combat for their rights,
those who are creators of civilizations,
from whom I learned the alphabets of life.

Acknowledgement

Thanks to my family who supported me unconditionally in the first chapter of my life.
Thanks to Thomas Studer who opened a new chapter in my life!
Thanks to the beautiful Alps that embraced me and my sorrows kindly!
During my journey I faced elves, without whom my way could not be continued at certain times:
Bettina Choffat, Eveline Lehmann, Ulrike Wild and Maryam Yousefi.
Thanks to the Logic and Theory Group members for all nice chats and coffee breaks.
I would also appreciate the gentle guidance from Pro. Federico Faroldi and Prof. George Metcalfe.

Abstract

This thesis consists of two main chapters which connect two areas of deontic logic and justification logic. Deontic logic is logic of normative concepts, where the deontic reading of modal operator for $\Box A$ is as "A is obligatory". On the other hand, justification logics replace an explicit term with modal operator such that instead of $\Box A$, justification logics features formulas of the form $t : A$, read as "A is justified by reason t", or "A is obligatory for the reason t", in deontic context.

In the second chapter, we focus on the category of non-normal modal logics and provide an explicit version of two logics in this category, the weakest non-normal modal logic E and logic EM, which is an extension of logic E by adding rule of monotonicity.

The main motivation for this attempt is raised from the following arguments initiated in the area of deontic logic. First, we consider paradoxes raising from deontic interpretation of modal operator in normal modal logic, which leads us towards non-normal modal logics.

The second issue is hyperintensionality. Considering the fact that deontic modals are hyperintensional, i.e., they can distinguish between logically equivalent formulas, one can see that traditional modal logic cannot provide an appropriate formalization of deontic situations. Since justification terms are hyperintensional in nature, we introduce novel justification logics as hyperintensional analogues to non-normal modal logics. We establish soundness and completeness with respect to various models and we study the problem of realization.

In the third chapter, we focus on conditional obligations in deontic logic. We review standard deontic logic and the well known Chisholm's puzzle confronting this system. For this reason, we turn into dyadic deontic logic which is often argued to be better than standard deontic logic at representing conditional and contrary-to-duty obligations.

We consider the alethic-deontic system (AD) and present a justification counterpart for this system (JAD) by replacing the alethic $\Box$-modality with proof terms and the dyadic deontic $\bigcirc$-modality with justification terms. The explicit representation of strong factual detachment (SFD) is given and finally soundness and completeness of the system (JAD) with respect to basic models and preference models is established.

Contents

Chapter 1

Introduction

1.1 Justification Logic and Deontic Logic

1.1.1 Justification Logic

Justification logic [4, 21] is a variant of modal logic that replaces the implicit $\Box$-operator with explicit justifications. Instead of formulas $\Box A$, meaning, e.g., *A is known* or *A is obligatory*, the language of justification logic features formulas of the form $t : A$ that stand for *t justifies the agent's knowledge of A* or *A is obligatory for reason t*, where t is a so-called justification term.

The first justification logic, the Logic of Proofs [1], has been developed by Artemov in order to provide a classical provability semantics for the modal logic S4 (and thus also for intuitionistic logic) [1, 24].

Starting with the work of Fitting [14], several interpretations of justification logic have been presented that combine justifications with traditional possible world models [3, 23, 26]. This opened the door for numerous applications of justification logic, e.g., in epistemic and deontic contexts [2, 6, 20, 37, 39, 42].

1.1.2 About Deontic Logic

The word *deon* is rooted in the Greek expression "déon" ($\delta\epsilon o\nu$) which means "what is binding" or "what is proper". The Austrian philosopher Mally developed a system "fundamental principles the logic of ought" in 1920's and called his theory "Deontik" [29]. Jeremy Bentham used the word "deontology" for the "science of morality" [5]. *Deontic logic* is an area of logic that involves normative (deontic) concepts including norms and norm systems, and normative reasoning. Normative concepts include the following concepts:

- obligation (duty, requirement, ought);
- permission (permissibility, may);
- prohibition (may not, forbidden).

Deontic logic is also concerned with the relations between normative concepts, axiological concepts (concept of value such as "good" or "better than"), and agent-evaluative concepts (like "blameworthy" and "praiseworthy"). Thus the formal language of deontic logic contain, in addition to propositional connectives, logical operators representing deontic concepts.

1.2 Connections between Justification Logic and Deontic Logic

1.2.1 Justification Counterpart for Non-normal Modal Logics

In this thesis, the main focus is on the deontic context of justification logic and using this logic to overcome deontic puzzles that face modal logic. In the second chapter, we discuss deontic puzzles formulated in modal logic and we show how replacing the modal operator with a justification term will overcome the puzzles.

We first observe normal modal logic and a puzzle which is raised from rules and axioms regarding normality. So we move to non-normal modal logics and we see that there are still puzzles steam from the fact that modal operator in modal logic is not hyperintensional and since justification terms are hyperintensional by nature, we come to the idea of introducing a justification counterpart for the weakest non-normal modal logic.

Under concept of hyperintensionality, here is the key idea to overcome puzzles. Justification logics are parametrized by a constant specification, which is a set

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

A constant specification CS is called *axiomatically appropriate* if for each axiom A there is a constant c such that $(c, A) \in \mathsf{CS}$. Instead of the rule of necessitation, justification logics include a rule called *axiom necessitation* saying that one is allowed to infer $c : A$ if $(c, A) \in \mathsf{CS}$. Hence, In epistemic settings, we can calibrate the reasoning power of the agents by adapting the constant specification. [1]

Faroldi and Protopopescu [11, 13] suggest using this mechanism also in deontic settings in order to avoid the usual paradoxes.

There is a problem with restricting the constant specification. Namely, the resulting constant specification will not be *axiomatically appropriate*, i.e. there will be axioms that are not justified by any term. This implies, however, that the Internalization property (saying that a justification logic internalizes its own notion of proof) does not hold, which is a problem for several reasons.

First, Internalization is needed to obtain completeness with respect to fully explanatory models. That is models where each formula that is obligatory (or believed) in the sense of the modal $\Box$ operator has a justification.

Further, Internalization is often required to obtain completeness when a form of the D axiom is present [23, 26, 30]. In deontic settings, this is often the case since obligations are supposed not to contradict each other. Hence restricting the constant specification leads to deductive systems that are not complete.

Conflicting obligations in justification logic have been studied in [9]. Recently, it turned out that this approach can also be used to analyze an epistemic paradox of quantum physics [40].

Moreover, Internalization is essential to obtain realization results. A justification logic realizes a modal logic if, given any theorem F of the modal logic, each occurrence of $\Box$ in F can be replaced with some justification term such that the resulting formula is a theorem of the justification logic. Realization is an important property connecting implicit and explicit modalities.

In the second chapter, we introduce two novel justification logics $\mathsf{JE}_{\mathsf{CS}}$ and $\mathsf{JEM}_{\mathsf{CS}}$ that are the explicit counterparts of the non-normal modal logics E and EM, respectively. On a technical level, the main novelty of our work

[1] It is important to consider axiom necessitation as a rule and not an axiom schema, even though it is a rule without premises. If we considered $c : A$ as an axiom for each $(c, A) \in \mathsf{CS}$, then the notion of an axiom would depend on the notion of a constant specification, which depends on the notion of an axiom. In order to avoid this circularity, we introduce axiom necessitation as a rule.

is the introduction of two types of terms for $\mathsf{JE}_{\mathsf{CS}}$ and $\mathsf{JEM}_{\mathsf{CS}}$. This makes it possible to formalize the characteristic principle of $\mathsf{JE}_{\mathsf{CS}}$ and $\mathsf{JEM}_{\mathsf{CS}}$ as an axiom (and not as a rule) and, therefore, our logics have the Internalization property. Note that we are not the first to use two types of terms. In [22], terms for representing proofs and terms justifying consistency have been combined in constructive justification logic.

We show soundness and completeness of $\mathsf{JE}_{\mathsf{CS}}$ and $\mathsf{JEM}_{\mathsf{CS}}$ with respect to basic models, modular models and fully explanatory modular models. Moreover, we show that the justification logics $\mathsf{JE}_{\mathsf{CS}}$ and $\mathsf{JEM}_{\mathsf{CS}}$ realize the modal logics E and EM, respectively.

From a technical perspective, the case of realizing E is particularly interesting because there we have to deal with a rule that does not respect the polarities of subformulas.

1.2.2 Justification Counterpart for Dyadic Deontic Logic

Using justification logic for resolving deontic puzzles is already discussed by Faroldi in [10–13] 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 [9].

In this chapter we first review the Monadic Deontic Logic (**MDL**) which works with one modal operator relational models for the sematics. We see that a puzzle called *Chisholm's puzzle* cannot be solved in this system. As a result Dyadic Deontic Logic (**DDL**) is viewed where two alethic ($\Box$)and deontic ($\bigcirc$) operators are employed. The semantics used for this system is *preference models* where the set of states is ordered according to a betterness relation and we see how this system is able to solve Chisholm's puzzle.

Having in mind hyperintensionality of deontic modalities, we come up with the idea of representing 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*. Here we extend the idea of using two types of terms in a way so that justification terms represent conditional obligations.

Axiomatization of the justification counterpart of the minimal **DDL** system JAD 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 $\mathsf{JAD}_{\mathsf{CS}}$ with respect to basic models and then preference models are established. The problem with explicit non-normal logics is that the logic is too weak and hardly derives a formula. In this chapter, 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.

So one of the main motivations to develop $\mathsf{JAD}_{\mathsf{CS}}$ is to construct explicit reasons for according-to-duty and contrary-to-duty obligations. An example in chapter 2 shows how such reasons are constructed based on given terms.

Contrary to duty conditionals have been discussed in deontic logic since the publication of Chisholm [7] and is known as "contrary-to-duty" (CTD) problem. According to Chisholm, the problem is raised from formulating those obligations which are generated when some other obligations are violated. To formulate such problems, an ordering on the set of worlds, in terms of preference was created. [17, 27, 28, 41]

Chapter 2

Non-normal Modal Logics and their Justification Counterparts

2.1 Introduction

In this chapter normal and non-normal modal logics are considered as well as neighborhood semantics for non-normal modal logics. Some philosophical puzzles confronting normal modal logic are presented which provide good reason for heading toward non-normal modal logic in deontic interpretation of modality. On the other hand, hyperintensionality is shown as a strong motivation to use justification logic in deontic context. An example for why deontic obligations are hyperintensional is given and finally, a justification counterpart of two varient of non-normal modal logics are presented as systems JE and JEM for modal logics E and EM, where E is the smallest non-normal modal logic and EM is equipped with monotonicity. Soundness and completeness of these two systems with respect to neighborhood models are established. In the last part of this chapter the realization of E in JE is provided.

2.2 Normal Modal Logic

In this section we introduce the language of modal logic and relational semantics.

2.2.1 Syntax of Modal Logic

Definition 1. Formulas of the language of modal logic are defined inductively as follows:

$$\mathsf{Fm} := \bot \mid P_i \mid F \to F \mid \Box F \ .$$

Where $P_i \in \mathsf{Prop}$ is an atomic proposition, $\Box F$ is read as *box* F. We also add a new abbreviation:

$$\Diamond F := \neg \Box \neg F$$

read as diamond F. Notice that unary operators $\neg, \Box, \Diamond$ bind stronger than binary connectors.

Definition 2 (Relational frame and model)**.** A *relational frame* is a tuple (W, R), where W is a nonempty set of states, and $R \subseteq W \times W$ is a relation on W. A *relational model* is a triple $\mathcal{M} = (W, R, V)$ where (W, R) is a relational frame and $V : \mathsf{Prop} \to \mathcal{P}(W)$ is a *valuation function* assigning a set of states to each atomic proposition.

Definition 3 (Truth under relational model)**.** Let $\mathcal{M} = (W, R, V)$ be a relational model. Truth of a modal formula $A \in \mathsf{Fm}$ at state w in $\mathcal{M}$ is defined inductively as follows:

- $\mathcal{M}, w \Vdash p$ iff $w \in V(p)$ for $p \in \mathsf{Prop}$;
- $\mathcal{M}, w \Vdash \neg A$ iff $\mathcal{M}, w \nVdash A$;
- $\mathcal{M}, w \Vdash A \to B$ iff $\mathcal{M}, w \nVdash A$ or $\mathcal{M}, w \Vdash B$;
- $\mathcal{M}, w \Vdash \Box A$ iff for all $v \in W$ if wRv then $\mathcal{M}, v \Vdash A$.

From this we can conclude the truth of $\Diamond A$ as follows:

- $\mathcal{M}, w \Vdash \Diamond A$ iff there is a $v \in W$ such that wRv and $\mathcal{M}, v \Vdash A$.

Definition 4 (Truth set)**.** Let $\mathcal{M} = (W, R, V)$ be a relational model, for each formula $A \in \mathsf{Fm}$ the *truth set* of A in $\mathcal{M}$, denoted by $|A|^{\mathcal{M}}$, is the set of all worlds in which A is true:

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

Definition 5 (Normal modal logic)**.** The smallest set of formulas containing all instances of:

- (CL) classical logic tautologies;
- (K) $\Box(A \to B) \to (\Box A \to \Box B)$;

which is closed under rules:

- (MP) from $A, A \to B$ infer B;
- (Nec) from A infer $\Box A$;

is called *normal modal logic* and is denoted by $\mathbf{K}$. In fact $\mathbf{K}$ is the smallest normal modal logic.

2.2.2 Philosophical Puzzles

One of the features of a normal modal logic is that it is closed under the rule of necessitation, that is if F is valid, then so is $\Box F$. Hence together with axiom K, we can easily derive the rule of monotonicity: Suppose $A \to B$ is valid. By necessitation, we get $\Box(A \to B)$. By axiom K and modus ponens we conclude $\Box A \to \Box B$.

Pacuit [31] mentions several interpretations of $\Box$ for which the validities and rules of inference of normal modal logic can be questioned. A well-known example is the paradox of gentle murder [15], where $\Box$ is read as *ought to*. Consider the statements:

Jones murders Smith. (2.1)

Jones ought not to murder Smith. (2.2)

If Jones murders Smith, then Jones ought to murder Smith gently. (2.3)

These sentences seem to be consistent. However, from (2.1) and (2.3) we infer

Jones ought to murder Smith gently. (2.4)

Moreover, we have the following valid implication

If Jones murders Smith gently, then Jones murders Smith. (2.5)

By the rule of monotonicity, (2.5) implies

If Jones ought to murder Smith gently,
then Jones ought to murder Smith. (2.6)

Now (2.4) and (2.6) together yield

$$\text{Jones ought to murder Smith.} \tag{2.7}$$

This contradicts (2.2). This argument suggests that deontic modal logic should not validate the rules of normal modal logic and thus a semantics different from Kripke semantics is needed. The traditional approach for models of non-normal modal logics is to use neighborhood semantics which is defined in the next section.

Justification logics are parametrized by a constant specification. Hence In epistemic settings, we can calibrate the reasoning power of the agents by adapting the constant specification. Faroldi and Protopopescu [11, 13] suggest using this mechanism also in deontic settings in order to avoid the usual paradoxes. For instance, they discuss Ross' paradox [38], which is:

$$\text{You ought to mail the letter.} \tag{2.8}$$

implies

$$\text{You ought to mail the letter or burn it.} \tag{2.9}$$

The reason is as before. It is a classical validity that

$$\textit{you mail the letter} \text{ implies } \textit{you mail the letter or burn it.} \tag{2.10}$$

By the monotonicity rule we find that (2.8) implies (2.9).

Fardoli and Protopopescu avoid this paradox by restricting the constant specification such that although (2.10) is a logical validity, there will no justification term for it. Thus the rule of monotonicity cannot be derived and there is no paradox.

2.2.3 Hyperintensionality

One of the reasons why Faroldi prefers justification logic over using neighborhood models is that he claims that deontic modalities are *hyperintensional* [10], i.e. they can distinguish between logically equivalent formulas. We first look at one definition of hyperintensionality [13].

Definition 6 (L-hyperintensionality or noncongruentiality)**.** A context **C** is L-hyperintensional if substitution of logically equivalent propositions in **C** cannot happen, i.e.,

$$\Vdash A \leftrightarrow B$$

but

$$\not\models \mathbf{C}(A) \leftrightarrow \mathbf{C}(B).$$

In the following we provide an example to show that the usual modal operator is not hyperintensional [11].

Example 7. Consider the following sentences:

$$\text{You ought to drive.} \tag{2.11}$$

$$\text{You ought to drive or to drive and drink.} \tag{2.12}$$

Intuitively sentences (2.11) and (2.12) are not equivalent, yet their formalizations in modal logic are so. If we represent (2.11) by $\Box A$ and (2.12) by $\Box(A \vee (A \wedge B))$, then we have $A \leftrightarrow A \vee (A \wedge B)$ by propositional reasoning and by the rule of equivalence we infer $\Box A \leftrightarrow \Box(A \vee (A \wedge B))$.

However, hyperintensionality is one of the distinguishing features of justification logics: they are hyperintensional by design. Even if A and B are logically equivalent, we may have that a term t justifying A does not justify 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 $A \to B$ and an application of modus ponens. Thus in justification logic, terms do distinguish between equivalent formulas, which, according to Faroldi, makes it a suitable framework for deontic reasoning.

2.3 Non-normal Modal Logics

In every relational model, the following rules and formulas are valid:

(Dual)	$\Box A \leftrightarrow \neg\Diamond\neg A$
(M)	$\Box(A \wedge B) \to (\Box A \wedge \Box B)$
(C)	$(\Box A \wedge \Box B) \to \Box(A \wedge B)$
(K)	$\Box(A \to B) \to (\Box A \to \Box B)$
(N)	$\Box\top$
(Nec)	from A infer $\Box A$
(RE)	from $A \leftrightarrow B$ infer $\Box A \leftrightarrow \Box B$
(RM)	from $A \to B$ infer $\Box A \to \Box B$

Since in some interpretations of modal language, the validity of some formulas above can be questioned, we consider modal languages excluding these rules and formulas. Modal logics that do not include one or more of the above rules and formulas are called *non-normal modal logics*. (Dual) and (RE) are valid in class of neighborhood frames. The smallest set of formulas containing all instances of classical logic tautologies, (Dual) and is closed under rules (RE) and (MP) is called logic E. The other logical systems will be extensions of E, for example logic EM results from adding rule (RM).

2.3.1 Neighborhood Semantics

Neighborhood models for modal logic are simply models which associate each state from the set of worlds, with a subset space over the set of worlds.

Definition 8 (Neighborhood frame)**.** Let W be a non-empty set of states. A function $N : W \to \mathcal{P}(\mathcal{P}(W))$ is called a *neighborhood function*. A pair (W, N) is called a *Neighborhood frame*.

Remark 9. *One can consider a neighborhood function as a relation. Let* $N : W \to \mathcal{P}(\mathcal{P}(W))$*, then* N *corresponds to* $R_N \subseteq W \times \mathcal{P}(W)$ *such that for any* $w \in W$*,* $X \in \mathcal{P}(W)$*,* wR_NX *if and only if* $X \in N(w)$*.*

Definition 10 (Neighborhood model)**.** A neighborhood model is a tuple $\mathcal{M} = (W, N, V)$, where:

- W is a set of states;
- N is a neighborhood function;
- $V : \mathsf{Prop} \to \mathcal{P}(W)$ is a valuation function assigning a set of states to each atomic proposition.

Definition 11 (Truth under neighborhood model)**.** Suppose $\mathcal{M} = (W, N, V)$ is a neighborhood model and $w \in W$. Truth of formula $A \in \mathsf{Fm}$ at w is defined by induction on structure of A as follows:

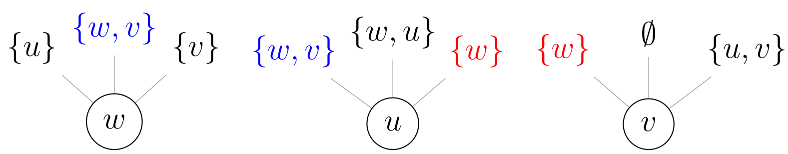
- $\mathcal{M}, w \Vdash p$ iff $w \in V(p)$ for $p \in \mathsf{Prop}$;
- $\mathcal{M}, w \Vdash \neg A$ iff $\mathcal{M}, w \nVdash A$;
- $\mathcal{M}, w \Vdash A \to B$ iff $\mathcal{M}, w \nVdash A$ or $\mathcal{M}, w \Vdash B$;

- $\mathcal{M}, w \Vdash \Box A$ iff $|A|^{\mathcal{M}} \in N(w)$;
- $\mathcal{M}, w \Vdash \Diamond A$ iff $W - |A|^{\mathcal{M}} \notin N(W)$;

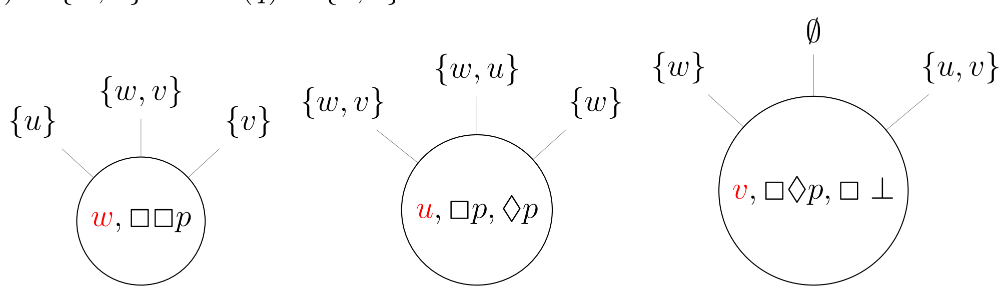
where $|A|^{\mathcal{M}}$ is the *truth set* of A, which is all the worlds in which A is true. Here is an example of truth in a model.

Example 12 (A neighborhood frame)**.** For $W = \{w, u, v\}$, suppose a neighbourhood function as follows:

- $N(w) = \{\{u\}, \{w, v\}, \{v\}\}$
- $N(u) = \{\{w, v\}, \{w\}, \{w, u\}\}$
- $N(v) = \{\{u, v\}, \{w\}, \emptyset\}$



A neighborhood model is now defined by setting valuation V such that $V(p) = \{w, u\}$ and $V(q) = \{u, v\}$.



Since $|\Diamond p|^{\mathcal{M}} = \{u, v\} \in N(v)$, we have: $\mathcal{M}, v \Vdash \Box\Diamond p$
Since $|\Box p|^{\mathcal{M}} = \{u\} \in N(w)$, we have: $\mathcal{M}, w \Vdash \Box\Box p$

2.4 Justification Counterpart for Non-normal Modal Logics

To define the language of our novel justification logic $\mathsf{JE}_{\mathsf{CS}}$, we extend the usual language of justification logic by introducing two types of terms. We consider *proof terms* and *justification terms*, which are inductively built-up from countably many proof constants and variables. So if we denote proof constants by α_i and proof variables by ξ_i, the set of proof terms is defined inductively as follows:

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

Justification terms have the following form:

$$t ::= \mathsf{e}(\lambda) \ .$$

where λ is a proof term. We denote the set of proof terms by PTm and the set of justification terms by JTm. Therefore, the set of all terms is $\mathsf{Tm} := \mathsf{PTm} \cup \mathsf{JTm}$. We use λ, κ, γ for elements of PTm and r, s, t for elements of JTm.

Let Prop be a countable set of atomic propositions. Formulas are inductively defined as follows:

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

where $P_i \in \mathsf{Prop}$, $\lambda \in \mathsf{PTm}$, and $t \in \mathsf{JTm}$. We use Fm for the set of formulas. $\lambda : F$ is read as λ *proofs* F and $[t]F$ is read as t *justifies* F. The axioms of JE are:

j	$\lambda : (F \to G) \to (\kappa : F \to \lambda \cdot \kappa : G)$
j+	$(\lambda : F \vee \kappa : F) \to (\lambda + \kappa) : F$
jt	$\lambda : F \to F$
j4	$\lambda : F \to !\lambda : \lambda : F$
je	$(\lambda : (F \to G) \wedge \lambda : (G \to F)) \to ([\mathsf{e}(\lambda)]F \to [\mathsf{e}(\lambda)]G)$
je+	$([\mathsf{e}(\lambda)]F \vee [\mathsf{e}(\kappa)]F) \to [\mathsf{e}(\lambda + \kappa)]F$

Note that the axioms **j**, **j+**, **jt**, and **j4** are exactly the axioms of the Logic of Proofs. Indeed, dropping **je** and **je+** from $\mathsf{JE}_{\mathsf{CS}}$ and restricting the language to proof terms (hence excluding justification terms) yields the Logic of Proofs.

Axiom **je** shows how justification terms $\mathsf{e}(\lambda)$ are constructed based on proof terms λ; and axiom **je+** is similar to axiom **j+** but for justification terms. It shows that the operation $+$ combines two proof terms such that if $\mathsf{e}(\lambda)$ or $\mathsf{e}(\kappa)$ provides evidence for F, the combined evidence $\mathsf{e}(\lambda + \kappa)$ remains evidence for F.

As we will see later, the axiom **je+** is only used to prove completeness of the logic JE w.r.t. fully explanatory models. It is not needed to establish our other (completeness) results.

In order to define the deductive system for our logic, we first need the notion of a constant specification.

Definition 13 (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 $(\alpha, A) \in \mathsf{CS}$.

Definition 14 (Logic $\mathsf{JE}_{\mathsf{CS}}$)**.** For a constant specification CS, the logic $\mathsf{JE}_{\mathsf{CS}}$ is defined by a Hilbert-style system with the axioms JE and the inference rules modus ponens (MP) and axiom necessitation ($\mathrm{AN}_{\mathsf{CS}}$), given by:

$$\frac{}{\alpha : A} \quad \text{where } (\alpha, A) \in \mathsf{CS} \ .$$

We write $\mathsf{JE}_{\mathsf{CS}} \vdash A$ to express that a formula A is provable in $\mathsf{JE}_{\mathsf{CS}}$. If the deductive system is clear from the context and we only want to stress the constant specification, we simply use $\vdash_{\mathsf{CS}} A$. When the constant specification does not matter or is clear from the context, we drop the subscript CS and write $\vdash A$.

It is a standard result that justification logics with an axiomatically appropriate constant specification internalize their own notion of proof [1,4,21].

Lemma 15 (Internalization)**.** *Let CS be an axiomatically appropriate constant specification. For any formula A with $\vdash A$, there exists a proof term λ such that $\vdash \lambda : A$.*

Moreover, justification logics enjoy a deduction theorem [1,4,21].

Lemma 16 (Deduction)**.** *Let CS be an arbitrary constant specification. For any set Δ of formulas and for any formulas A and B,*

$$\Delta, A \vdash B \quad \textit{iff} \quad \Delta \vdash A \to B \ .$$

2.4.1 Semantics

Let us now turn to semantics. In order to present basic evaluations for $\mathsf{JE}_{\mathsf{CS}}$, we need some operations on sets of formulas.

Definition 17. Let X, Y be sets of formulas and λ be a proof term. We define the following operations:

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

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

$$X \odot Y := \{F \mid F \to G \in X \text{ and } G \to F \in X \text{ for some } G \in Y\} .$$

Definition 18 (Basic evaluation)**.** Let CS be an arbitrary constant specification. A *basic evaluation* for $\mathsf{JE}_{\mathsf{CS}}$ is a function ε that maps atomic propositions to 0 or 1

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

and maps terms to a set of formulas:

$$\varepsilon\colon \mathsf{PTm} \cup \mathsf{JTm} \to \mathcal{P}(\mathsf{Fm}) ,$$

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

1. $\varepsilon(\lambda) \cdot \varepsilon(\kappa) \subseteq \varepsilon(\lambda \cdot \kappa)$;
2. $\varepsilon(\lambda) \cup \varepsilon(\kappa) \subseteq \varepsilon(\lambda + \kappa)$;
3. $F \in \varepsilon(\lambda)$ if $(\lambda, F) \in \mathsf{CS}$;
4. $\lambda : \varepsilon(\lambda) \subseteq \varepsilon(!\lambda)$;
5. $\varepsilon(\lambda) \odot \varepsilon(\mathsf{e}(\lambda)) \subseteq \varepsilon(\mathsf{e}(\lambda))$;
6. $\varepsilon(\mathsf{e}(\lambda)) \cup \varepsilon(\mathsf{e}(\kappa)) \subseteq \varepsilon(\mathsf{e}(\lambda + \kappa))$.

Definition 19 (Truth under a basic evaluation)**.** We define truth of a formula F under a basic evaluation ε inductively as follows:

1. $\varepsilon \nVdash \bot$;
2. $\varepsilon \Vdash P$ iff $\varepsilon(P) = 1$ for $P \in \mathsf{Prop}$;

3. $\varepsilon \Vdash F \to G$ iff $\varepsilon \nVdash F$ or $\varepsilon \Vdash G$;

4. $\varepsilon \Vdash \lambda : F$ iff $F \in \varepsilon(\lambda)$;

5. $\varepsilon \Vdash [t]F$ iff $F \in \varepsilon(t)$.

Definition 20 (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 21 (Basic model)**.** Given an arbitrary CS, a *basic model* for $\mathsf{JE}_{\mathsf{CS}}$ is a basic evaluation that is *factive*.

As expected, we have soundness and completeness with respect to basic models. The following theorem is established in the following.

Theorem 22 (Soundness w.r.t. basic models)**.** *The Logic* $\mathsf{JE}_{\mathsf{CS}}$ *is sound with respect to basic models. For an arbitrary constant specification* CS *and any formula* F*,*

$$\mathsf{JE}_{\mathsf{CS}} \vdash F \implies \varepsilon \Vdash F \textit{ for any basic model } \varepsilon \ .$$

Proof. As usual, the proof is by induction on the length of $\mathsf{JE}_{\mathsf{CS}}$ derivations and a case distinction on the last rule. The only interesting case is when F is an instance of **je**. Suppose

$$\varepsilon \Vdash \lambda : (A \to B) \quad \text{and} \quad \varepsilon \Vdash \lambda : (B \to A) \quad \text{and} \quad \varepsilon \Vdash [\mathsf{e}(\lambda)]A \ .$$

Thus we have

$$(A \to B) \in \varepsilon(\lambda) \quad \text{and} \quad (B \to A) \in \varepsilon(\lambda) \quad \text{and} \quad A \in \varepsilon(\mathsf{e}(\lambda)) \ .$$

By Definition 17 we find $B \in \varepsilon(\lambda) \odot \varepsilon(\mathsf{e}(\lambda))$. Hence, by the definition of basic model we get $B \in \varepsilon(\mathsf{e}(\lambda))$, which is $\varepsilon \Vdash [\mathsf{e}(\lambda)]B$. □

To prove the completeness theorem, we need to know that $\mathsf{JE}_{\mathsf{CS}}$ is consistent.

Lemma 23. *For any constant specification* CS*,* $\mathsf{JE}_{\mathsf{CS}}$ *is consistent.*

Proof. As usual, one can show that $\mathsf{JE}_{\mathsf{CS}}$ is a conservative extension of classical propositional logic. This immediately yields consistency of $\mathsf{JE}_{\mathsf{CS}}$. □

Definition 24. A set of formulas Γ is called $\mathsf{JE}_{\mathsf{CS}}$-*consistent* if for each finite subset $\Sigma \subseteq \Gamma$, we have $\nvdash_{\mathsf{CS}} \bigwedge \Sigma \rightarrow \bot$. The set Γ is *maximal* $\mathsf{JE}_{\mathsf{CS}}$-*consistent* if Γ is consistent and none of its proper supersets is.

As usual, any consistent set can be extended to a maximal consistent set.

Lemma 25 (Lindenbaum)**.** *For each $\mathsf{JE}_{\mathsf{CS}}$-consistent set Δ, there exists a maximal $\mathsf{JE}_{\mathsf{CS}}$-consistent set $\Gamma \supseteq \Delta$.*

Lemma 26. *For any constant specification CS and maximal $\mathsf{JE}_{\mathsf{CS}}$-consistent set Γ, there is a canonical basic model ε^c induced by Γ that is defined as follows:*

$\varepsilon^c(P) := 1$, *if* $P \in \Gamma$ *and* $\varepsilon^c(P) := 0$, *if* $P \notin \Gamma$*;*

$\varepsilon^c(\lambda) := \{F \mid \lambda : F \in \Gamma\}$*;*

$\varepsilon^c(t) := \{F \mid [t]F \in \Gamma\}$.

Proof. First we have to establish that ε^c is a basic evaluation. We only show the condition

$$\varepsilon(\lambda) \odot \varepsilon(\mathsf{e}(\lambda)) \subseteq \varepsilon(\mathsf{e}(\lambda)). \tag{2.13}$$

Suppose $B \in \varepsilon(\lambda) \odot \varepsilon(\mathsf{e}(\lambda))$, which means there is a formula $A \in \varepsilon^c(\mathsf{e}(\lambda))$ with $(A \rightarrow B) \in \varepsilon^c(\lambda)$ and $(B \rightarrow A) \in \varepsilon^c(\lambda)$. By the definition of ε^c, we have

$$\lambda : (A \rightarrow B) \in \Gamma \quad \text{and} \quad \lambda : (B \rightarrow A) \in \Gamma \quad \text{and} \quad [\mathsf{e}(\lambda)]A \in \Gamma\ .$$

Since Γ is a maximal consistent set and

$$(\lambda : (A \rightarrow B) \wedge \lambda : (B \rightarrow A)) \rightarrow ([\mathsf{e}(\lambda)]A \rightarrow [\mathsf{e}(\lambda)]B)$$

is an instance of **je**, we obtain $[\mathsf{e}(\lambda)]B \in \Gamma$. Hence $B \in \varepsilon^c(\mathsf{e}(\lambda))$ and (2.13) is established.

Next, a truth lemma can be established as usual by induction on formula complexity. For all formulas F,

$$F \in \Gamma \quad \text{iff} \quad \varepsilon^c \Vdash F\ . \tag{2.14}$$

Finally, we show that our basic evaluation ε^c is factive and hence a basic model. Suppose $\varepsilon^c \Vdash \lambda : F$. Hence $\lambda : F \in \Gamma$. Since Γ is maximal consistent, we get by axiom **jt** that $F \in \Gamma$. By (2.14) we conclude $\varepsilon^c \Vdash F$. □

Using the Lindenbaum lemma, the canonical basic model and the established truth lemma (2.14), we immediately get the following completeness result.

Theorem 27 (Completeness w.r.t. basic models)**.** *Let* CS *be an arbitrary constant specification. The logic* $\mathsf{JE}_{\mathsf{CS}}$ *is complete with respect to basic models. For any formula* F*,*

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

2.4.2 Neighborhood Semantics and Modular Models

The main purpose of modular models is to connect justification logic to traditional modal logic. To define modular models for $\mathsf{JE}_{\mathsf{CS}}$, we start with a neighborhood model (like for the modal logic E) and assign to each possible world a basic evaluation. This, however, is not enough since these basic evaluations may have nothing to do with the neighborhood structure of the model. Hence we introduce the following principle:

having a specific justification for F must yield
F is obligatory in the sense of the neighborhood structure.

This principle was first introduced in epistemic contexts and is, therefore, called *justification yields belief* (JYB).

Definition 28 (Quasi-model)**.** A quasi-model for $\mathsf{JE}_{\mathsf{CS}}$ is a triple

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

where W is a non-empty set of worlds, N is a neighborhood function and ε is an evaluation function that maps each world to a basic evaluation ε_w.

Definition 29 (Truth in quasi-model)**.** Let $\mathcal{M} = \langle W, N, \varepsilon \rangle$ be a quasi-model. *Truth of a formula at a world* w in a quasi-model is defined inductively as follows:

1. $\mathcal{M}, w \nVdash \bot$;
2. $\mathcal{M}, w \Vdash P$ iff $\varepsilon_w(P) = 1$, for $P \in \mathsf{Prop}$;
3. $\mathcal{M}, w \Vdash F \to G$ iff $\mathcal{M}, w \nVdash F$ or $\mathcal{M}, w \Vdash G$;

4. $\mathcal{M}, w \Vdash \lambda : F$ iff $F \in \varepsilon_w(\lambda)$;

5. $\mathcal{M}, w \Vdash [t]F$ iff $F \in \varepsilon_w(t)$.

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

Remark 30. *The neighborhood function plays no rule in the definition of truth in quasi-models. Hence truth in quasi-models is local to a possible world. Let $\mathcal{M} = \langle W, N, \varepsilon \rangle$ be a quasi-model. For any $w \in W$ and any formula F,*

$$\mathcal{M}, w \Vdash F \quad \textit{iff} \quad \varepsilon_w \Vdash F \ . \tag{2.15}$$

Definition 31 (Factive quasi-model)**.** A quasi-model $\mathcal{M} = \langle W, N, \varepsilon \rangle$ is *factive* if for each world w, we have that for any formula $\lambda : F$,

$$\mathcal{M}, w \Vdash \lambda : F \quad \text{implies} \quad \mathcal{M}, w \Vdash F \ .$$

Remark 32. *Let $\mathcal{M} = \langle W, N, \varepsilon \rangle$ be a quasi-model. The* truth set *of a formula F, denoted by $|F|^{\mathcal{M}}$, is the set of all worlds in which F is true, i.e.,*

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

Further, we define

$$\Box_w := \{F \mid |F|^{\mathcal{M}} \in N(w)\} \quad .$$

Looking back at neighborhood models for E*, it is easy to see that $F \in \Box_w$ means (modulo the different language that we are using) that $\Box F$ holds at world w. As a result, we can formulate the principle of justification yields belief as follows:*

$$\textit{for any } t \in \mathsf{JTm} \textit{ and } w \in W, \textit{ we have that } \varepsilon_w(t) \subseteq \Box_w \ . \tag{JYB}$$

Definition 33 (Modular model)**.** A $\mathsf{JE}_{\mathsf{CS}}$ *modular model* is a quasi-model for $\mathsf{JE}_{\mathsf{CS}}$ that is factive and satisfies (JYB).

$\mathsf{JE}_{\mathsf{CS}}$ is sound and complete with respect to modular models. A proof of the following theorem is given as follows.

Theorem 34 (Soundness and completeness w.r.t. modular models)**.** *Let* CS *be an arbitrary constant specification. For each formula F we have*

$$\mathsf{JE}_{\mathsf{CS}} \vdash F \quad \textit{iff} \quad \mathcal{M} \Vdash F \textit{ for all } \mathsf{JE}_{\mathsf{CS}} \textit{ modular models } \mathcal{M}.$$

Proof. To prove soundness, suppose $\mathcal{M} = \langle W, N, \varepsilon \rangle$ is a $\mathsf{JE}_{\mathsf{CS}}$ modular model, and $\mathsf{JE}_{\mathsf{CS}} \vdash A$. We need to show that A is true in every world $w \in W$. Assume that ε_w is a basic model. Then by soundness with respect to basic models we get $\varepsilon_w \Vdash A$ and by (2.15) we conclude $\mathcal{M}, w \Vdash A$. It remains to show that ε_w indeed is a basic model, i.e. that it is factive. Suppose $\varepsilon_w \Vdash \lambda : F$. By (2.15) we get $\mathcal{M}, w \Vdash \lambda : F$. By factivity of modular models we get $\mathcal{M}, w \Vdash F$ and by (2.15) again we conclude $\varepsilon_w \Vdash F$.

For completeness, suppose that $\mathsf{JE}_{\mathsf{CS}} \nvdash F$. Since $\mathsf{JE}_{\mathsf{CS}}$ is complete with respect to basic models, there is a $\mathsf{JE}_{\mathsf{CS}}$-basic model ε with $\varepsilon \nVdash F$. Now we construct a quasi-model $\mathcal{M} := \langle \{w\}, N, \varepsilon' \rangle$ with $\varepsilon'_w := \varepsilon$ and

$$N(w) = \{|G|^{\mathcal{M}} \mid G \in \varepsilon'_w(t), \text{ for any } t \in \mathsf{JTm}\}.$$

By (2.15) we find $\mathcal{M}, w \nVdash F$. It only remains to show that $\mathcal{M}$ is a modular model: Factivity follows immediately from (2.15) and the fact that ε is factive. To show (JYB), we suppose $F \in \varepsilon'_w(t)$. By the definition of N we get $|F|^{\mathcal{M}} \in N(w)$, which means $F \in \Box_w$. □

It is natural to ask whether every obligatory formula in a modular model is justified by a justification term.

Definition 35 (Fully explanatory modular model)**.** A $\mathsf{JE}_{\mathsf{CS}}$ modular model $\mathcal{M} = \langle W, N, \varepsilon \rangle$ is *fully explanatory* if for any $w \in W$ and any formula F,

$$|F|^{\mathcal{M}} \in N(w) \quad \text{implies} \quad F \in \varepsilon_w(t) \text{ for some } t \in \mathsf{JTm}\ .$$

The fully explanatory property can be seen as the converse of justification yields belief. In fully explanatory models we have that for each world w,

$$\bigcup_{t \in \mathsf{JTm}} \varepsilon_w(t) = \Box_w\ .$$

For any axiomatically appropriate constant specification CS, we can show that $\mathsf{JE}_{\mathsf{CS}}$ is sound and complete with respect to fully explanatory $\mathsf{JE}_{\mathsf{CS}}$ modular models. In order to obtain this, we need monotonicity of the e-operation with respect to + as expressed in axiom **je+**. The proof is presented in the following. Before starting to prove the theorem, we need an auxiliary notion:

Definition 36 (Proof set)**.** Let M_{JE} be the set of all maximal $\mathsf{JE}_{\mathsf{CS}}$-consistent sets of formulas. We set

$$\mathsf{M}_{\mathsf{JE}} := \{\Gamma \mid \Gamma \text{ is a maximal } \mathsf{JE}_{\mathsf{CS}}\text{-consistent set }\}\ .$$

For any formula F we define $\|F\| := \{\Gamma \mid \Gamma \in \mathsf{M}_{\mathsf{JE}} \text{ and } F \in \Gamma\}$, called the *proof set* of F.

Proof sets share a number of properties, which are given in the following lemma.

Lemma 37. *For formulas F, G following properties hold:*

1. $\|F \wedge G\| = \|F\| \cap \|G\|$;
2. $\|\neg F\| = \mathsf{M}_{\mathsf{JE}} \setminus \|F\|$;
3. $\|F \vee G\| = \|F\| \cup \|G\|$;
4. $\|F\| \subseteq \|G\|$ *iff* $\vdash F \rightarrow G$;
5. $\vdash (F \leftrightarrow G)$ *iff* $\|F\| = \|G\|$;
6. $\|\lambda : G\| \subseteq \|G\|$ *for any proof term* λ.

Proof. Let only show claim 4. The claim from right to left immediately follows from closure of maximal consistent sets under modus ponens. For the other direction, suppose $\|F\| \subseteq \|G\|$, but not $\vdash F \rightarrow G$. Then $\neg(F \rightarrow G)$ is consistent and by Lindenbaum's Lemma there is a maximal consistent set $\Gamma \ni \neg(F \rightarrow G)$. This means $F, \neg G \in \Gamma$. Since $F \in \Gamma$ and $\|F\| \subseteq \|G\|$, we get $G \in \Gamma$, which contradicts $\neg G \in \Gamma$. ☐

Theorem 38 (Soundness and completeness for fully explanatory modular models)**.** *Let* CS *be an axiomatically appropriate constant specification.* $\mathsf{JE}_{\mathsf{CS}}$ *is sound and complete with respect to fully explanatory* $\mathsf{JE}_{\mathsf{CS}}$ *modular models.*

Proof. Soundness is a direct consequence of soundness for the class of $\mathsf{JE}_{\mathsf{CS}}$ modular models.

To prove completeness, we define a canonical model $\mathcal{M}^c := \langle W^c, N^c, \varepsilon^c \rangle$ by

- $W^c := \mathsf{M}_{\mathsf{JE}}$;
- $N^c \colon W^c \rightarrow \mathcal{P}(\mathcal{P}(W^c))$, such that for each $\Gamma \in W^c$,

$$\|F\| \in N^c(\Gamma) \text{ iff } [\mathsf{e}(\gamma)]F \in \Gamma \text{ for some } \mathsf{e}(\gamma) \in \mathsf{JTm} \ ;$$

- $\varepsilon^c_\Gamma(t) := \{F \mid [t]F \in \Gamma\}$ and $\varepsilon^c_\Gamma(\lambda) := \{F \mid \lambda : F \in \Gamma\}$.

Before establishing that this canonical model is a fully explanatory modular model, we show that the neighborhood function is well-defined. The issue is that different formulas may have the same proof set. Thus we need to show the following lemma.

Lemma 39. *Let* CS *be axiomatically appropriate. The neighborhood mapping* N^c *is well-defined: for any* $\Gamma \in \mathsf{M}_{\mathsf{JE}}$ *and any formulas* F, G*, if* $\|F\| \in N^c(\Gamma)$ *and* $\|F\| = \|G\|$*, then there is a term* $\mathsf{e}(\lambda) \in \mathsf{JTm}$ *such that* $[\mathsf{e}(\lambda)]G \in \Gamma$*.*

Proof. Let F, G be two formulas such that $\|F\| = \|G\|$. For some $\Gamma \in \mathsf{M}_{\mathsf{JE}}$, suppose $\|F\| \in N^c(\Gamma)$. By the definition of the canonical model we have $[\mathsf{e}(\gamma)]F \in \Gamma$ for some $\mathsf{e}(\gamma) \in \mathsf{JTm}$. By Lemma 37, we have $\vdash_{\mathsf{JE}} F \leftrightarrow G$ and so $\vdash_{\mathsf{JE}} G \to F$ and $\vdash_{\mathsf{JE}} F \to G$. Since CS is axiomatically appropriate, there are proof terms δ_1, δ_2 such that $\vdash_{\mathsf{JE}} \delta_1 : (F \to G)$ and $\vdash_{\mathsf{JE}} \delta_2 : (G \to F)$. By the **j+** axiom, there is a term $\lambda = (\delta_1 + \delta_2 + \gamma)$ such that $\vdash_{\mathsf{JE}} \lambda : (F \to G)$ and $\vdash_{\mathsf{JE}} \lambda : (G \to F)$. By maximal consistency of Γ we get

$$([\mathsf{e}(\lambda)]F \to [\mathsf{e}(\lambda)]G) \in \Gamma \tag{2.16}$$

Further, we get by axiom **je+** and maximal consistency of Γ that $[\mathsf{e}(\lambda)]F \in \Gamma$ and thus by (2.16) we conclude $[\mathsf{e}(\lambda)]G \in \Gamma$. □

Next we can establish the truth lemma.

Lemma 40 (Truth lemma)**.** *For each formula* F*, we have* $|F|^{\mathcal{M}^c} = \|F\|$*.*

Proof. As usual the proof is by induction on the structure of F. We only show the case when F is $[t]G$. We have the following equivalences: $\Gamma \in |[t]G|^{\mathcal{M}^c}$ iff $\mathcal{M}^c, \Gamma \Vdash [t]G$ iff $G \in \varepsilon^c_\Gamma(t)$ iff $[t]G \in \Gamma$ iff $\Gamma \in \|[t]G\|$. □

Now we show that the canonical model is a modular model. First, we show that $W^c \neq \emptyset$. Recall that by Lindenbaum's Lemma, for every consistent set of formulas Γ, there exist a maximally consistent set of formulas that contains Γ. Since the empty set is consistent, by Lindenbaum's Lemma, there is a maximal consistent set that contains the empty set and is an element of W^c.

Next we show factivity. Suppose $\mathcal{M}^c, \Gamma \Vdash \lambda : G$. By the truth lemma we get $\lambda : G \in \Gamma$. Since Γ is maximally consistent, we obtain by axiom **jt** that $G \in \Gamma$. Again by the truth lemma we conclude $\mathcal{M}^c, \Gamma \Vdash G$.

Now we show that the canonical model satisfies justification yields belief (JYB). Suppose $F \in \varepsilon^c_\Gamma(t)$ for some justification term t, some formula F, and some $\Gamma \in W^c$. The term t has the form $\mathsf{e}(\lambda)$ and by the definition of

ε^c_Γ we find $[\mathsf{e}(\lambda)]F \in \Gamma$. By the definition of N^c we obtain $\|F\| \in N^c(\Gamma)$. Thus, using the the truth lemma, we get $|F|^{\mathcal{M}^c} \in N^c(\Gamma)$. Thus (JYB) is established.

It remains to show that the canonical model is fully explanatory. Suppose $|F|^{\mathcal{M}^c} \in N^c(\Gamma)$ for some formula F and some $\Gamma \in W^c$. By the truth lemma we find $\|F\| \in N^c(\Gamma)$. By the definition of N^c, this implies $[t]F \in \Gamma$ for some justification term t. By the definition of ε^c_Γ we finally conclude $F \in \varepsilon^c_\Gamma(t)$. $\square$

2.5 Justification Counterpart of Monotonic Non-normal Logic

There are several applications for which the modal logic E is too weak and one considers the extension of E with the axiom $\Box(A \wedge B) \to (\Box A \wedge \Box B)$ or, equivalently, with the rule

$$\frac{A \to B}{\Box A \to \Box B} \quad .$$

The resulting logic is called EM. In this section we introduce an explicit counterpart JEM of the modal logic EM.

First, we adapt the language as follows. Proof terms are given as before but without $+$:

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

The set of *justification terms* is built up inductively, starting from a countable set of justification variables x_i, by:

$$t ::= x_i \mid t + t \mid \mathsf{m}(\lambda, t)$$

where λ is a proof term. Formulas are then built using this extended set of justification terms. It will always be clear from the context whether we work with the basic language for JE or with the extended language for JEM.

The axioms of JEM consist of the axioms **j**, **jt**, and **j4** together with

jm $\quad \lambda : (F \to G) \to ([t]F \to [\mathsf{m}(\lambda, t)]G)$.

$\mathbf{j+_1}$ $\quad ([t]F \vee [s]F) \to [t + s]F$.

For a constant specification CS, we now consider axioms of JEM and the system $\mathsf{JEM}_{\mathsf{CS}}$ consists of the axioms of JEM plus the rules of modus ponens and axiom necessitation. Note that Internalization and the Deduction theorem hold for $\mathsf{JEM}_{\mathsf{CS}}$, too. Axiom $\mathbf{j}+_1$ will be used in the realization proof, but we do not need $+$ for proof terms in JEM and thus we dispense with axiom $\mathbf{j}+$. For JEM, we can establish completeness w.r.t. fully explanatory models without using axiom $\mathbf{je}+$, thus we do not include it in JEM.

A *basic evaluation for* $\mathsf{JEM}_{\mathsf{CS}}$ is defined similar to a basic evaluation for $\mathsf{JE}_{\mathsf{CS}}$ with the conditions for $+$ on proof terms and for e dropped and with the additional requirements that for arbitrary terms $\lambda \in \mathsf{PTm}$ and $t, s \in \mathsf{JTm}$:

1. $\varepsilon(\lambda) \cdot \varepsilon(t) \subseteq \varepsilon(\mathsf{m}(\lambda, t))$;
2. $\varepsilon(t) \cup \varepsilon(s) \subseteq \varepsilon(t + s)$.

Further we define a *monotonic basic model* (for $\mathsf{JEM}_{\mathsf{CS}}$) as a basic evaluation for $\mathsf{JEM}_{\mathsf{CS}}$ that is factive.

Similar to $\mathsf{JE}_{\mathsf{CS}}$, we can show that $\mathsf{JEM}_{\mathsf{CS}}$ is sound and complete with respect to monotonic basic models.

Theorem 41. *Let* CS *be an arbitrary constant specification. The logic* $\mathsf{JEM}_{\mathsf{CS}}$ *is sound and complete with respect to monotonic basic models. For any formula* F*,*

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

Proof. The proof is similar to soundness and completeness of logic $\mathsf{JE}_{\mathsf{CS}}$ with respect to basic models. □

Now we are going to adapt modular models to $\mathsf{JEM}_{\mathsf{CS}}$. A neighborhood function N for a non-empty set of worlds W is called *monotonic* provided that for each $w \in W$ and for each $X \subseteq W$,

$$\text{if } X \in N(w) \text{ and } X \subseteq Y \subseteq W \text{ then } Y \in N(w).$$

A *monotonic quasi-model* for $\mathsf{JEM}_{\mathsf{CS}}$ is defined like a quasi-model for $\mathsf{JE}_{\mathsf{CS}}$ but we use a monotonic neighborhood function and each world is mapped to a basic evaluation for $\mathsf{JEM}_{\mathsf{CS}}$. A *monotonic modular model* is then defined like a modular model but the underlying quasi-model is required to be monotonic. As for $\mathsf{JE}_{\mathsf{CS}}$ we get completeness or $\mathsf{JEM}_{\mathsf{CS}}$ with respect to monotonic modular models.

Theorem 42. *Let* CS *be an arbitrary constant specification. For each formula* F *we have*

$$\mathsf{JEM}_{\mathsf{CS}} \vdash F \quad \textit{iff} \quad \mathcal{M} \Vdash F \textit{ for all } \mathsf{JEM}_{\mathsf{CS}} \textit{ monotonic modular models } \mathcal{M}.$$

Proof. The proof is similar to soundness and completeness of logic $\mathsf{JE}_{\mathsf{CS}}$ with respect to modular models. □

Theorem 43. *Let* CS *be an axiomatically appropriate constant specification.* $\mathsf{JEM}_{\mathsf{CS}}$ *is sound and complete with respect to fully explanatory* $\mathsf{JEM}_{\mathsf{CS}}$ *monotonic modular models.*

Proof. Soundness is a result of soundness of logic $\mathsf{JE}_{\mathsf{CS}}$ with respect to fully explanatory models.

To achieve completeness with respect to fully explanatory monotonic modular models, one needs some additional construction to guarantee that the neighborhood function constructed in the canonical model is monotonic.

For any set $\mathcal{U} \subseteq \mathcal{P}(W)$, we say $\mathcal{U}$ is *supplemented* or *monotonic*, if $X \in \mathcal{U}$ and $X \subseteq Y \subseteq W$ then $Y \in \mathcal{U}$. So for any $\mathcal{X} \subseteq \mathcal{P}(W)$, we denote the closure of $\mathcal{X}$ under supplementation by $(\mathcal{X})^{mon}$. Moreover, a *proof set* is defined as:

$$\|F\| := \{\Gamma \mid \Gamma \in \mathsf{M}_{\mathsf{JEM}} \text{ and } F \in \Gamma\},$$

where $\mathsf{M}_{\mathsf{JEM}}$ is the set of all maximal $\mathsf{JEM}_{\mathsf{CS}}$-consistent sets.

Now we define the canonical model $\mathcal{M}^c_{mon} := \langle W^c, N^c_{mon}, \varepsilon^c \rangle$, such that:

- $W^c := \mathsf{M}_{\mathsf{JEM}}$;
- $N^c_{mon} := (N^c_{min})^{mon}$, such that:

$$N^c_{min}(\Gamma) = \{\|F\| \mid [t]F \in \Gamma, \text{ for some } t \in \mathsf{JTm}\} \ ;$$

- $\varepsilon^c_\Gamma(t) := \{F \mid [t]F \in \Gamma\}$ and $\varepsilon^c_\Gamma(\lambda) := \{F \mid \lambda : F \in \Gamma\}$.

We will only show that N^c_{min} is well-defined and that $\mathcal{M}^c_{mon}$ is fully explanatory. The rest of the completeness proof is similar to the case for $\mathsf{JE}_{\mathsf{CS}}$.

To establish that N^c_{min} is well defined, assume that F, G are two formulas such that $\|F\| = \|G\|$ with $\|F\| \in N^c(\Gamma)$ for some $\Gamma \in \mathsf{M}_{\mathsf{JEM}}$. Thus $[s]F \in \Gamma$ for some justification term s. By Lemma 37 we find $\vdash_{\mathsf{JEM}_{\mathsf{CS}}} F \to G$. Since CS is axiomatically appropriate, there is a proof term λ with $\vdash_{\mathsf{JEM}_{\mathsf{CS}}} \lambda : (F \to G)$. By axiom **jm**, we conclude $[\mathsf{m}(\lambda, s)]G \in \Gamma$.

To show that $\mathcal{M}^c_{mon}$ is fully explanatory, suppose $|G|^{\mathcal{M}^c_{mon}} \in N^c_{mon}(\Gamma)$ for some formula G and some $\Gamma \in \mathsf{M}_{\mathsf{JEM}}$. By truth lemma for $\mathcal{M}^c_{mon}$, we have $\|G\| \in N^c_{mon}(\Gamma)$. By definition of N^c_{mon} it means that either $\|G\| \in N^c_{min}(\Gamma)$ or there exists a formula H such that $\|H\| \in N^c_{min}(\Gamma)$ and $\|H\| \subseteq \|G\|$. In the former case by definition of canonical model $[t]G \in \Gamma$ for some $t \in \mathsf{JTm}$. In the latter case, we find $[t]H \in \Gamma$ for some $t \in \mathsf{JTm}$. Moreover, by Lemma 37 we obtain $\vdash_{\mathsf{JEM}_{\mathsf{CS}}} H \to G$. Since CS is axiomatically appropriate, there is a proof term λ such that $\vdash_{\mathsf{JEM}_{\mathsf{CS}}} \lambda : (H \to G)$. By axiom **jm**, there is a term $\mathsf{m}(\lambda, t)$ such that $\vdash_{\mathsf{JEM}_{\mathsf{CS}}} [t]H \to [\mathsf{m}(\lambda, t)]G$. We conclude $[\mathsf{m}(\lambda, t)]G \in \Gamma$. $\square$

2.6 Realization

This section is concerned with the exact relationship between some non-normal modal logic M and its explicit counterpart J. Let Fm^{M} denote the set of formulas from modal logic and Fm^{J} the set of all justification logic formulas (for E or for EM) that do not contain subformulas of the form $\lambda : F$. There is the so-called forgetful translation $^\circ$ from Fm^{J} to Fm^{M} given by

$$\bot^\circ := \bot \qquad P^\circ := P \qquad (A \to B)^\circ := A^\circ \to B^\circ \qquad ([t]A)^\circ := \Box A^\circ \ .$$

However, we are mainly interested in the converse direction. A *realization* is a mapping from Fm^{M} to Fm^{J} such that for all $A \in \mathsf{Fm}^{\mathsf{M}}$, we have $(r(A))^\circ = A$.

Now the question is whether a realization theorem holds, i.e. given a modal logic M and a justification logic J, does there exist a realization r such that for all $A \in \mathsf{Fm}^{\mathsf{M}}$, we have that $\mathsf{M} \vdash A$ implies $\mathsf{J} \vdash r(A)$?

In order to establish such a realization theorem, we need the notion of a schematic constant specification.

Definition 44. A constant specification CS is called *schematic* if it satisfies the following: for each constant c, the set of axioms $\{A \mid (c, A) \in \mathsf{CS}\}$ consists of all instances of one or several (possibly zero) axioms schemes of the justification logic.

Schematic constant specifications are important in the context of substitutions, where a substitution replaces atomic propositions with formulas, proof variables with proof terms, and justification variables with justification terms. The following lemma is standard [21].

Lemma 45. *Let CS be a schematic constant specification. We have for any set of formulas Δ, any formula A, and any substitution σ*

$$\Delta \vdash A \quad \textit{implies} \quad \Delta\sigma \vdash A\sigma \ .$$

In order to show a realization result, we further need a cut-free sequent calculus for the given modal logic. The system GE is given by the following propositional axioms and rules, the structural rules, and the rule (**RE**). If we replace (**RE**) with (**RM**), we obtain the system GM. In these systems, a *sequent* is an expression of the form $\Gamma \supset \Delta$ where Γ and Δ are finite multisets of formulas.

Propositional axioms and rules:

$$P \supset P \qquad\qquad \bot \supset$$

$$\frac{\Gamma \supset \Delta, A \qquad B, \Gamma \supset \Delta}{A \to B, \Gamma \supset \Delta}\,(\to\supset) \qquad\qquad \frac{A, \Gamma \supset \Delta, B}{\Gamma \supset \Delta, A \to B}\,(\supset\to)$$

Structural rules:

$$\frac{\Gamma \supset \Delta}{A, \Gamma \supset \Delta}\,(w \supset) \qquad\qquad \frac{\Gamma \supset \Delta}{\Gamma \supset \Delta, A}\,(\supset w)$$

$$\frac{A, A, \Gamma \supset \Delta}{A, \Gamma \supset \Delta}\,(c \supset) \qquad\qquad \frac{\Gamma \supset \Delta, A, A}{\Gamma \supset \Delta, A}\,(\supset c)$$

Modal rules:

$$\frac{A \supset B \qquad B \supset A}{\Box A \supset \Box B}\,(\mathbf{RE}) \qquad\qquad \frac{A \supset B}{\Box A \supset \Box B}\,(\mathbf{RM})$$

The systems GE and GM are sound and complete [19, 25].

Theorem 46. *For each modal logic formula A, we have*

1. $\mathsf{GE} \vdash\ \supset A$ *iff* $\mathsf{E} \vdash A$*;*
2. $\mathsf{GM} \vdash\ \supset A$ *iff* $\mathsf{EM} \vdash A$*.*

2.6.1 Realization of Modal Logic E in $\mathsf{JE}_{\mathsf{CS}}$

To realize the non-normal modal logic E in $\mathsf{JE}_{\mathsf{CS}}$, we need the following notions. Let $\mathcal{D}$ be a GE-proof of $\supset A$. We say that occurrences of $\Box$ in $\mathcal{D}$ are *related* if they occur in the same position in related formulas of premises and conclusions of a rule instance in $\mathcal{D}$. We close this relationship of related occurrences under transitivity.

All occurrences of $\Box$ in $\mathcal{D}$ naturally split into disjoint *families* of related $\Box$-occurrences.

We call a family of $\Box$-occurrences *essential* if at least one of its members is a $\Box$-occurrence introduced by an instance of (**RE**).

We say two essential families are *equivalent* if there is an instance of (**RE**) rule which introduces $\Box$-occurrences to each of these two families.

We close this relationship of equivalent families under transitivity. This equivalence relation makes a partition on the set of all essential families. Hence by a *class of equivalent essential families* we mean the set of all essential families which are equivalent.

Theorem 47 (Constructive realization of logic E)**.** *For any axiomatically appropriate and schematic constant specification* CS*, there exist a realization* r *such that for each formula* $A \in \mathsf{Fm}^{\mathsf{M}}$*, we have*

$$\mathsf{GE} \vdash \supset A \quad \textit{implies} \quad \mathsf{JE}_{\mathsf{CS}} \vdash r(A) \ .$$

We will not present the full proof of the realization theorem. The essence is the same as in the proof of the constructive realization theorem for the Logic of Proofs [1, 21].

Let $\mathcal{D}$ be the GE-proof of $\supset A$. The realization r is constructed by the following algorithm. We reserve a large enough set of proof variables as *provisional variables*.

1. For each non-essential family of $\Box$-occurrences, replace all occurrences of $\Box$ by $[\mathsf{e}(\xi)]$ such that each family has a distinct proof variable ξ.

2. For a class of equivalent essential families F, enumerate all instances of **RE** rules which introduce a $\Box$-occurrence to this class of families. Let n_f denote the number of all such **RE** rule instances. Replace each $\Box$ of this class of families with a justification term $[\mathsf{e}(\zeta_1 + ... + \zeta_{n_f})]$ where each ζ_i is a provisional variable. Applying this step for all classes of equivalent essential families yields a derivation tree $\mathcal{D}'$ labeled by Fm^{J}-formulas.

3. Replace all provisional variables in $\mathcal{D}'$ from the leaves toward the root. By induction on the depth of a node in $\mathcal{D}'$, we show that after each

replacement, the resulting sequent of this step is derivable in $\mathsf{JE}_{\mathsf{CS}}$ where for finite multisets Γ and Δ of Fm^{J}-formulas, derivability of $\Gamma \supset \Delta$ means $\Gamma \vdash_{\mathsf{CS}} \bigvee \Delta$.

According to the enumeration defined in 2, the ith occurrence of **RE** rule in $\mathcal{D}'$ is labelled by:

$$\frac{A \supset B \qquad B \supset A}{[\mathsf{e}(\kappa_1 + \ldots + \zeta_i + \ldots + \kappa_{n_f})]A \supset [\mathsf{e}(\kappa_1 + \ldots + \zeta_i + \ldots + \kappa_{n_f})]B}\ (\mathbf{RE})$$

where the κ's are proof terms and ζ_i is a provisional variable.

By I.H. we have $A \vdash_{\mathsf{CS}} B$ and $B \vdash_{\mathsf{CS}} A$. By the Deduction Theorem we get $\vdash_{\mathsf{CS}} A \to B$ and $\vdash_{\mathsf{CS}} B \to A$. By the internalization lemma there are proof terms such $\lambda_{i_1}, \lambda_{i_2}$ that $\vdash_{\mathsf{CS}} \lambda_{i_1} : (A \to B)$ and $\vdash_{\mathsf{CS}} \lambda_{i_2} : (B \to A)$. Replace ζ_i globaly in the whole derivation $\mathcal{D}'$ with $(\lambda_{i_1} + \lambda_{i_2})$.

Now by axiom **j+** we conclude

$$\vdash_{\mathsf{CS}} (\kappa_1 + \ldots + (\lambda_{i_1} + \lambda_{i_2}) + \ldots + \kappa_{n_f}) : (A \to B)$$

and similarly

$$\vdash_{\mathsf{CS}} (\kappa_1 + \ldots + (\lambda_{i_1} + \lambda_{i_2}) + \ldots + \kappa_{n_f}) : (B \to A).$$

By axiom **je** we find

$$\vdash_{\mathsf{CS}} [\mathsf{e}(\kappa_1 + \ldots + (\lambda_{i_1} + \lambda_{i_2}) + \ldots + \kappa_{n_f})]A \to [\mathsf{e}(\kappa_1 + \ldots + (\lambda_{i_1} + \lambda_{i_2}) + \ldots + \kappa_{n_f})]B$$

Note that since CS is schematic and by Lemma 45, replacing ζ_i with $(\lambda_{i_1} + \lambda_{i_2})$ in $\mathcal{D}'$ does not affect already established derivability results.

2.6.2 Realization of Modal Logic EM in $\mathsf{JEM}_{\mathsf{CS}}$

In order to realize the modal logic EM in $\mathsf{JEM}_{\mathsf{CS}}$, we need some technical notions about occurrences of $\Box$-operators.

We assign a *positive or negative polarity* to each subformula occurrence within a fixed formula A as follows:

1. To the only occurrence of A in A we assign the positive polarity.

2. If a polarity is assigned to a subformula of the form $B \to C$ in A, then the same polarity is assigned to C and opposite polarity is assigned to B.

3. If a polarity is already assigned to a subformula of the form $\Box B$ in A, then the same polarity is assigned to B.

Let $\Box B$ be a sub-formula of A. If $A \in \Delta$ in a sequent $\Gamma \supset \Delta$, then the $\Box$-operator of $\Box B$ has the same *polarity* as the subformula occurrence of $\Box B$ in A. If $A \in \Gamma$ in a sequent $\Gamma \supset \Delta$, then the $\Box$-operator of $\Box B$ has the opposite *polarity* as the subformula occurrence of $\Box B$ in A.

Remark 48. *All rules of* GM *respect the polarities of* $\Box$*-operators. The rule* (**RM**) *introduces negative* $\Box$*-occurrence to the left side, and positive* $\Box$*-occurrence to the right side of the conclusion.*

In the following we consider the system GM. Let $\mathcal{D}$ be a derivation in GM. Again, we say that occurrences of $\Box$ in $\mathcal{D}$ are *related* if they occur in the same position in related formulas of premises and conclusions of a rule instance in $\mathcal{D}$. We close this relationship of related occurrences under transitivity.

All occurrences of $\Box$ in $\mathcal{D}$ naturally split into disjoint *families* of related $\Box$-occurrences. We call such a family *essential* if at least one of its members is a positive $\Box$-occurrence introduced by an instance of (**RM**).

Now we are ready to formulate and prove the realization theorem.

Definition 49 (Normal realization)**.** A realization is called *normal* if all negative occurrences of $\Box$ are realized by distinct justification variables.

Theorem 50 (Constructive realization)**.** *For any axiomatically appropriate and schematic constant specification* CS*, there exist a normal realization* r *such that for each formula* $A \in \mathsf{Fm}^{\mathsf{M}}$*, we have*

$$\mathsf{GM} \vdash\ \supset A \quad \textit{implies} \quad \mathsf{JEM}_{\mathsf{CS}} \vdash r(A)\ .$$

Let $\mathcal{D}$ be the GM-proof of $\supset A$. The realization r is constructed by the following algorithm. We reserve a large enough set of justification variables as *provisional variables.*

1. For each non-essential family of $\Box$-occurrences, replace all occurrences of $\Box$ by $[x]$ such that each family has a distinct justification variable.

2. For an essential family of $\Box$-occurrences, enumerate all occurrences of (**RM**) rules that introduce a $\Box$-operator to this family. Let n be the number of such occurrences. Replace each $\Box$-occurrence of this family with $[v_1+...+v_n]$ where each v_i is a fresh provisional variable. Applying this step for all essential families yields a derivation tree $\mathcal{D}'$ labeled by Fm^{J}-formulas.

3. Replace all provisional justification variables in $\mathcal{D}'$ from the leaves toward the root. By induction on the depth of a node in $\mathcal{D}'$, we show that after each replacement, the resulting sequent of this step is derivable in $\mathsf{JEM}_{\mathsf{CS}}$.

   Let us show the case of an instance of (**RM**) with number i in an essential family. The corresponding node in $\mathcal{D}'$ is labelled by

$$\frac{A \supset B}{[x]A \supset [v_1 + \ldots + v_i + \ldots + v_n]B}\ (\mathbf{RM})$$

   where the v's are justification terms and v_i is a justification variable. By I.H. we get $A \vdash_{\mathsf{CS}} B$. By the Deduction Theorem we get $\vdash_{\mathsf{CS}} A \to B$ and Internalization yields a proof term λ with $\vdash_{\mathsf{CS}} \lambda : (A \to B)$. By **jm** we get $\vdash_{\mathsf{CS}} [x]A \to [\mathsf{m}(\lambda, x)]B$. Hence, again by the Deduction Theorem, we find $[x]A \vdash_{\mathsf{CS}} [\mathsf{m}(\lambda, x)]B$ and thus

$$[x]A \vdash_{\mathsf{CS}} [v_1 + \ldots + \mathsf{m}(\lambda, x) + \ldots + v_n]B$$

   by axiom $\mathbf{j+}_1$. Substitute $\mathsf{m}(\lambda, x)$ for v_i everywhere in $\mathcal{D}'$. By Lemma 45 this does not affect the already established derivabilty results since CS is schematic.

In the following we provide some examples of realization.

Example 51. We realize the following theorem of E in JE:

$$\Box A \to (\Box B \to \Box A).$$

Consider the derivation in E:

$$\dfrac{\dfrac{\dfrac{\dfrac{A \supset A \qquad A \supset A}{\Box A \supset \Box A}\ (\mathbf{RE})}{\Box A, \Box B \supset \Box A}\ (w \supset)}{\Box A \supset \Box B \to \Box A}\ (\supset\to)}{\supset \Box A \to (\Box B \to \Box A)}\ (\supset\to)$$

Let λ be a proof term such that $\lambda : (A \to A)$ is provable. We find the following realization. Note that (**je**) denotes several reasoning steps.

$$\dfrac{\dfrac{\dfrac{\dfrac{A \supset A \qquad A \supset A}{[\mathsf{e}(\lambda)]A \supset [\mathsf{e}(\lambda)]A}\,(\mathbf{je})}{[\mathsf{e}(\lambda)]A, [\mathsf{e}(\xi_0)]B \supset [\mathsf{e}(\lambda)]A}\,(w \supset)}{[\mathsf{e}(\lambda)]A \supset [\mathsf{e}(\xi_0)]B \to [\mathsf{e}(\lambda)]A}\,(\supset\to)}{\supset [\mathsf{e}(\lambda)]A \to ([\mathsf{e}(\xi_0)]B \to [\mathsf{e}(\lambda)]A)}\,(\supset\to)$$

Note that this is already a simplification. Following the realization procedure exactly as given in the proof, would yield

$$[\mathsf{e}(\lambda+\lambda)]A \to ([\mathsf{e}(\xi_0)]B \to [\mathsf{e}(\lambda+\lambda)]A).$$

Example 52. Realize $\Box\Box A \to \Box\Box A$ in JE. We find the following derivation in E:

$$\dfrac{\dfrac{A \supset A \qquad A \supset A}{\Box A \supset \Box A}\,(\mathbf{RE}) \qquad \dfrac{A \supset A \qquad A \supset A}{\Box A \supset \Box A}\,(\mathbf{RE})}{\Box\Box A \supset \Box\Box A}\,(\mathbf{RE})$$

We obtain the following realization, where again λ is a proof term with $\lambda : (A \to A)$ and κ is a proof term with $\kappa : ([\mathsf{e}(\lambda+\lambda)]A \to [\mathsf{e}(\lambda+\lambda)]A)$ being provable.

$$\dfrac{\dfrac{A \supset A \qquad A \supset A}{[\mathsf{e}(\lambda+\lambda)]A \supset [\mathsf{e}(\lambda+\lambda)]A} \qquad \dfrac{A \supset A \qquad A \supset A}{[\mathsf{e}(\lambda+\lambda)]A \supset [\mathsf{e}(\lambda+\lambda)]A}}{[\mathsf{e}(\kappa)][\mathsf{e}(\lambda+\lambda)]A \supset [\mathsf{e}(\kappa)][\mathsf{e}(\lambda+\lambda)]A}$$

Again, we used a simplification. The exact procedure would yield

$$[\mathsf{e}(\kappa+\kappa)][\mathsf{e}((\lambda+\lambda)+(\lambda+\lambda))]A \supset [\mathsf{e}(\kappa+\kappa)][\mathsf{e}((\lambda+\lambda)+(\lambda+\lambda)]A.$$

Example 53. Realize $\Box(A \to A) \to \Box(B \to B)$ by JE. We find the following derivation in E:

$$\dfrac{\dfrac{\dfrac{\dfrac{B \supset B}{A \to A, B \supset B}\,(w \supset)}{A \to A \supset B \to B}\,(\supset\to) \qquad \dfrac{\dfrac{A \supset A}{B \to B, A \supset A}\,(w \supset)}{B \to B \supset A \to A}\,(\supset\to)}{\Box(A \to A) \supset \Box(B \to B)}\,(\mathbf{RE})}{\supset \Box(A \to A) \to \Box(B \to B)}\,(\supset\to)$$

Let λ_1 and λ_2 be proof terms such that $\lambda_1 : ((A \to A) \to (B \to B))$ and $\lambda_2 : ((B \to B) \to (A \to A))$ are provable. We find the following realization:

$$\dfrac{\dfrac{\dfrac{\dfrac{B \supset B}{A \to A, B \supset B}\,(w\supset)}{A \to A \supset B \to B}\,(\supset\to) \qquad \dfrac{\dfrac{A \supset A}{B \to B, A \supset A}\,(w\supset)}{B \to B \supset A \to A}\,(\supset\to)}{[\mathsf{e}(\lambda_1 + \lambda_2)](A \to A) \supset [\mathsf{e}(\lambda_1 + \lambda_2)](B \to B)}\,(\mathbf{je})}{\supset [\mathsf{e}(\lambda_1 + \lambda_2)](A \to A) \to [\mathsf{e}(\lambda_1 + \lambda_2)](B \to B)}\,(\supset\to)$$

Example 54. We realize the axiom scheme $\mathbf{M} : \Box(A \wedge B) \to (\Box A \wedge \Box B)$ in JEM. We start with its derivation in EM:

$$\dfrac{\dfrac{\dfrac{\dfrac{A \supset A}{A \wedge B \supset A}\,(w\supset)}{\Box(A \wedge B) \supset \Box A}\,(\mathbf{RM}) \qquad \dfrac{\dfrac{B \supset B}{A \wedge B \supset B}\,(w\supset)}{\Box(A \wedge B) \supset \Box B}\,(\mathbf{RM})}{\Box(A \wedge B) \supset (\Box A \wedge \Box B)}\,(\to\wedge)}{\supset \Box(A \wedge B) \to (\Box A \wedge \Box B)}\,(\supset\to)$$

We find the following realization in JEM:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \supset A}{A \wedge B \supset A}\,(w\supset)}{A \wedge B \to A}\,(\supset\to)}{[x](A \wedge B) \supset [\mathsf{m}(\lambda, x)]A}\,(\mathbf{jm}) \qquad \dfrac{\dfrac{\dfrac{B \supset B}{A \wedge B \supset B}\,(w\supset)}{A \wedge B \to B}\,(\supset\to)}{[x](A \wedge B) \supset [\mathsf{m}(\kappa, x)]B}\,(\mathbf{jm})}{[x](A \wedge B) \supset ([\mathsf{m}(\lambda, x)]A \wedge [\mathsf{m}(\kappa, t)]B)}\,(\to\wedge)}{\supset [x](A \wedge B) \to ([\mathsf{m}(\lambda, x)]A \wedge [\mathsf{m}(\kappa, t)]B)}\,(\supset\to)$$

where λ, κ are proof terms with

$$\vdash_{\mathsf{JEM}} \lambda : (A \wedge B \to A) \quad \text{and} \quad \vdash_{\mathsf{JEM}} \kappa : (A \wedge B \to B)\ .$$

Example 55. Now we consider the formula $\Box A \vee \Box B \to \Box(A \vee B)$ with the following derivation:

$$\dfrac{\dfrac{\dfrac{\dfrac{A \supset A}{A \supset A, B}\,(\supset w)}{A \supset A \vee B}\,(\to\vee)}{\Box A \supset \Box(A \vee B)}\,(\mathbf{RM}) \qquad \dfrac{\dfrac{\dfrac{B \supset B}{B \supset A, B}\,(\supset w)}{B \supset A \vee B}\,(\to\vee)}{\Box B \supset \Box(A \vee B)}\,(\mathbf{RM})}{\Box A \vee \Box B \supset \Box(A \vee B)}\,(\vee\to)$$

We find the following realization tree:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \supset A}{A \supset A, B}(\supset w)}{A \supset A \vee B}(\to \vee)}{\supset A \to A \vee B}(\supset\to)}{[x]A \supset [v_1+v_2](A \vee B)}(\mathbf{jm}) \quad \dfrac{\dfrac{\dfrac{\dfrac{B \supset B}{B \supset B, A}(\supset w)}{B \supset A \vee B}(\to \vee)}{\supset B \to A \vee B}(\supset\to)}{[y]B \supset [v_1+v_2](A \vee B)}(\mathbf{jm})}{[x]A \vee [y]B \supset [v_1+v_2](A \vee B)}(\vee \to)
$$

Now we substitute the provisional variables v_1, v_2 by terms $v_1 = \mathsf{m}(\lambda, x)$ and $v_2 = \mathsf{m}(\kappa, y)$ where λ, κ are proof terms that

$$\vdash_{\mathsf{JEM}} \ \lambda : (A \to A \vee B) \text{ and } \vdash_{\mathsf{JEM}} \ \kappa : (B \to A \vee B) \ .$$

Hence we obtain

$$[x]A \vee [y]B \supset [\mathsf{m}(\lambda, x) + \mathsf{m}(\kappa, y)](A \vee B) \ .$$

Example 56. We realize formula $\Box(\Box A \wedge \Box B) \to (\Box\Box A \wedge \Box\Box B)$ in JEM. We start with the following derivation where we do not mention all rule applications.

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \supset A}{\Box A \supset \Box A}(\mathbf{RM})}{\Box A \wedge \Box B \supset \Box A}(w \supset)}{\Box(\Box A \wedge \Box B) \supset \Box\Box A}(\mathbf{RM}) \quad \dfrac{\dfrac{\dfrac{B \supset B}{\Box B \supset \Box B}(\mathbf{RM})}{\Box A \wedge \Box B \supset \Box B}(w \supset)}{\Box(\Box A \wedge \Box B) \supset \Box\Box B}(\mathbf{RM})}{\Box(\Box A \wedge \Box B) \supset \Box\Box A \wedge \Box\Box B}(\to \wedge)}{\supset \Box(\Box A \wedge \Box B) \to (\Box\Box A \wedge \Box\Box B)}(\supset\to)
$$

We find the following derivation for suitable proof terms $\lambda_1, \lambda_2, \gamma_1, \gamma_2$:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \supset A}{[x]A \supset [\mathsf{m}(\lambda_1, x)]A}(\mathbf{jm})}{[x]A \wedge [y]B \supset [\mathsf{m}(\lambda_1, x)]A}(w \supset)}{\begin{array}{c}[z]([x]A \wedge [y]B) \supset \\ [\mathsf{m}(\gamma_1, z)][\mathsf{m}(\lambda_1, x)]A\end{array}}(\mathbf{jm}) \quad \dfrac{\dfrac{\dfrac{B \supset B}{[y]B \supset [\mathsf{m}(\lambda_2, y)]B}(\mathbf{jm})}{[x]A \wedge [y]B \supset [\mathsf{m}(\lambda_2, y)]B}(w \supset)}{\begin{array}{c}[z]([x]A \wedge [y]B) \supset \\ [\mathsf{m}(\gamma_2, z)][\mathsf{m}(\lambda_2, y)]B\end{array}}(\mathbf{jm})}{[z]([x]A \wedge [y]B) \supset ([\mathsf{m}(\gamma_1, z)][\mathsf{m}(\lambda_1, x)]A \wedge [\mathsf{m}(\gamma_2, z)][\mathsf{m}(\lambda_2, y)]B)}(\to \wedge)}{\supset ([z]([x]A \wedge [y]B) \to ([\mathsf{m}(\gamma_1, z)][\mathsf{m}(\lambda_1, x)]A \wedge [\mathsf{m}(\gamma_2, z)][\mathsf{m}(\lambda_2, y)]B))}(\supset\to)
$$

Chapter 3

Dyadic Deontic Logic and its Justification Counterpart

3.1 Introduction

In this chapter we first regard Monadic Deontic Logic (MDL), which works with one modal operator. We raise up the well-known *Chisholm's puzzle* to indicate that MDL fails to overcome this puzzle while formulating the so called *Chisholm's set*. As a result, we turn to Dyadic Deontic Logic (DDL), where two modal operators are employed and due to having different deontic conditionals, the logic DDL overcomes the Chisholm's puzzle by formulating the Chisholm's set sentences by dyadic conditionals. The logic DDL uses *Preference models* in semantics, which we adopt in the justification version. Finally, we introduce an explicit version of DDL by using two types of terms, namely *proof terms* and *justification terms*.

3.2 Monadic Deontic Logic

Monadic Deontic Logic (MDL) or Standard Deontic Logic is in fact modal logic **KD** which works with one modal operator shown by $\bigcirc$ [32]. So $\bigcirc A$ is read as *A is obligatory*.

3.2.1 Syntax

The language of MDL is generated by the set of formulas which is inductively defined as follows:

$$F := P_i \mid \neg F \mid F \to F \mid \bigcirc F$$

where $P_i \in \mathsf{Prop}$ and $\bigcirc F$ is read as "F is obligatory". Axiom schemas of this system is as follows:

Axioms of classical propositional logic	CL
$\bigcirc(A \to B) \to (\bigcirc A \to \bigcirc B)$	($\bigcirc$-K)
$\bigcirc A \to \neg \bigcirc \neg A$	($\bigcirc$-D)

$$\frac{A \quad A \to B}{B} \ (\mathrm{MP}) \qquad\qquad \frac{A}{\bigcirc A} \ (\bigcirc\text{-Necessitation})$$

3.2.2 Semantics

For the semantics, MDL uses Kripke models $\mathcal{M} = (W, R, V)$, where R is serial.

Definition 57 (Relational model)**.** A relational model is a tuple $\mathcal{M} = (W, R, V)$, such that:

- W is a non-empty set of worlds;
- $R \subseteq W \times W$ is a binary relation on set W, such that R is serial, i.e.,

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

- V is a valuation function $V : \mathsf{Prop} \to \mathcal{P}(W)$ that assigns each atomic proposition p a subset of W, namely set of worlds in which p is true.

Not to mention that relation R is understood as a relation of deontic alternativeness. We read wRv as v is an ideal alternative to w, or v is a good successor of w. In fact w is good in the sense that it complies with all the obligations true in v.

Definition 58 (Truth under relational model)**.** Let $w, v \in W$. Truth of a formula $A \in \mathsf{Fm}$ is defined as follows:

- for the propositional cases it is in the standard way

- $\mathcal{M}, w \Vdash \bigcirc A$ iff for all $v \in W$ if wRv then $\mathcal{M}, v \Vdash A$.

This means $\bigcirc A$ is true at w in model $\mathcal{M}$ if A is true at all w's ideal alternatives.

In continue, we consider the deficacy of MDL to overcome the following puzzle.

3.2.3 Chisholm's Set

Chisholm [7] 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 [27, 28, 41]. Here is an example of Chisholm's set. Consider the following sentences:

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** [32].

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)$E \rightarrow \bigcirc R$	(2.2) $\bigcirc (E \rightarrow R)$	(3.2) $\bigcirc (E \rightarrow R)$	(4.2)$E \rightarrow \bigcirc R$
(1.3) $\bigcirc (\neg E \rightarrow \neg R)$	(2.3)$\neg E \rightarrow \bigcirc \neg R$	(3.3) $\bigcirc (\neg E \rightarrow \neg R)$	(4.3)$\neg E \rightarrow \bigcirc \neg R$
(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

$$P \to (\neg P \to Q) \tag{3.1}$$

is a propositional tautology. Using (3.1) 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 (3.1) 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 (3.1) that (4.4) implies (4.2).

3.3 Dyadic Deontic Logic

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 [8]. 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 [18]. In the following is shown how the *Chisholm's set* is solved by preference models.

3.3.1 Syntax

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: [34]

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

such that $P_i \in \mathsf{Prop}$, $\Box F$ is read as "F is setteled true" and $\bigcirc(F/G)$ as "F is obligatory, given G". $P(F/G)$ is a short form for $\neg \bigcirc (\neg F/G)$, read as "F is permissible, given G, and $\Diamond F$ 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.

3.3.2 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 AD, two types of modal operators are used: the alethic $\Box$-operator and dyadic deontic $\bigcirc$-operator.

System AD with the two operators $\Box$ and $\bigcirc$ is axiomatized as follows:

Axioms of classical propositional logic	CL
S5-scheme axioms for $\Box$	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)

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

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

- The axioms containing one operator $\Box$. These are axiom schemas of S5, namely K, T, and 5.
  - (K): $\Box(A \to B) \to (\Box A \to \Box B)$
  - (T): $\Box A \to A$
  - (5): $\Diamond A \to \Box \Diamond A$
- 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 from system AD:

$$
\begin{array}{ll}
\text{if } A \leftrightarrow B \text{ then } \bigcirc(C/A) \leftrightarrow \bigcirc(C/B) & \text{(LLE)} \\
\text{if } A \rightarrow B \text{ then } \bigcirc(A/C) \rightarrow \bigcirc(B/C) & \text{(RW)} \\
\bigcirc(B/A) \wedge \bigcirc(C/A) \rightarrow \bigcirc(B \wedge C/A) & \text{(AND)} \\
\bigcirc(C/A) \wedge \bigcirc(C/B) \rightarrow \bigcirc(C/A \vee B) & \text{(OR)} \\
\bigcirc(C/A) \wedge \bigcirc(D/B) \rightarrow \bigcirc(C \vee D/A \vee B) & \text{(OR')}
\end{array}
$$

3.3.3 Preference Models

Now we consider semantics of Dyadic Deontic Logic based on *preference models*.

Definition 59 (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 60 (Truth under preference model)**.** Given a preference model $\mathcal{M} = (W, \preceq, V)$ for $w, \in W$ and $A, B \in \mathsf{Fm}$, the truth for formulas under $\mathcal{M}$ is defined as follows:

- for the propositional casesit is in the 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 best$\|B\| \subseteq \|A\|$;

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

3.3.4 Chisholm's Set Revisited

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		$\bullet w_4,\ \ R$	

3.3.5 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]:

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

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

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

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":

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

3.4 Justification Version of Dyadic Deontic Logic

3.4.1 Justification Version of System AD

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

Definition 61. 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 62 (Axiom schemas of JAD)**.**

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
$\lambda : F \to !\lambda : \lambda : F$	$\mathsf{j4}$
$\neg\lambda : A \to ?\lambda : (\neg\lambda : A)$	$\mathsf{j5}$
$[t](B/A) \to \Delta : [t](B/A)$	(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 63 (Constant specification)**.** A *constant specification* CS is any subset:

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

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

Definition 64 (System $\mathsf{JAD_{CS}}$)**.** For a constant specification CS, the system $\mathsf{JAD_{CS}}$ is defined by a Hilbert-style system with the axioms of JAD and the following inference rules:

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

As usual in justification logic [1,4,21], $\mathsf{JAD_{CS}}$ internalizes its own notion of proof.

Lemma 65 (Internalization)**.** *Let CS be an axiomatically appropriate constant specification. For any formula A with $\mathsf{JAD_{CS}} \vdash A$, there exists a proof term λ such that $\mathsf{JAD_{CS}} \vdash \lambda : A$.*

To have a better understanding of the axiomatic system of JAD, 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 66. 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 $\mathsf{JAD_{CS}}$ 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 67. The explicit version of

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

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

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

Example 68. The explicit version of

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

strong factual detachment is derivable in $\mathsf{JAD_{CS}}$ 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
$[t](A/B) \to [\mathsf{e}(t, \gamma)](A/B \wedge \top)$	(Ext)
$[\mathsf{e}(t, \gamma)](A/B \wedge \top)$	(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)

3.4.2 Semantics

In this section we adopt *preference models* for system JAD. In this direction we keep the tradition of justification logic for the proof terms and add the preference model structure for justification terms. We first consider the following operations on the sets of formulas and sets of pairs of formulas in order to define basic evaluations.

Definition 69. 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:

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

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

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

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

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

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

Definition 70 (Basic evaluation)**.** A *basic evaluation* for $\mathsf{JAD}_{\mathsf{CS}}$ 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:

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

  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:

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

  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 71 (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 72 (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 73 (Basic model)**.** Given an arbitrary CS, a *basic model* for $\mathsf{JAD}_{\mathsf{CS}}$ is a basic evaluation that is *factive.*

3.5 Soundness and Completeness of $\mathsf{JAD}_{\mathsf{CS}}$

Theorem 74. *System* $\mathsf{JAD}_{\mathsf{CS}}$ *is sound with respect to the class of all basic models.*

Proof. The proof is by induction on the length of derivations in $\mathsf{JAD_{CS}}$. 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 $\mathsf{JAD_{CS}} \vdash F$ 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 75. *System* $\mathsf{JAD_{CS}}$ *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:

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

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 $\varepsilon^c \Vdash \lambda : F$, by 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. $\square$

3.6 Preference Models

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

Definition 76 (Quasi-model)**.** A quasi-model for $\mathsf{JAD_{CS}}$ 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. Two worlds w_1 and w_2 are *incomparable*, $w_1 \| w_2$, if $w_1 \npreceq w_2$ and $w_2 \npreceq w_1$. We say w_1 and w_2 are *equally good* if $w_1 \preceq w_2$ and $w_2 \preceq w_1$.

- ε is an evaluation function that assigns a basic evaluation ε_w to each world w.

Definition 77 (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 78. *As usual for quasi-models for justification logic [3,21,23], truth is local, i.e., for a quasi-model* $\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$ *and* $w \in W$*, we have for any* $F \in \mathsf{Fm}$*:*

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

Remark 79. *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 80 (Two notions of "best")**.** *There are two ways to formalize the notion of "best world" respecting optimality and maximality [35]:*

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

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

- $\mathrm{best}\|A\|$ *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 81 (Preference model)**.** A *preference model* is a quasi-model where ε_w is factive and satisfies the following condition:

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

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

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

Definition 82 (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 83. *$max_{\preceq}(\|A\|) = opt_{\preceq}(\|A\|)$ 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 $w \in opt_{\preceq}(\|A\|)$. □

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

Theorem 84. *System $\mathsf{JAD_{CS}}$ 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 $\mathsf{JAD} \vdash A$. We show that A is true in every world $w \in W$. By soundness of JAD 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 $\mathsf{JAD} \nvdash A$. By completeness of JAD 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 $\text{best}\|C\| \subseteq \|B\|$ since $\text{best}\|C\| = \emptyset$.

□

Remark 85. *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* $\text{best}\|C\| = \emptyset$.

However, by proving the following theorem we get desired results analogous to result in [36].

Theorem 86. *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 84 holds.

Corollary 87. *System* $\mathsf{JAD_{CS}}$ *is sound and complete with respect to preference models with a total betterness relation.*

By Lemma 83 this implies completeness of $\mathsf{JAD_{CS}}$ with respect to preference models under max rule.

Corollary 88. *System* $\mathsf{JAD_{CS}}$ *is sound and complete with respect to preference models under max rule.*

Definition 89 (Fully explanatory preference models)**.** A preference model is *fully explanatory* if the converse of statement (JYB) holds, it means for any world w and any formulas A, B:

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

For a completeness proof of system $\mathsf{JAD_{CS}}$ with respect to fully explanatory preference models, we need a detour through the selection function semantics.

Definition 90 (Selection function model)**.** A *selection function model* is a preference model where the relation $\preceq$ is replaced with a selection function $\mathfrak{f} : \mathsf{Fm} \to \wp(W)$, which assigns any formula $A \in \mathsf{Fm}$ a subset of $\|A\|$.
One can consider this subset to be the best A-worlds. So the selection function model is a triple $\mathcal{M} = \langle W, \mathfrak{f}, \varepsilon\rangle$ where $\mathfrak{f}$ meets the following conditions:

- if $\|A\| = \|B\|$ then $\mathfrak{f}(A) = \mathfrak{f}(B)$ (Syntax independence)
- $\mathfrak{f}(A) \subseteq \|A\|$ (Inclusion)
- $\mathfrak{f}(A) \cap \|B\| \subseteq \mathfrak{f}(A \wedge B)$ (Chernoff)

Thus the truth of formulas under a selection function model is defined in the same way as in preference models and (JYB) is as follows:

$$\mathcal{M}, w \Vdash [t](A/B) \text{ implies } \mathfrak{f}(B) \subseteq \|A\| .$$

A selection function model is called *fully explanatory* if the following holds:

$$\mathfrak{f}(B) \subseteq \|A\| \text{ implies } (A,B) \in \varepsilon(t) \text{ for some } t \in \mathsf{JTm}\ .$$

Theorem 91. *The system* $\mathsf{JAD}_{\mathsf{CS}}$ *is sound and complete with respect to the class of selection function models* $\mathcal{M} = \langle W, \mathfrak{f}, \varepsilon \rangle$*.*

Proof. The soundness proof is in the same way as in the theorem 25. The proof of completeness is in the same way as in preference models except for constructing $\mathcal{M} = \langle w_1, \mathfrak{f}, \varepsilon' \rangle$ such that $\varepsilon'_{w_1} := \varepsilon$ and

$$\mathfrak{f}(A) = \emptyset, \forall A \in \mathsf{Fm}$$

It's easy to check that $\mathfrak{f}$ has the properties: Syntax independence, Inclusion, and Chernoff. To show (JYB) let $A, B \in \mathsf{Fm}$ and for some $t \in \mathsf{JTm}$ $(A,B) \in \varepsilon(t)$. Obviously we have $\mathfrak{f}(B) \subseteq \|A\|$. □

Theorem 92. *For an axiomatically appropriate* CS*, the system* $\mathsf{JAD}_{\mathsf{CS}}$ *is sound and complete with respect to the class of fully explanatory selection function models.*

Proof. Soundness is obtained from soundness with respect to selection function models.

To show completeness, let $\mathsf{M}_{\mathsf{JAD}}$ be the set of all maximal JAD-consistent sets of formulas, and for $\omega \in \mathsf{M}_{\mathsf{JAD}}$, let:

$$\omega/G := \{F \mid [t](F/G) \in \omega, \text{ for some } t \in \mathsf{JTm}\}.$$

$$\omega/\Box := \{F \mid \lambda : F \in \omega \text{ for some } \lambda \in \mathsf{PTm}\}.$$

Now define the canonical model generated by ω

$$\mathcal{M}^\omega = \langle W^\omega, \mathfrak{f}^\omega, \varepsilon^\omega \rangle$$

as follows:

- $W^\omega := \{\Delta \in \mathsf{M}_{\mathsf{JAD}} \mid \Delta \supseteq \omega/\Box\}$;
- $\mathfrak{f}^\omega : \mathsf{Fm} \to \mathcal{P}(\mathsf{M}_{\mathsf{JAD}})$ such that:

  $$\mathfrak{f}^\omega(G) := \{\Delta \in \mathsf{M}_{\mathsf{JAD}} \mid \Delta \supseteq \omega/G\};$$

  for each $\Delta \in W^\omega$ we have:

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

Accordingly, the *truth set* of F in canonical model induced by $\omega \in \mathsf{M_{JAD}}$ is defined as follows:

$$\|A\|^{\mathcal{M}^\omega} = \{\Delta \in W^\omega \mid \mathcal{M}^\omega, \Delta \Vdash A\}.$$

We now show the truth lemma.

Theorem 93 (Truth lemma)**.** *Let ω be a fixed element of $\mathsf{M_{JAD}}$, and $\mathcal{M}^\omega$ be the canonical model generated by ω. For any $\Delta \in W^\omega$ and $A \in \mathsf{Fm}$ we have:*

$$\mathcal{M}^\omega, \Delta \Vdash A \textit{ iff } A \in \Delta$$

Proof. The proof is by induction on structure of A. In case A is atomic proposition, one can immediately see the result from the definition of canonical model. The Boolean cases are handled in the standard way. We just consider the cases with alethic and deontic operators. Let $A := \lambda : B$, then $\mathcal{M}^\omega, \Delta \Vdash \lambda : B$ iff $B \in \varepsilon^\omega_\Delta(\lambda)$ iff $A \in \Delta$. For deontic operator, suppose $A := [t](B/C)$, then $\mathcal{M}^\omega, \Delta \Vdash [t](B/C)$ iff $(B, C) \in \varepsilon^\omega_\Delta(t)$ iff $A \in \Delta$.

□

In the following we show that $\mathcal{M}^\omega$ is indeed a selection function model for every $\omega \in \mathsf{M_{JAD}}$, which means to check the following properties:

- Syntax independence: if $\|A\|^{\mathcal{M}^\omega} = \|B\|^{\mathcal{M}^\omega}$ then $\mathfrak{f}^\omega(A) = \mathfrak{f}^\omega(B)$.
  To show this, we first show the following lemma:

  Lemma 94. *Let $\omega \in \mathsf{M_{JAD}}$ be fixed. For any $\Delta \in W^\omega$ and arbitrary $F, G \in \mathsf{Fm}$ we have:*

  $$\vdash_{\mathsf{JAD_{CS}}} (F \leftrightarrow G) \textit{ iff } (F \in \Delta \Leftrightarrow G \in \Delta)$$

  Proof. Suppose $\vdash_{\mathsf{JAD_{CS}}} (F \leftrightarrow G)$. The result immediately follows from closure of maximal consistent sets under modus ponens. For the other direction, suppose $(F \in \Delta \Leftrightarrow G \in \Delta)$, for any $\Delta \in W^\omega$ but not $\vdash_{\mathsf{JAD_{CS}}} (F \to G)$. Then $\neg(F \to G)$ is consistent and by Lindenbaum's Lemma there is a maximal consistent set $\Gamma \ni \neg(F \to G)$. This means $F, \neg G \in \Gamma$. Since $F \in \Gamma$ and $\Gamma \in W^\Gamma$, we get $G \in \Gamma$, which contradicts $\neg G \in \Gamma$. □

Now to show syntax independence suppose $\|A\|^{\mathcal{M}^\omega} = \|B\|^{\mathcal{M}^\omega}$ for arbitrary $A, B \in \mathsf{Fm}$. This means $A \in \Delta$ iff $B \in \Delta$ for any $\Delta \in W^\omega$ by truth lemma. Hence by lemma above we get $\vdash_{\mathsf{JAD_{CS}}} (A \leftrightarrow B)$. Since CS is axiomatically appropriate, there is a $\lambda \in \mathsf{PTm}$, such that $\vdash_{\mathsf{JAD_{CS}}} \lambda : (A \leftrightarrow B)$. Thus $\{\lambda : (A \leftrightarrow B)\}$ is consistent and there is a maximal set $\Gamma \in \mathsf{M_{JAD}}$ we get $\lambda : (A \leftrightarrow B) \in \Gamma$. By axiom (Ext) and maximal consistency of Γ, we have

$$([t](C/A) \to [\mathsf{e}(t, \lambda)](C/B)) \in \Gamma.$$

This means if $\Gamma \supseteq \omega/A$ then $\Gamma \supseteq \omega/B$. The converse also holds by replacing B with A. Thus we have $\Gamma \in \mathfrak{f}^\omega(A)$ iff $\Gamma \in \mathfrak{f}^\omega(B)$.

- Inclusion: $\mathfrak{f}^\omega(A) \subseteq \|A\|$.
  By axiom (Id), and maximal consistency of ω, for any arbitrary A we have $[i](A/A) \in \omega$ and thus $A \in \omega/A$. By definition of $\mathcal{M}^\omega$ we get $\Delta \in \mathfrak{f}^\omega(A)$ implies $A \in \Delta$ and as a result $\mathfrak{f}^\omega(A) \subseteq \|A\|$.

- Chernoff: $\mathfrak{f}^\omega(B) \cap \|A\| \subseteq \mathfrak{f}^\omega(A \wedge B)$.
  Suppose $\Delta \in (\mathfrak{f}^\omega(B) \cap \|A\|)$ for arbitrary formulas A, B. Then $\Delta \in \mathfrak{f}^\omega(B)$ and $\Delta \in \|A\|$. So $\omega/B \subseteq \Delta$ and $A \in \Delta$. We show that $\Delta \in \mathfrak{f}^\omega(A \wedge B)$ by knowing that:

$$\mathfrak{f}^\omega(A \wedge B) = \{\Delta \in \mathsf{M_{JAD}} : \omega/(A \wedge B) \subseteq \Delta\},$$

  where

$$\omega/(A \wedge B) = \{F : [t](F/A \wedge B) \in \omega\}.$$

  Now we show $\omega/(A \wedge B) \subseteq \Delta$. Suppose $F \in \omega/(A \wedge B)$, which means $[t](F/A \wedge B) \in \omega$. By axiom (Sh) and maximal consistency of ω, we have $[\nabla t](B \to F/A) \in \omega$ and by axiom (COK) and maximal consistency of ω, again we get: $([s](A/B) \to [\nabla t \cdot s](F/B)) \in \omega$. Since $A \in \Delta$ and CS is axiomatically appropriate, there is a $\lambda \in \mathsf{PTm}$, such that $\mathsf{JAD_{CS}} \vdash \lambda : A$ and by maximal consistency of Δ we have $(\lambda : A) \in \Delta$. Again by maximal consistency of Δ and by axiom (Nec), $[\mathsf{n}(\lambda)](A/B) \in \Delta$. Hence by maximal consistency of ω we get $[\nabla t \cdot \mathsf{n}(\lambda)](F/B)) \in \omega$ which means $F \in \Delta$.

To show (JYB), consider the following observation. Let $\omega \in \mathsf{M_{JAD}}$ fixed and

$$\Delta/\Box := \{F : \lambda : F \in \Delta \text{ for some } \lambda \in \mathsf{PTm}\}.$$

We define a relation $R \subseteq \mathsf{M_{JAD}} \times \mathsf{M_{JE}}$ such that :

$$\Delta R \Gamma \text{ iff } \Delta/\Box \subseteq \Gamma,$$

then by having $\mathsf{S5}$ schemata, we find out R is an equivalence relation on set $\mathsf{M_{JAD}}$. Since R is symmetric, we have:

$$\Delta/\Box \subseteq \Gamma \text{ implies } \Gamma/\Box \subseteq \Delta.$$

As a result $\omega \supseteq \Delta/\Box$ for any $\Delta \in W^\omega$. Now suppose $\mathcal{M}^\omega, \Delta \Vdash [t](A/B)$ for some $\Delta \in W^\omega$. By Truth lemma we have $[t](A/B) \in \Delta$ which means $\Delta : [t](A/B) \in \Delta$, and by the observation above we have $[t](A/B) \in \omega$. We need to show that $\mathfrak{f}^\omega(B) \subseteq \|A\|^{\mathcal{M}^\omega}$. Suppose $\Gamma \in \mathfrak{f}^\omega(B)$ for some $\Gamma \in W^\omega$. By definition of canonical model we have $\Gamma \supseteq \omega/B$ and since $A \in \omega/B$ we get $A \in \Gamma$ which means $\Gamma \in \|A\|$.

For fully explanatory property of $\mathcal{M}^\omega$, we show that:

$$\mathfrak{f}^\omega(B) \subseteq \|A\| \text{ implies } \mathcal{M}^\omega, \Delta \Vdash [t](A/B) \text{ for some } t \in \mathsf{JTm}.$$

Suppose $\mathfrak{f}^\omega(B) \subseteq \|A\|$ for some $A, B \in formulas$. It means for any $\Gamma \in \mathfrak{f}^\omega(B)$, we have $A \in \Gamma$. So:

$$\forall\ \Gamma \in W^\omega, \Gamma \supseteq \omega/B \text{ implies } A \in \Gamma$$

which says that A belongs to every maximal consistent extension of Γ/B. We first need to show the following lemma.

Lemma 95. *Let $B \in \Delta$ for $B \in \mathsf{Fm}$ and $\Delta \in W^\omega$. Then consider $\Gamma = \omega/B := \{C : [t](C/B) \in \omega$ for some $t \in \mathsf{JTm}\}$ we show that:*

- *Γ is consistent and can be extended to a maximal consistent set called Γ^+;*
- *$\Gamma^+ \in W^\omega$.*

Now by the Lindenbaum's lemma and compactness of our logic [33], A is derivable from finite set

$$\{C_1, ..., C_n | [t](C_i/B) \in \omega \text{ for some } t \in \mathsf{JTm}\}$$

$$\vdash_{\mathsf{JAD_{CS}}} (C_1 \wedge ... \wedge C_n) \rightarrow A,$$

by using rule (AND), there is a term $s \in \mathsf{JTm}$ such that

$$[s](C_1 \wedge ... \wedge C_n/B) \in \omega$$

and by rule (RW) we get

$$[t](A/B) \in \omega, \text{ for some } t \in \mathsf{JTm}.$$

□

Theorem 96 (Completeness of $\mathsf{JAD_{CS}}$ w.r.t. fully explanatory selection function models). $\mathsf{JAD_{CS}}$ *is complete with respect to fully explanatory selection function models. For any formula $F \in \mathsf{Fm}$, we have*

$$\mathsf{JAD_{CS}} \vdash F \quad \textit{implies} \quad \mathcal{M} \Vdash F \textit{ for all selection function models } \mathcal{M}.$$

Proof. Suppose towards contradiction that $\mathsf{JAD_{CS}} \nvdash F$. By propositional reasoning we have $\mathsf{JAD_{CS}} \nvdash \neg F \rightarrow \bot$. It means that $\{\neg F\}$ is $\mathsf{JAD_{CS}}$-consistent and hence contained in a $\mathsf{JAD_{CS}}$ maximal consistent set ω. The canonical model induced by ω is a selection function model. By the truth lemma, we have

$$\mathcal{M}^\omega, \omega \Vdash \neg F,$$

since $\omega \in W^\omega$. In other words $\mathcal{M}^\omega \nVdash F$, and thus F is not valid by soundness of $\mathsf{JAD_{CS}}$ w.r.t. selection function models. □

The first step was to show the soundness and completeness of system $\mathsf{JAD_{CS}}$ with respect to the class of fully explanatory selection function models, and the second step is to show that the two semantics systems, selection function semantics and preference model semantics are equivalent. Half of this equivalence is easier and is established in the following theorem.

Definition 97 (Equivalence of class of models)**.** Two class of models $\mathcal{C}_1, \mathcal{C}_2$ for $\mathsf{JAD_{CS}}$ are called *logically equivalent* if for any $\mathcal{M}_1 \in \mathcal{C}_1$ there is an equivalent $\mathcal{M}_2 \in \mathcal{C}_2$ and vice versa, such that for all formula such $F \in \mathsf{Fm}$, we have $\mathcal{M}_1 \Vdash F$ if and only if $\mathcal{M}_2 \Vdash F$.

Theorem 98. *For any preference model $\mathcal{M} = \langle W, \preceq, \varepsilon \rangle$ under opt rule, there is an equivalent selection function model $\mathcal{M}' = \langle W, \mathfrak{f}, \varepsilon \rangle$, where W and ε are the same sets of worlds and $\mathfrak{f}$ meets syntax independence, inclusion and Chernoff. Moreover, if $\mathcal{M}$ is fully explanatory then $\mathcal{M}'$ is also fully explanatory.*

Proof. In case of opt rule, for any $\mathcal{M} = \langle W, \preceq, \varepsilon\rangle$, define $\mathcal{M}' = \langle W, \mathfrak{f}, \varepsilon\rangle$ by the following definition of $\mathfrak{f}$:

$$\forall A \in \mathsf{Fm}, \quad \mathfrak{f}(A) = opt_{\preceq}(\|A\|)$$

In case of max rule, the function is defined by:

$$\forall A \in \mathsf{Fm}, \quad \mathfrak{f}(A) = max_{\preceq}(\|A\|)$$

One can easily check the conditions, for any $A, B \in \mathsf{Fm}$:

- Syntax independence: If $\|A\| = \|B\|$ then $opt_{\preceq}(\|A\|) = opt_{\preceq}(\|B\|)$ and so $\mathfrak{f}(A) = \mathfrak{f}(B)$;
- Inclusion: $opt_{\preceq}(\|A\|) \subseteq \|A\|$ and so $\mathfrak{f}(A) \subseteq \|A\|$;
- Chernoff: $\mathfrak{f}(A) \cap \|B\| = opt_{\preceq}(\|A\|) \cap \|B\|$, on the other hand we have $\mathfrak{f}(A \wedge B) = opt_{\preceq}(\|A \wedge B\|) = opt_{\preceq}(\|A\| \cap \|B\|)$. Since $opt_{\preceq}(\|A\|) \cap \|B\| \subseteq opt_{\preceq}(\|A\| \cap \|B\|)$ we get $\mathfrak{f}(A) \cap \|B\| \subseteq \mathfrak{f}(A \wedge B)$.

□

To show the equivalence of two semantics systems, it is enough to show that for every selection function model, there exist a preference model.

Theorem 99. *For every selection function model $\mathcal{M} = \langle W, \mathfrak{f}, \varepsilon\rangle$ where $\mathfrak{f}$ meets syntax independence, inclusion and Chernoff, there is a preference model $\mathcal{M}' = \langle W', \preceq', \varepsilon'\rangle$ which is equivalent to $\mathcal{M}$ under opt rule.*

Proof. We start the proof with a remark providing some essential definitions.

Remark 100. *Let X be a collection of nonempty sets. A* choice function *is a function f defined on X, such that for every $A \in X$, $f(A)$ is an element of A. The axiom of choice states that for any set X of nonempty sets, there exists a choice function f defined on X, that maps each set of X to an element of that set. Formally it states:*

$$\emptyset \notin X \textit{ implies } \exists f : X \to \bigcup X, \textit{ s.t. } \forall A \in X (f(A) \in A).$$

Each choice function on a collection X of nonempty sets is an element of the cartesian product *of the sets in X. A family of sets is considered as any function g on arbitrary domain I that maps each $i \in I$ to a set X_i. We show* *the range of this function as* $\{X_i\}_{i\in I}$*. For a family of sets* $\{X_i\}_{i\in I}$*, a* choice function *on a family of sets* $\{X_i\}_{i\in I}$*, is a function:*

$$g : I \to \bigcup_{i\in I} X_i \quad \textit{such that} \quad \forall i \in I, g(i) \in X_i$$

which chooses an element from each set in X_i*. The set of all choice functions on the family of* $\{X_i\}_{i\in I}$ *is called the general cartesian product of* $\{X_i\}_{i\in I}$ *defined as follows:*

$$\Pi_{i\in I}\{X_i\} := \{g : I \to \bigcup_{i\in I} X_i \mid \forall i \in I, g(i) \in X_i\}$$

Note that if $I = \emptyset$ *then* g *is* $\emptyset$*.*

Let $\mathcal{M} = \langle W, \mathfrak{f}, \varepsilon\rangle$. For any $w \in W$ define:

$$\mathcal{Y}_w = \{\|C\| \subseteq W \mid w \in (\|C\| - \mathfrak{f}(C)) \text{ for some } C \in \mathsf{Fm}\} \quad \text{and} \quad F_w := \prod \mathcal{Y}_w$$

By assuming the axiom of choice, we define the following $\mathcal{M}' = \langle W', \preceq', \varepsilon'\rangle$:

- $W' = \{\langle w, g\rangle \mid w \in W, g \in F_w\}$
- $\langle w, g\rangle \preceq \langle v, g'\rangle$ iff $v \notin \text{Rng}(g)$
- $\varepsilon'(p) = \{\langle w, g\rangle : w \in \varepsilon(p)\}$
- $\varepsilon'_{\langle w,g\rangle}(\lambda) = \varepsilon_w(\lambda)$
- $\varepsilon'_{\langle w,g\rangle}(t) = \varepsilon_w(t)$

where $\text{Rng}(g)$ denotes the range of g. We first show that $W' \neq \emptyset$, i.e., for all $w \in W$ there exists a $g \in F_w$. Since $W \neq \emptyset$ there exists $w \in W$. In case $\mathcal{Y}_w = \emptyset$ then $F_w = \{\emptyset\}$. In case $\mathcal{Y}_w \neq \emptyset$, $\mathcal{Y}_w$ is a collection of non-empty sets and by axiom of choice there exists a function $g \in F_w$.

Now to show (JYB) in $\mathcal{M}'$, suppose $\mathcal{M}', \langle w, g\rangle \Vdash [t](A/B)$, so $(A, B) \in \varepsilon'_{\langle w,g\rangle}$ and $(A, B) \in \varepsilon_w$. Applying (JYB) for $\mathcal{M}$ we have $\mathfrak{f}(B) \subseteq \|A\|$. By lemma 37, $v \in \mathfrak{f}(B)$ if and only if there exist $g' \in F_w$ such that $\langle v, g'\rangle \in opt_{\preceq}(\|B\|)$. Since $\mathcal{M}, v \Vdash A$ we get $\mathcal{M}', \langle v, g'\rangle \Vdash A$ and as a result $\langle v, g'\rangle \in \|A\|$ which means $opt_{\preceq}(\|B\|) \subseteq \|A\|$. $\square$

As a result we get the following corollary.

Corollary 101. *For every selection function model* $\mathcal{M} = \langle W, \mathfrak{f}, \varepsilon \rangle$ *where* $\mathfrak{f}$ *meets syntax independence, inclusion and Chernoff, there is a preference model* $\mathcal{M}' = \langle W', \preceq', \varepsilon' \rangle$ *which is equivalent to* $\mathcal{M}$ *under opt rule. Consequently the completeness of system* JAD *with respect to preference models results from the completeness of this system with respect to selection function models.*

Conclusion

In this thesis the connection between justification logic and deontic logic is established and explicit versions of systems which tackle some deontic puzzles are presented. 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.

In the second chapter, we have presented two new justification logics $\mathsf{JE}_{\mathsf{CS}}$ and $\mathsf{JEM}_{\mathsf{CS}}$ as explicit counterparts of the non-normal modal logics E and EM, respectively.

Having a justification analogue of the modal logic E is particularly important in deontic contexts since, according to Faroldi [10], deontic modalities are hyperintensional.

Note that $\mathsf{JE}_{\mathsf{CS}}$ is hyperintensional even if it includes the axiom of equivalence **je**. Assume $[\mathsf{e}(\lambda)]F$ and let G be equivalent to F. Then $[\mathsf{e}(\lambda)]G$ only holds if λ proves the equivalence of F and G. Thus, in general, for any λ with $[\mathsf{e}(\lambda)]F$ one can find a G such that G is equivalent to F but $[\mathsf{e}(\lambda)]G$ does not hold.

On a technical level, the main novelty in our work is the introduction of two types of terms. This facilitates the formulation of axiom **je**, which corresponds to the rule of equivalence. Having this principle as an axiom (and not as a rule) in justification logic is important to obtain Internalization (Lemma 15).

We have established soundness and completeness of logics $\mathsf{JE}_{\mathsf{CS}}$ and $\mathsf{JEM}_{\mathsf{CS}}$ with respect to basic models, modular models and fully explanatory modular models.

We have shown that for an axiomatically appropriate and schematic constant specification CS, the justification logics $\mathsf{JE}_{\mathsf{CS}}$ and $\mathsf{JEM}_{\mathsf{CS}}$ realize the modal logics E and EM, respectively. The realization proof for $\mathsf{JEM}_{\mathsf{CS}}$ is standard, whereas the realization proof for $\mathsf{JE}_{\mathsf{CS}}$ required some new ideas since the rule (**RE**) does not respect polarities of $\Box$-occurences.

In the third chapter, we consider dyadic deontic logic that aims to formulate deontic conditionals. We introduced an explicit version $\mathsf{JAD_{CS}}$ of the alethic-deontic system AD, which features dyadic modalities to capture deontic conditionals. Semantics for AD is given in terms of preference models, where the set of worlds is ordered according to a betterness relation. The language of $\mathsf{JAD_{CS}}$ includes proof terms for the alethic modality and justification terms for the deontic modality.

In the sense of using two types of terms, we used the same idea of second chapter, however non-normal logics presented in the second chapter are very weak with respect to deductive power. To overcome this deficiency, the goal of second chapter was an attempt towards stronger systems.

We established soundness and completeness of $\mathsf{JAD_{CS}}$ with respect to basic models and preference models, and fully explanatory preference models.

Bibliography

[1] Sergei Artemov. Explicit provability and constructive semantics. *BSL*, 7(1):1–36, March 2001.

[2] Sergei Artemov. Justified common knowledge. *TCS*, 357(1–3):4–22, 2006.

[3] Sergei Artemov. The ontology of justifications in the logical setting. *Studia Logica*, 100(1–2):17–30, April 2012.

[4] Sergei Artemov and Melvin Fitting. *Justification Logic: Reasoning with Reasons*. Cambridge Tracts in Mathematics. Cambridge University Press, 2019.

[5] Jeremy Bentham. *Deontology ; Together with a Table of the Springs of Action ; and the Article on Utilitarianism*. New York: Oxford University Press, 1983.

[6] Samuel Bucheli, Roman Kuznets, and Thomas Studer. Justifications for common knowledge. *Applied Non-Classical Logics*, 21(1):35–60, 2011.

[7] Roderick M. Chisholm. Contrary-To-Duty Imperatives and Deontic Logic. *Analysis*, 24(2):33–36, 12 1963.

[8] Sven Danielsson. *Preference and obligation. : Studies in the logic of ethics.* Filosofiska föreningen, 1968.

[9] Federico Faroldi, Meghdad Ghari, Eveline Lehmann, and Thomas Studer. Impossible and conflicting obligations in justification logic. In A. Marra, F. Liu, P. Portner, and F. Van De Putte, editors, *Proceedings of DEON 2020*, 2020.

[10] Federico L. G. Faroldi. Deontic modals and hyperintensionality. *Logic Journal of the IGPL*, 27:387–410, 2019.

[11] Federico L. G. Faroldi. *Hyperintensionality and Normativity*. Springer, 2019.

[12] Federico L. G. Faroldi and Tudor Protopopescu. All-things-considered ought via reasons in justification logic, 2018. Preprint.

[13] Federico L. G. Faroldi and Tudor Protopopescu. A hyperintensional logical framework for deontic reasons. *Logic Journal of the IGPL*, 27:411–433, 2019.

[14] Melvin Fitting. The logic of proofs, semantically. *APAL*, 132(1):1–25, 2005.

[15] James William Forrester. Gentle murder, or the adverbial samaritan. *Journal of Philosophy*, 81(4):193–197, 1984.

[16] Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leon van der Torre. *Handbook of Deontic Logic and Normative System, Volume 2*. College Publications, 2021.

[17] William H. Hanson. Semantics for deontic logic. *Logique Et Analyse*, 8:177–190, 1965.

[18] Bengt Hansson. An analysis of some deontic logics. *Noûs*, 3(4):373–398, 1969.

[19] Andrzej Indrzejczak. Admissibility of cut in congruent modal logics. *Logic and Logical Philosophy*, 20(3):189–203, 2011.

[20] Ioannis Kokkinis, Petar Maksimović, Zoran Ognjanović, and Thomas Studer. First steps towards probabilistic justification logic. *Logic Journal of IGPL*, 23(4):662–687, 2015.

[21] R. Kuznets and T. Studer. *Logics of Proofs and Justifications*. Studies in logic. College Publications, 2019.

[22] Roman Kuznets, Sonia Marin, and Lutz Straßburger. Justification logic for constructive modal logic. *Journal of Applied Logic*, 8:2313–2332, 2021.

[23] Roman Kuznets and Thomas Studer. Justifications, ontology, and conservativity. In Thomas Bolander, Torben Braüner, Silvio Ghilardi, and Lawrence Moss, editors, *Advances in Modal Logic, Volume 9*, pages 437–458. College Publications, 2012.

[24] Roman Kuznets and Thomas Studer. Weak arithmetical interpretations for the logic of proofs. *Logic Journal of the IGPL*, 24(3):424–440, 2016.

[25] René Lavendhomme and Thierry Lucas. Sequent calculi and decision procedures for weak modal systems. *Studia Logica: An International Journal for Symbolic Logic*, 66(1):121–145, 2000.

[26] Eveline Lehmann and Thomas Studer. Subset models for justification logic. In R. Iemhoff, M. Moortgat, and R. de Queiroz, editors, *Logic, Language, Information, and Computation - WoLLIC 2019*, pages 433–449. Springer, 2019.

[27] David K. Lewis. *Counterfactuals*. Cambridge, MA, USA: Blackwell, 1973.

[28] David K. Lewis. Semantic analyses for dyadic deontic logic. In Sören Stenlund, editor, *Logical Theory and Semantic Analysis: Essays Dedicated to Stig Kanger on His Fiftieth Birthday*, pages 1–14. Reidel, 1974.

[29] Ernst Mally. *Grundgesetze des Sollens*, pages 227–324. Springer Netherlands, Dordrecht, 1971.

[30] E. Pacuit. A note on some explicit modal logics. 2005.

[31] Eric Pacuit. *Neighborhood Semantics for Modal Logic*. Cham, Switzerland: Springer, 2017.

[32] X. Parent and L. van der Torre. *Introduction to Deontic Logic and Normative Systems*. Texts in logic and reasoning. College Publications, 2018.

[33] Xavier Parent. On the strong completeness of åqvist's dyadic deontic logic g. In *International Workshop on Deontic Logic in Computer Science*, 2008.

[34] Xavier Parent. A complete axiom set for Hansson's deontic logic DSDL2. *Logic Journal of the IGPL*, 18(3):422–429, 2010.

[35] Xavier Parent. Maximality vs. optimality in dyadic deontic logic. *J. Philos. Log.*, 43(6):1101–1128, 2014.

[36] Xavier Parent. Completeness of Åqvist's systems E and F. *Rev. Symb. Log.*, 8(1):164–177, 2015.

[37] Bryan Renne. *Dynamic Epistemic Logic with Justification*. PhD thesis, City University of New York, May 2008.

[38] Alf Ross. Imperatives and logic. *Philosophy of Science*, 11(1):30–46, 1944.

[39] Thomas Studer. Decidability for some justification logics with negative introspection. *JSL*, 78(2):388–402, June 2013.

[40] Thomas Studer. A conflict tolerant logic of explicit evidence. *Logical Investigations*, 27:124–144, 05 2021.

[41] Bas C. van Fraassen. The logic of conditional obligation. *Journal of Philosophical Logic*, 1(3/4):417–438, 1972.

[42] Chao Xu, Yanjing Wang, and Thomas Studer. A logic of knowing why. *Synthese*, 198:1259–1285, 2021.