Diplomarbeit

Extensionalität für Verpflichtungen in Åqvists System F

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Technische Mathematik

eingereicht von

Dominik Pichler
Matrikelnummer 01325451

ausgeführt am Institut für Logic and Computation (TU Wien)
eingereicht an der Fakultät für Mathematik und Geoinformation der
Technischen Universität Wien

Betreuung
Betreuerin: Univ.Prof. Dr. Agata Ciabattoni
Mitwirkung: Dr. Xavier Parent

Wien, 20.10.2022

(Unterschrift Verfasser) (Unterschrift Betreuerin)

Master’s Thesis

Extensionality of obligations in Åqvist’s system F

for attainment of the academic degree of

Diplom-Ingenieur

as part of the study

Technical Mathematics

submitted by

Dominik Pichler
Student ID 01325451

written at the institute of Logic and Computation (TU Wien)
submitted to the Faculty of Mathematics and Geoinformation of TU Wien

Supervision
Main Advisor: Univ.Prof. Dr. Agata Ciabattoni
Co-Advisor: Dr. Xavier Parent

Vienna, 20.10.2022

(Signature author) (Signature main advisor)

Declaration

I hereby declare that this master's thesis is my original work. I have made no use of sources, materials or assistance other than those which had been openly and fully acknowledged in the text.

Vienna, 20.10.2022

Dominik Pichler

Acknowledgments

I would like to express my gratitude to all the people whose support let me reach this point in my life. First, I want to thank my thesis main advisor, Univ. Prof. Dr. Agata Ciabattoni, who has guided me through the last year with her invaluable patience and feedback. I also extend my deepest appreciation to my co-advisor, Dr. Xavier Parent, for the plentiful discussions and valuable comments throughout the last year, without them, this thesis would not have been possible. I am also very grateful to my coworkers at the TU Wien, especially my office mates, who significantly impacted and inspired me over these last few months. Finally, I would like to give a special thanks to my family and my partner Tanja. Without your support, I would have never gotten this far.

Abstract

The dyadic deontic logic system **F** is one of the best known preference-based deontic logics. It was introduced as a propositional logic, by Lennart Åqvist to offer a solution to contrary-to-duty paradoxes plaguing the field of deontic logic. This thesis addressed the challenging topic of extending the system **F** to first-order. We construct a first-order dyadic deontic logic system extending propositional **F**, which includes equality and definite descriptions in its language and an extensional dyadic deontic operator. To better understand equality, definite descriptions and extensional operators and why we want our system to include them, we investigate them in detail. We show which characterising properties a logic system has to fulfil to express those notions meaningfully and accurately. Furthermore, we demonstrate which problems must be circumvented when defining such a system. We provide two different first-order dyadic deontic logic systems extending the propositional system **F**. For each of them we introduce semantics, using Kripke models, and a Hilbert calculus. Moreover, we show that the Hilbert axiomatisations are sound in their respective semantics, that is if there is a derivation of a formula φ in the calculus from a set of premises Γ, then φ is a semantical consequence of Γ.

Keywords: Dyadic deontic logic, First-order modal logic, Definite descriptions, Extensionality, Åqvist's system **F**

Contents

1 Introduction

1.1 Problem Statement

Deontic logic is the branch of logic concerned with obligation, permission and related normative notions. Besides its theoretical interest, deontic logic finds applications in many fields ranging from law to artificial intelligence. The best-known propositional deontic logic is standard deontic logic SDL introduced by von Wright in [19], which set the groundwork for later deontic logics. SDL extends classical propositional logic with the operators $\bigcirc(.)$ and $P(.)$ to express that something ought to be and that something is permitted, respectively. For example, the SDL formula

$$PA \supset \bigcirc B$$

is read as "If A is permitted, then it ought to be that B". SDL is also called monadic deontic logic since its deontic operators are one-place.

SDL suffers, however, from many so-called paradoxes. On the one hand, there are derivable formulas which are counterintuitive from a common-sense reading. On the other hand, some formulas are underivable in SDL even though they should be derivable. Furthermore, SDL cannot deal with contrary-to-duty (CTD) reasoning, which concerns norms that prescribe what to do in cases of violation or sub-ideality. Contrary-to-duty reasoning is essential to our moral thinking, as in the statement, "If you are guilty, you should confess". Chisholm's paradox is one of the most famous paradoxes regarding the derivability of counterintuitive formulas and CTD reasoning, see [2]. This paradox deals with four logically independent statements which cannot be formalised in SDL without being either logically dependent or contradictory to each other. Those four statements are:

- It ought to be that Jones goes to the assistance of his neighbours.
- It ought to be that if Jones does go, he tells them he is coming.
- If Jones does not go, then he ought not to tell them he is coming.
- Jones does not go.

To solve this paradox, various deontic logics have been introduced, see e.g. the two volumes of the handbook [6] and [7]. Considering that obligations are often given as conditional statements, as "If A happens, then there is the obligation to do B", a binary obligation operator $\bigcirc(./.)$ was introduced by von Wright in 1956 to express under which condition an obligation has to be met, see [20]. This so-called dyadic deontic operator, expressing a conditional obligation $\bigcirc(B/A)$, read as: "B is obligatory, given A", gave rise to the field of Dyadic Deontic Logic, or DDL for short. The idea for the dyadic deontic operator was to introduce a counterpart to the material implication $\supset$ of classical logic used in SDL, which does not share certain properties suited for a deontic context. For example, in contrast to a material implication, the dyadic deontic formula $\bigcirc(B/A)$, together with the formula A, should in general, not imply the formula $\bigcirc B$. Furthermore, the so-called strengthening of the antecedent "$\bigcirc(B/A)$ implies $\bigcirc(B/A \wedge C)$" should also not apply. A natural language example of why the latter should not apply would be: "The obligation to cook dinner under the condition that you have guests over" does not imply "The obligation to cook dinner under the condition that you have guests over and your kitchen is on fire".

Out of the field of economics, rational choice theory was developed, see [12], which is based on the idea that individuals use their self-interests to make choices that will provide them with the most significant benefit. This gave rise to preference-based systems for DDL, analysing deontic modalities using possible world semantics, see [11] and [3]. A preference relation ranks the possible worlds in terms of betterness or comparative goodness, where one world gets ranked better than another iff it violates fewer obligations than the other. In those systems, a conditional obligation $\bigcirc(B/A)$ is true when the so-called best worlds in which A is true are worlds in which B is true as well.

The preference-based semantics for dyadic deontic logic were formulated into a modal logic by Åqvist [1] and Lewis [14]. Åqvist's modal setting also contains a modal operator $\square(.)$ to express necessitation. For example, the formula

$$\square B \supset \square \bigcirc (B/A)$$

can be read as "The necessity of B implies the necessity of the obligation for B under the condition A." One of the landmark systems in DDL is Åqvist's system **F** introduced in [1]. System **F** can adequately address contrary-to-duty obligations. For example, the four statements of Chisholm's paradox can be formulated without being logically dependent on each other or arriving at a contradiction. The solution to this paradox, written in the syntax of the system **F** is

$$\bigcirc(g/\top),\ \bigcirc(t/g),\ \bigcirc(\neg t/\neg g) \quad \text{and} \quad \neg g,$$

where g represents that Jones goes to his neighbour's assistance, t represents that Jones tells his neighbour that he is coming, and $\top$ stands for a true statement. The paradox is avoided by writing the antecedents of the second and third statements of Chisholm's paradox as the condition of the conditional obligation operator instead of antecedents of a material implication.

Another way to solve certain deontic paradoxes is to extend a propositional deontic logic system to the first-order level. As in classical first-order logic, in first-order deontic logic, quantifiers $\forall$ and $\exists$ are used to quantify over variables of the language to talk about the existing objects in the logic, for example, to express sentences of the form "One has an obligation to help under the condition that there exists a person that needs help." The paradox of gentle murder, introduced in [5] by Forrester, is concerned with the following three logically independent and non-contradictory statements:

- It is obligatory that Smith not murders Jones.
- It is obligatory that, if Smith murders Jones, Smith murders Jones gently.
- Smith murders Jones.

Formalising those three statements in propositional SDL results in the obligation for Smith to murder Jones. This is not only an undesired outcome in itself but also in direct contradiction to the first statement. Sinnott-Armstrong demonstrated one possible solution for this paradox in the 1985 article [17]. He shows that by adding quantifiers to SDL, we can formalise the three statements of Forrester's paradox without creating a contradictory or counterintuitive statement. The idea behind the solution is that the second statement can be written with the help of quantifiers as

$$\exists x M(x, j, s) \supset \exists x(M(x, j, s) \wedge \bigcirc G(x)),$$

where $M(x, j, s)$ represents that x is an act of murder by Smith of Jones and $G(x)$ represents that x is done gently. This paradox of propositional deontic logic, which cannot be dealt with propositional SDL, shows us how introducing quantifiers to deontic logic also leads to the solution of certain paradoxes.

When adding quantifiers and terms to a propositional logic system, one can introduce an equality symbol $(.) = (.)$ to build formulas describing when two terms are equal. Furthermore, definite descriptions, which are phrases describing a unique act, object or person, can be added to a logic system to describe terms via a unique property. Syntactically, a definite description is a term built from a formula. This is done in the form of $\iota x \varphi(x)$

and is read as "the unique person/object x for which $\varphi(x)$ is true". It is reasonable to add equality and definite descriptions to a deontic logic since, more often than not, if we talk about another person, we describe this person in the form of a unique feature instead of naming them by their name, for example, "The author of The Lord of the Rings" or "My next door neighbour". Equality and definite descriptions are, at first glance, straightforward concepts. We all seem to know what it means for two objects in our everyday life to be equal or what it means to describe an object through a unique feature. The concepts become more difficult to describe if we talk about abstract formulas and if the logic containing them allows a form of Extensionality, which is a natural requirement for deontic logic.

The principle of Extensionality is dealt with in many aspects of mathematics, logic and even our everyday life. For someone familiar with set theory, the first thing that comes to mind if they hear the word Extensionality is probably the axiom of Extensionality.

$$\forall x \forall y(x = y) \leftrightarrow \forall z(z \in x \leftrightarrow z \in y)$$

This axiom declares that two sets are equal if and only if they contain exactly the same elements. In other words, a set is determined uniquely by its members. This represents the main idea behind Extensionality in mathematics: two objects are considered equal if they have the same external properties, even though they might be defined differently. While there are many uses of Extensionality in mathematics, not just in set theory, one could argue that this all comes down to the axiom of Extensionality since set theory serves as its base. For example the functions $f(x) = x^2 + 10x + 25$ and $g(x) = (x+5)^2$, even though they are defined differently, are extensionally equal [1], because f and g always produce the same value given the same input.

The basic idea of Extensionality in logic is that if we have two terms that are seen as equal, we can replace one term with the other in a formula without changing the truth value of that formula. When dealing with equality, it makes sense for two equal terms not to change the validity of certain formulas if we replace one term with the other. The point made by Lou Goble in [10] is that it is natural to consider the deontic operator extensional. Since in a deontic logic, our terms usually range over people, we would like for two formulas containing obligations to be equivalent if they just differ by the use of different terms for the same person. There are arguments to be made that the conditional obligation operator of DDL can be considered extensional as well since obligations and conditions should not depend on how or by which name we describe a person, an object or an act. When we talk about an obligation using a definite description, we mean the obligation for the person or object we are describing, not for the description itself. The

[1] if f and g are defined on the same domain

idea of looking at people or objects independently of their description is something we want when talking about obligations. For example, if a judge gives the statement, "The accused had the obligation to help, given that he was present at the time", then the judge refers with "the accused" to the specific person in the courtroom, not to any other person in a different scenario.

This thesis aims to introduce a first-order dyadic deontic logic system, containing equality and definite descriptions, extending the propositional system **F**, in which the conditional obligation operator is extensional in both arguments. This raises the following hindrances to overcome. Firstly, Åqvist's system **F** is only a propositional dyadic deontic logic system. Therefore we have to extend **F** and its semantics into a first-order dyadic deontic logic system without changing the underlying system **F**.
Another problem lies in the semantics given to **F**. The propositional preference-based semantics used for Åqvist's system **F** evaluate formulas contained inside the scope of the conditional obligations operator at the best worlds in which the condition is true. Those worlds can be different from the world we started in. Hence even though obligations could refer to a world different from the current world, we nevertheless want them to refer to the same person or object we are discussing at the current world. We want to keep the interpretation of our variables inside of an obligation bound to the world we consider the obligation in. When we use definite descriptions to describe an object or a person who has the duty to fulfil an obligation in a possible world semantics, we have to be careful for our definite description to not "lose" the person or object we are talking about. This challenge is going to be a big part of this thesis.
The third problem lies on the syntactic level. This problem was discussed for SDL by Lou Goble in [8]. Goble states that one has to be careful when dealing with extensional deontic operators in a first-order deontic logic system containing equality and definite description. Indeed, without adding restrictions on the axiomatisation of the propositional rules of SDL, it results in the collapse of the deontic operator. This means that every formula A is equivalent to $\bigcirc A$, which renders the deontic operator irrelevant.

This brings us to the **main research question** of this thesis:

> Is it possible to create a first-order dyadic deontic logic system extending the propositional system **F**, with an extensional deontic operator $\bigcirc$, while avoiding its collapse?

Considering that **F** also contains the modal operator $\Box$, whose Extensionality can be put into question, the thesis **second research question** arises:

> Is it possible to create a first-order dyadic deontic logic system extending the propositional system **F**, with an extensional deontic operator $\bigcirc$ and a non-extensional modal operator $\Box$?

1.2 Thesis Results and Structure

In order to answer both question, we build two first-order dyadic deontic logic systems extending propositional **F** in which the conditional obligation operator is extensional in both arguments. To answer the first research question, we will build a system as described in this question, called $\mathbf{F}_1^{\forall}$. This system will extend system **F** with one minor change. Afterwards, we tackle the second research question by building a second system, called $\mathbf{F}_2^{\forall}$, with the help of the already established system. This system will extend system **F** with more changes than the first system.
This thesis is mainly based on the paper [9] by Lou Goble that introduces a first-order monadic deontic logic system containing equality, definite description and an extensional obligation operator while still avoiding its collapse. The move to a first-order dyadic setting offers additional challenges, which we discuss in detail in Chapter 2 of this thesis. This thesis consists of six chapters and is structured as follows.

Chapter 2 introduces Åqvist's system **F** and its semantics and explains why this thesis focuses on this particular system. Afterwards, we explore the axioms we want to add to our first-order extension that let us consider every operator in this system extensional and the axioms that characterise definite descriptions. We also give a formal proof showing that adding the axioms defining Extensionality and definite descriptions to a first-order version of the system **F** results in the collapse of the deontic and modal operator similarly to what has been shown by Lou Goble in [10]. Afterwards, we discuss further problems that can arise when defining a first-order dyadic deontic logic system containing an extensional deontic operator and how to circumvent them.

Chapter 3 defines the language, syntax and semantics of our first system, called $\mathbf{F}_1^{\forall}$. The semantics are based on the propositional preference-based semantics of Åqvist's system **F** and let us define which formulas are valid in $\mathbf{F}_1^{\forall}$. Then we show the properties that the semantics of $\mathbf{F}_1^{\forall}$ fulfil.

Chapter 4 defines the notion of a formal proof and a provable formula for $\mathbf{F}_1^{\forall}$. This is done in the form of a first-order Hilbert axiomatisation, called $\mathbf{HF}_1^{\forall}$, based on the Hilbert axiomatisation of system **F**, but also including the axioms described in Chapter 2. We then present the connection between $\mathbf{HF}_1^{\forall}$ and $\mathbf{F}_1^{\forall}$ in the form of a soundness proof, which shows that every provable formula in the Hilbert axiomatisation is valid in $\mathbf{F}_1^{\forall}$. This shows us that $\mathbf{F}_1^{\forall}$ is fully extensional, which means that every operator of this system fulfils the Extensionality axioms defined in Chapter 2.

Chapter 5 uses the syntax and adjusts the semantics and the Hilbert axiomatisation of the previous chapters to introduce a second system, called $\mathbf{F}_2^{\forall}$, and a second Hilbert axiomatisation, called $\mathbf{HF}_2^{\forall}$. In this system, the modal operator $\Box$ cannot be considered extensional anymore, making it a non-extensional operator while maintaining its obligation operator's Extensionality. In this chapter we give a soundness proof of the second established Hilbert axiomatisation in accordance with their semantics and also discuss the similarities and main differences between $\mathbf{F}_1^{\forall}$ and $\mathbf{F}_2^{\forall}$ and between $\mathbf{HF}_1^{\forall}$ and $\mathbf{HF}_2^{\forall}$.

Chapter 6 ends this thesis with a summary of the findings. We also give directions for future work in the form of further research questions.

2 Preliminaries

This chapter aims to introduce Åqvist's system **F**, to explain how it is defined and why this thesis focuses on **F**. We also discuss first-order dyadic deontic logic and what axioms a first-order dyadic deontic logic has to fulfil for its operators to be considered extensional and how to add definite descriptions. In this chapter, we also revisit the problem of the deontic collapse described by Goble, but in this case, for the system **F**. We describe new issues that arise when we extend the system **F** into first-order with an extensional dyadic deontic operator and how this will affect the definition of the semantics built in this thesis.

2.1 Hilbert Systems and Semantics

A Hilbert-style deduction system, named after David Hilbert and also called a Hilbert system or Hilbert axiomatisation, consists of a set of logical axioms and a set of inference rules. Logical axioms are formulas obtained through certain formula schemas, where a formula schema describes an infinite set of formulas of a fixed pattern. For example, every formula that can be derived by substituting φ in the formula schema $\neg(\varphi \wedge \neg\varphi)$ with a formula of the syntax of the system **F** is an axiom in **F**. A rule of inference describes which formulas can be derived given one or more formulas. One example of a rule of inference contained in most Hilbert systems is the rule of modus ponens which allows us to derive a formula φ from the already derived formulas $\psi \supset \varphi$ and ψ. Every formula that can be derived by starting with one or more finitely many axioms and using finitely many rules of inference is called provable (in symbols: $\vdash \varphi$) in that respective Hilbert system, and the finite sequence of axioms and rules building this derivation is called a formal proof.

Formal semantics, also called the semantics of a logic, are used to give formulas an interpretation and determine their meaning. The semantics of a logic can be defined in many different ways. Still, in general, they are algebraic structures which assign a certain truth value to a formula to define if this formula is valid (in symbols: $\models \varphi$).

Optimally, even though the semantics and the Hilbert axiomatisation are independently defined, the set of all provable formulas is equal to the set of all valid formulas. In other words, one has to show that for every formula φ, the notions $\vdash \varphi$ and $\models \varphi$ are equivalent.

Showing for every formula φ that $\vdash \varphi$ implies $\models \varphi$ is called a soundness proof and shows that every formula derivable in the Hilbert axiomatisation is valid in the semantics. Proving for every formula φ that $\models \varphi$ implies $\vdash \varphi$, is called a completeness proof and shows that every valid formula is provable in the Hilbert system.

2.2 Åqvist's System F

Åqvist introduced three different propositional dyadic deontic logic (DDL) systems called **E**, **F** and **G** respectively. The three systems are built on the following syntax:

Definition 2.2.1. *The language $\mathscr{L}$ used for Åqvist's systems consists of the following:*

- *A countable set of propositional variables* $\mathscr{P} := \{p, q, r, \ldots\}$
- *Two logical connectives* $\wedge, \neg$
- *A binary obligation operator* $\bigcirc(./.)$
- *A unary modal operator* $\Box$

A string of symbols of the language $\mathscr{L}$ is called a formula if they are arranged in a certain finite order. The formula definition is given inductively.

Definition 2.2.2. *A finite string of symbols φ is called a formula if $\varphi \in \mathscr{P}$ or if it is of the form $\psi \wedge \chi$, $\neg\psi$, $\bigcirc(\psi/\chi)$, or $\Box\psi$, given two already established formulas ψ and χ.*

Definition 2.2.3. *The symbols $\vee, \supset, \leftrightarrow, \Diamond$ and P are defined the following way:*
Let φ and ψ be formulas in the above-defined language, then

- $\varphi \vee \psi := \neg(\neg\varphi \wedge \neg\psi)$
- $\varphi \supset \psi := \neg(\varphi \wedge \neg\psi)$
- $\varphi \leftrightarrow \psi := (\varphi \supset \psi) \wedge (\psi \supset \varphi)$
- $\Diamond\varphi := \neg\Box\neg\varphi$, *read as "$\varphi$ is possible"*
- $P(\varphi/\psi) := \neg\bigcirc(\neg\varphi/\psi)$, *read as "$\varphi$ is permitted given ψ"*

The systems **E**, **F** and **G** contain the modal logic **S5** as a sublogic, see [13]. **S5** is characterised by the axioms $\Box(\varphi \supset \psi)$, $\Box\varphi \supset \varphi$ and $\Diamond\varphi \supset \Box\Diamond\varphi$. The Hilbert axiomatisation given to the system **E** is:

Axioms:

All truth-functional tautologies	(**PL**)
S5-schemata for $\Box$ and $\Diamond$	(**S5**)
$\bigcirc(\varphi \supset \chi/\psi) \supset (\bigcirc(\varphi/\psi) \supset \bigcirc(\chi/\psi))$	(**COK**)
$\bigcirc(\varphi/\psi) \supset \Box\bigcirc(\varphi/\psi)$	(**Abs**)
$\Box\varphi \supset \bigcirc(\varphi/\psi)$	(**Nec**)
$\Box(\varphi \leftrightarrow \psi) \supset (\bigcirc(\chi/\varphi) \leftrightarrow \bigcirc(\chi/\psi))$	(**Ext**)
$\bigcirc(\varphi/\varphi)$	(**Id**)
$\bigcirc(\varphi/\psi \wedge \chi) \supset \bigcirc(\chi \supset \varphi/\psi)$	(**Sh**)

Rules:

If $\vdash \varphi$ and $\vdash \varphi \supset \chi$ then $\vdash \chi$	(**MP**)
If $\vdash \varphi$ then $\vdash \Box\varphi$	(**N**)

The Hilbert axiomatisation given to the system **F** is the same as for system **E**, including the extra axiom:

$$\Diamond\psi \supset (\bigcirc(\varphi/\psi) \supset P(\varphi/\psi)) \quad (\textbf{D*})$$

The Hilbert axiomatisation given to the system **G** is the same as for system **F**, including the extra axiom:

$$(P(\psi/\varphi) \wedge \bigcirc(\psi \supset \chi/\varphi)) \supset \bigcirc(\chi/\varphi \wedge \psi) \quad (\textbf{SP})$$

The axioms of those systems represent statements whose validity is accepted in general. For example, the axiom **COK** says that the obligation for φ implying χ under the condition ψ implies that obligation for φ under the condition ψ implies the obligation for χ under the condition ψ. In other words, the obligation of an implication implies that the obligation of the antecedent implies the obligation of the consequent. The reason we focus specifically on the system **F** and not on the system **E**, which is a subsystem of system **F**, is that the system **F** contains the axiom **D***. It says that if ψ is possible, the obligation for φ under the condition ψ implies the permission for φ under the condition ψ. This axiom represents the dyadic deontic logic counterpart to the standard deontic logic axiom $\bigcirc\varphi \supset P\varphi$. This axiom is one of the leading causes of the deontic collapse in SDL, as shown in [8]. For similar reasons, we are focusing on system **F** rather than system **G**, which is an extension of system **F**. System **G** encompasses system **F** and adds the axiom **SP**, which states that if ψ is permitted and ψ implying χ is obligated, both under the condition φ then χ is obligated under the condition of φ and ψ. Although this axiom is interesting, it does not add anything new that would be relevant to the original problem proposed in Goble's paper.

Remark 2.2.4. *The formula $\Box\varphi \leftrightarrow \bigcirc(\bot/\neg\varphi)$ is provable for any formula φ.*

The semantics introduced for the systems **E**, **F** and **G** use preference models to define which formulas are valid. Those are structures of the following form:

Definition 2.2.5 (Preference Model)**.** *$M = \langle W, \succeq, \upsilon\rangle$ is called a preference model, where*

- *$W \neq \emptyset$ is a set of possible worlds*
- *$\succeq \subseteq W \times W$ is a binary relation on W called the betterness relation*
- *υ is a function which maps every propositional letter to a set of possible worlds.*

The definition for a formula φ being true at a world w in a model M (in symbols: $M, w \models \varphi$) is given through a double induction with the help of the so-called truth sets $||\varphi||^M$ and best truth sets $best(||\varphi||^M)$. Sets of the form $||\varphi||^M$ are sets of worlds representing all worlds of M in which φ is true. This means $M, w \models \varphi \Leftrightarrow w \in ||\varphi||^M$. The set $best(||\varphi||^M)$ contains the best worlds out of $||\varphi||^M$ according to the betterness relation $\succeq$. The best worlds can be defined in different, not always equivalent, ways.

Definition 2.2.6. *Let $M = \langle W, \succeq, \upsilon\rangle$ be a preference model and $W' \subseteq W$ be a set of worlds. Then we call*

$$opt(W') := \{w \in W' : \forall v \in W (v \in W' \Rightarrow v \succeq w)\}$$

the set of optimal worlds of W'.

The other common notion for best worlds in DDL is that of the maximal world:

Definition 2.2.7. *Let $M = \langle W, \succeq, \upsilon\rangle$ be a preference model and $W' \subseteq W$ be a set of worlds. Then we call*

$$max(W') := \{w \in W' : \forall v \in W ((v \in W' \wedge w \succeq v) \Rightarrow v \succeq w)\}$$

the set of maximal worlds of W'.

The definitions of optimality and maximality are taken from the paper [15]. In this thesis, we are focusing on optimal sets.

Definition 2.2.8 (Truth in **F**). *Given a preference model $M = \langle W, \succeq, \upsilon \rangle$, a formula φ and a world $w \in W$, we define what it means that φ is true in M at w:*

- *If $\varphi \in \mathscr{P}$ then $M, w \models \varphi$ iff $w \in \upsilon(\varphi)$*
- *If $\varphi = \neg\psi$, then $M, w \models \varphi$ iff $M, w \not\models \psi$*
- *If $\varphi = \psi \wedge \chi$, then $M, w \models \varphi$ iff $M, w \models \psi$ and $M, w \models \chi$*
- *If $\varphi = \Box\psi$, then $M, w \models \varphi$ iff $\forall v \in W : M, v \models \psi$*
- *If $\varphi = \bigcirc(\psi/\chi)$, then $M, w \models \varphi$ iff $best(||\chi||^M) \subseteq ||\psi||^M$*

A formula is called valid, if it is true at every world of every preference model.

The properties to which a class of models has to subscribe to such that it is sound and complete regarding one of Åqvist's systems varies from system to system. Some examples of properties used are *reflexivity*, *limitedness* and *totalness*:

Definition 2.2.9. *A preference model $M = \langle W, \succeq, \upsilon \rangle$ is called reflexive if*

$$\forall w \in W : w \succeq w.$$

holds. A preference model $M = \langle W, \succeq, \upsilon \rangle$ fulfils the limitedness property if for every formula φ

$$||\varphi||^M \neq \emptyset \Rightarrow best(||\varphi||^M) \neq \emptyset$$

holds. A preference model $M = \langle W, \succeq, \upsilon \rangle$ is called total if

$$\forall w, v \in W : w \succeq v \vee v \succeq w$$

holds.

The completeness of Åqvist's systems **E** and **F** with respect to their respective preference-based semantics is one of the oldest problems in deontic logic. In [15], Parent shows that the system **F** is sound and complete with respect to the class of preference models in which the betterness relation is limited, the class of those in which it is limited and reflexive, and the class of those in which it is limited and total. [1]

[1] for both definitions of best, respectively

Our first-order dyadic deontic logic's semantics are based on the class of preference models in which the betterness relation is limited and reflexive, and best worlds are defined as optimal worlds. [2] We now define the reflexive closure for an arbitrary binary relation to make our examples easier to read, since all models in this thesis will have the reflexivity property:

Definition 2.2.10. *Given a binary relation $B \subseteq X \times X$ we call $RC(B) := B \cup \{(z,z) : z \in X\}$ the reflexive closure of B.*

As mentioned in Chapter 1, the four statements of Chisholm's paradox can be expressed in the system **F** [3] without contradiction, as the formulas

$$\bigcirc(g/\top),\ \ \bigcirc(t/g),\ \ \bigcirc(\neg t/\neg g),\ \ \neg g,$$

where g represents that Jones goes to his neighbour's assistance, t represents that Jones tells his neighbour that he is coming, and $\top$ stands for a true statement. To check that those four formulas do not contradict each other, in other words, that they are all satisfiable simultaneously, we have to build a preference model in which all those four statements are true at the same world. Since this thesis focuses on the models in which the betterness relation is limited and reflexive, we will also give the following model those properties and call a world a best world if it is an optimal world.

Example 2.2.11. *Let be the preference model $M := \langle W, \succeq, \upsilon \rangle$ with*

$$\begin{aligned} W &:= \{w_1, w_2, w_3, w_4\} \\ \succeq &:= RC(\{(w_1,w_2),(w_1,w_3),(w_1,w_4),(w_2,w_4),(w_3,w_4)\}) \\ \upsilon(g) &:= \{w_2, w_4\} \\ \upsilon(t) &:= \{w_1, w_4\} \end{aligned}$$

In this model the four formulas $\bigcirc(g/\top)$, $\bigcirc(t/g)$, $\bigcirc(\neg t/\neg g)$ and $\neg g$ are true at the worlds w_1 and w_3, because $w_1 \notin \upsilon(g)$, $w_3 \notin \upsilon(g)$ and $best(||\top||) = \{w_4\} \subseteq \{w_2, w_4\} = ||g||$, $best(||g||) = \{w_4\} \subseteq \{w_1, w_4\} = ||t||$ and $best(||\neg g||) = \{w_3\} \subseteq \{w_2, w_3\} = ||\neg t||$.

In the figure below, an arrow pointed from a world w to a world v represents $w \succeq v$. Notice that one world is ranked better than another in this model if fewer obligations are violated. Violating an obligation in this context means that the condition of an obligation holds but

[2] If the same or similar results can be achieved with different properties and a different notion of best worlds is up to further research.

[3] also in the systems **E** and **G**

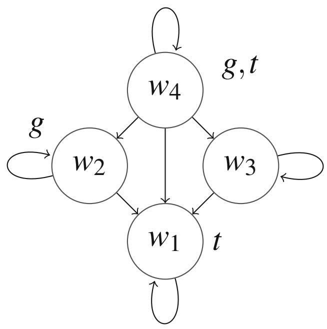

Figure 2.1: Solution to Chisholm's paradox

the obligation itself does not. This does not mean that the obligation is not valid in the model, since this is dependent on the evaluation of the formula in the best worlds where the condition holds true. For example, at world w_4, no given obligation is violated, but world w_2 violates the obligation "It ought to be that if Jones does go, he tells them he is coming".

2.3 First-Order DDL, Extensionality and Definite Descriptions

The solution to the Forrester paradox in first-order deontic logic, see p. 3, shows that the quantification over acts and people gives us much more freedom regarding the formalisation of obligations and permissions. In propositional deontic logic, an obligation stated as "Everyone is obliged to do the act A" and an obligation stated as "The act A should be done" cannot be formulated differently enough to highlight the distinction between those two statements. However, with the help of quantifiers, we could write the first statement as $\forall x \bigcirc A(x)$ and the second statement as $\bigcirc \exists x A(x)$.
In the case of dyadic deontic logic, adding quantifiers lets us be more precise with the obligation's meaning and the condition inside of the conditional obligation operator than in propositional DDL. Let $H(y)$ represent "y helps" and $A(x)$ represent "x is under attack":

1. $\forall x \forall y \bigcirc (H(y)/A(x))$
2. $\bigcirc(\forall x H(y)/\forall y A(x))$

The first formula purely focuses on the individuals alone. It could be used to formalise the statement:

"Every individual has an obligation to help under the condition that anyone is under attack."

The second one states an obligation for every person in an ideal world, seen as a collective, under the condition that something is happening to a person in an ideal world. It could be used to formalise the statement:

"We have the obligation to help under the condition that we are under attack."

When we talk about every individual, as in the above statement, we would like to imply a statement which talks about a specific existing person. For example, if "Tom exists" and the first statement holds true, then it should imply that "Everyone has an obligation to help under the condition Tom is attacked." holds true. On the other hand, even if "Tom exists" and the second statement hold true, we could argue that the statement "We have the obligation to help under the condition Tom is attacked" is generally not true. This example shows that in first-order dyadic deontic logic, we can emphasise the distinction between the obligation referring to every individual or to a collective, the same holds for the condition. [4]

When introducing Extensionality into our first-order logic we have to decide which axioms and/or rules our logic system has to fulfil to consider it extensional. One of the axiom schemas we will focus on in this thesis is *Universal Instantiation*:

$$\exists x(x=t) \supset (\forall x \varphi(x) \supset \varphi(t)) \qquad (\mathbf{UI})$$

It represents the statement: If there exists an object t, then if a formula φ is true for every existing object, the formula φ is true for the object t. We can view Universal Instantiation as the Extensionality of our logic regarding quantifiers. It lets us replace a variable inside a formula bound by a universal quantifier on the outside with an existing object. Let us give an example of the use of **UI** in a deontic context using the example given above:

"If Tom exists, then if every individual has an obligation to help under the condition that anyone is under attack", we can infer that "Every individual has an obligation to help under the condition that Tom is under attack".

This statement could be written as

$$\exists x(x=t) \supset (\forall x \forall y \bigcirc (H(y)/A(x)) \supset \forall y \bigcirc (H(y)/A(t))).$$

[4] One could also give different possible meanings to the formulas $\forall x \bigcirc (\forall y H(y)/A(x))$ and $\forall y \bigcirc (H(y)/\forall x A(x))$.

We should mention that outside the field of deontic logic, we would not like **UI** to be a valid formula schema, especially if the logic contains definite descriptions. For example, in the case of temporal logic [16], which is used to reason about time and temporal information, the validity of **UI** should not be accepted. Otherwise, given the two statements

- "The pope currently exists."
- "For every currently existing person, it will always be the case that they were born before 2025."

we could derive the statement: "It will always be the case that the pope was born before 2025." This statement is counterintuitive to the idea of the title of pope passing from person to person, and therefore at one point in time, or at least within the next 150 years, there is going to be a pope born after 2025.

When dealing with equality, it makes sense for two equal terms not to change the validity of a formula if we replace one term with the other. This brings us to the other formula schema we want to focus on, which we call the *Axiom of Replacement*:

$$t = s \supset (\varphi(t) \leftrightarrow \varphi(s)) \qquad \textbf{(I)}$$

This formula schema represents a form of replacement rule for equal objects. If t and s represent the same object, then the same formulas should hold if we use t instead of s and vice versa. In our everyday life, we often use the contraposition of this formula as a tool to argue why two things are not the same. For example, the sentence

"The key you are holding is not my house key since this one is silver and my house key is golden."

represents a statement that will generally be accepted without much protest. This is because if two objects do NOT share a certain property, in this case, colour, then they cannot be the same object or, in this case, the same key. In a deontic context, one could argue that **I** should hold without restrictions since obligations do not depend on how we describe a person. For example:

"If Tom is the person with the yellow shirt, then the obligation to help Tom is equivalent to the obligation to help the person with the yellow shirt."

UI and **I** are used by Goble in [9] to in his first-order logic system to characterise the Extensionality of his deontic operator. The validity of **UI** and **I** in the semantics of first-order dyadic deontic is not as straightforward as it seems. Problems arise if placing terms in a formula changes the meaning of those terms. This can happen if φ changes the world from which we originally started from. For example let w be a world at which the equality $t = s$ holds. The evaluation of the formula $\bigcirc\varphi(t)$ at w depends on the evaluation of $\varphi(t)$ at a best world v, which could be different from w. At this new world v the interpretation of the terms t and s must not be the same, hence we can not imply $\varphi(t) \leftrightarrow \varphi(s)$ at v. This would mean $t = s \not\supset \bigcirc\varphi(t) \leftrightarrow \bigcirc\varphi(s)$ at world w. Since this world change is the basic idea behind the interpretation of the modal operator $\Box$ and the obligation operator $\bigcirc$, the change can happen if the terms are contained in the scopes of such operators. Hence we want to craft the semantics for our first-order dyadic deontic logic system in such a way that we can talk about terms like definite descriptions in a deontic context without losing the meaning of those descriptions in the process of evaluating the formula in which the descriptions appear.[5]

One way of introducing definite description into a first-order system is by adding the following axiom schemas:

$$\forall y((\forall x(\varphi \leftrightarrow x = y)) \supset y = \imath x\varphi) \quad (\mathbf{D1})$$

$$\exists y(y = \imath x\varphi) \supset \exists! x\varphi \quad (\mathbf{D2})$$

D1 and **D2** are used by Thomason in [18] to syntactically characterise definite descriptions. **D1** states that "For every individual y if for every individual x the truth value of a formula φ is the same as for x being equal to y then y is the unique individual that can be characterised by the definite description $\imath x\varphi$. An exemplary statement for **D1** would be:

"If a person is the one and only who invented a certain object, we can call him or her "the inventor of that object".

D2 describes the other way around. It states that "If there exists an individual for which the definite description of $\imath x\varphi$ fits, he or she is the unique individual with the property of φ" a definitive statement for **D2** would be:

"If a person can be defined through "the inventor of a certain object", they are the only individual who invented that object."

[5]This idea is based on Goble's paper [9].

2.4 The Deontic Collapse

Goble has demonstrated that adding axioms and rules to a first-order monadic deontic logic, which let us consider the obligations to be extensional, and adding equality and definite descriptions to the logic, leads to the collapse of the deontic operator. This means that $\varphi \leftrightarrow \bigcirc\varphi$ is provable for every formula φ, which renders the deontic operator useless. A similar problem arises when we try to add Extensionality to a first-order dyadic deontic logic containing the axiom **D***. Using defining axioms for definite descriptions and wanting the Axiom of Replacement and Universal Instantiation to work in our first-order logic without placing restrictions on the axioms and rules of the propositional system **F**, we run into the problem that the operators $\square$ and $\bigcirc$ collapse into trivialities. Below we see a formal proof of this phenomenon. In short, the problem arises from the fact that adding definite descriptions to a system that is extensional to a certain point without restricting its original rules results in every formula implying its own obligation as well as its own necessity and vice versa.

We now give a formal proof for the collapse of the dyadic deontic operator $\bigcirc$ with axioms and rules of the system **F**, as well as the axioms **D1**, **D2**, **I** and **UI** and basic first-order rules for the quantifiers. We start by proving $\varphi \supset \square\varphi$ and $\varphi \supset \bigcirc(\varphi/\psi)$ where φ is an arbitrary formula which does not contain x as a free variable.

(a) $\varphi \supset \forall x[(x = t \wedge \varphi) \leftrightarrow x = t]$	(Tautology + Quantifier rules)
(b) $\forall y[\forall x((x = y \wedge \varphi) \leftrightarrow x = y) \supset y = \iota x(x = y \wedge \varphi))]$	(**D1**)
(c) $\exists y(y = t) \supset [\forall x[(x = t \wedge \varphi) \leftrightarrow x = t] \supset t = \iota x(x = t \wedge \varphi)]$	(b + **UI**)
(d) $\exists y(y = t) \supset [\varphi \supset t = \iota x(x = t \wedge \varphi)]$	(a + c)
(e) $t = \iota x(x = t \wedge \varphi) \supset [\square\exists y(y = t) \leftrightarrow \square\exists y(y = \iota x(x = t \wedge \varphi))]$	(**I**)
(f) $\exists y(y = t) \supset [\varphi \supset [\square\exists y(y = t) \leftrightarrow \square\exists y(y = \iota x(x = t \wedge \varphi))]]$	(d + e)
(g) $\square\exists y(y = t) \supset [\varphi \supset \square\exists y(y = \iota x(x = t \wedge \varphi))]$	(f + **S5**)
(h) $\exists y(y = \iota x(x = t \wedge \varphi)) \supset \exists! x(x = t \wedge \varphi)$	(**D2**)
(i) $\exists y(y = \iota x(x = t \wedge \varphi)) \supset \varphi$	(h + Quantifier rules)
(j) $\square\exists y(y = \iota x(x = t \wedge \varphi)) \supset \square\varphi$	(i + **N** + **S5**)
(k) $\square\exists y(y = t) \supset (\varphi \supset \square\varphi)$	(g + j)
(l) $\square\exists y(y = t)$	(Assumption)
(m) $\varphi \supset \square\varphi$	(l + k + **MP**)
(n) $\square\varphi \supset \bigcirc(\varphi/\psi)$	(**Nec**)
(o) $\varphi \supset \bigcirc(\varphi/\psi)$	(m + n)

We can see that assuming something simple like the necessary existence of a certain term

t will result in $\varphi \supset \Box\varphi$ and $\varphi \supset \bigcirc(\varphi/\psi)$ being provable for every formula φ and every formula ψ. Suppose we also assume that ψ is possible. In that case, we can even prove $\varphi \leftrightarrow \bigcirc(\varphi/\psi)$ which completely collapses the $\bigcirc$ operator and makes the obligation of a formula equivalent to the formula itself:

(a) $\neg\varphi \supset \bigcirc(\neg\varphi/\psi)$	(Proof above)
(b) $\neg\bigcirc(\neg\varphi/\psi) \supset \varphi$	(Contraposition of a)
(c) $P(\varphi/\psi) \supset \varphi$	(b + definition of P)
(d) $\Diamond\psi \supset (\bigcirc(\varphi/\psi) \supset P(\varphi/\psi))$	(**D***)
(e) $\Diamond\psi \supset (\bigcirc(\varphi/\psi) \supset \varphi)$	(c + d)
(f) $\Diamond\psi$	(Assumption)
(g) $\bigcirc(\varphi/\psi) \supset \varphi$	(e + f + **MP**)

One could argue that this problem arises from the Extensionality of the $\Box$ operator since we used the axioms **I** and **UI** on formulas containing the operator $\Box$. What if we only allow them on formulas not containing $\Box$? Do we still run into the same problem? The answer is yes, which we show in the following proof. We now assume that the axioms **I** and **UI** hold only for formulas in which the replaced variable or term is not contained inside the scope of a $\Box$ operator. We can still construct a proof for $\varphi \supset \bigcirc(\varphi/\psi)$, where ψ is any formula and $\bigcirc(\varphi/\psi) \supset \varphi$:

(a) $\exists y(y=t) \supset [\varphi \supset t = \iota x(x=t \wedge \varphi)]$	(Same as before)
(b) $\exists y(y = \iota x(x=t \wedge \varphi)) \supset \varphi$	(Same as before)
(c) $\Box[\exists y(y = \iota x(x=t \wedge \varphi)) \supset \varphi]$	(b + **N**)
(d) $\bigcirc([\exists y(y = \iota x(x=t \wedge \varphi)) \supset \varphi]/\psi)$	(c + **Nec** + **MP**)
(e) $\bigcirc([\exists y(y = \iota x(x=t \wedge \varphi))]/\psi) \supset \bigcirc(\varphi/\psi)$	(d + **COK**)
(f) $t = \iota x(x=t \wedge \varphi) \supset [\bigcirc(\exists y(y=t)/\psi) \leftrightarrow \bigcirc(\exists y(y = \iota x(x=t \wedge \varphi))/\psi)]$	(**I**)
(g) $\Box\exists y(y=t)$	(Assumption)
(h) $\bigcirc(\exists y(y=t)/\psi)$	(g + **Nec** + **MP**)
(i) $t = \iota x(x=t \wedge \varphi) \supset \bigcirc(\exists y(y = \iota x(x=t \wedge \varphi))/\psi)$	(f + h)
(j) $t = \iota x(x=t \wedge \varphi) \supset \bigcirc(\varphi/\psi)$	(e + i)
(k) $\exists y(y=t)$	(g + **S5**)
(l) $\varphi \supset t = \iota x(x=t \wedge \varphi)$	(a + k + **MP**)
(m) $\varphi \supset \bigcirc(\varphi/\psi)$	(j + l)

The proof for $\bigcirc(\varphi/\psi) \supset \varphi$ is the same as in the previous case.
In this thesis we circumvent the collapse of the operators by placing a restriction on the rule **N** as seen in Goble's paper [9]. This restriction blocks the steps from (i) to (j) in the first proof and (b) to (c) in the third proof.

2.5 Global and Local Operators

This section discusses a new problem that arises when dealing with an extensional operator. We also highlight how this problems leads us to consider two different interpretations for the operator $\Box$.

We start this section with the following (informal) definition: Given a possible world semantics, then we call a modal operator global if its truth value is not dependent on the world it is evaluated in, in other words, an operator ⌑ is called global if for any formula φ the formula ⌑φ being true at one world of a model implies ⌑φ being true at all worlds of that model. If ⌑ is not a global operator, we call it a local operator. In the preference-based semantics of system **F**, both the $\Box$ operator and the $\bigcirc$ operator are global operators. Given this definition, we now look at an example of why we have to be careful when we use the Axiom of Replacement on formulas that include definite descriptions contained in a global operator.

Example 2.5.1 (Switching seats)**.** *Let us consider the scenario depicted in figure 2.2. At the dinner table are five seats. Because the apartment is small, there is not enough space for the person in the corner to get up and walk away easily without the person to their right making some room. For reasons of etiquette, the person to the right of the person sitting in the corner has the obligation to get up under the condition that the person in the corner gets up. Now let us define the following formulas*

- $L(x,y)$ *represents "x is sitting to the left of y"*
- $G(x)$ *represents "x gets up"*
- $C(x)$ *represents "x sits in the corner".*

As a result, we can write the above-described obligation as the formula

$$\bigcirc(G(\imath y L(\imath x C(x), y))/G(\imath x C(x))).$$

At world w the equalities $x_1 = \imath x C(x)$ and $x_2 = \imath y L(\imath x C(x), y)$ hold if the variable x_1 is assigned to the person b and the variable x_2 is assigned to the person i, since b is the *person sitting in the corner and to the left of i. Hence at the world w if* x_1 *gets up* x_2 *has the obligation to get up too, this is represented in the formula* $\bigcirc(G(x_2)/G(x_1))$*. We would therefore like for the formula* $\bigcirc(G(x_2)/G(x_1))$ *to be true at w as well.*

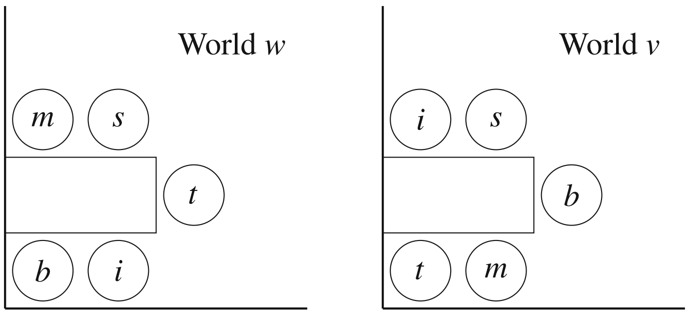

Figure 2.2: Switching seats

So far no new problem has been created. Still, this example is not only given to motivate why we should consider conditional obligations to be extensional but should also highlight another problem we encounter when we want our dyadic deontic operator to be extensional. Namely, we cannot expect an extensional operator to have a global interpretation. In general, our obligations will not be valid globally if they are valid in one world. By extending the example above, we can demonstrate how the Extensionality of our formulas and global interpretations of a conditional obligation do not mix well with definite descriptions. Let us consider the same scenario as before but with the persons sitting in a different order around the table, like at the world v. Hence the formula $\bigcirc(G(x_2)/G(x_1))$ should in general not be true at the world v even though $\bigcirc(G(\iota y L(\iota x C(x), y))/G(\iota x C(x)))$ is true at the world v. We can see that the obligation $\bigcirc(G(x_2)/G(x_1))$ is only locally true, even though $\bigcirc(G(\iota y L(\iota x C(x), y))/G(\iota x C(x)))$ holds at both worlds. This means that at the world w, x_2 has an obligation to get up under the condition that x_1 gets up, but this does not transcend to a world in which x_1 is not the person sitting to x_2's left. Since in the propositional Åqvist's system **F**, in contrast to SDL, the conditional obligation operator and the necessity operator are global operators, this is an entirely new aspect we have to consider, which is not considered in [9].

This example shows us that we can consider obligations extensional, but we must confine our interpretation to a local level! Even though we have an obligation for something in one world, this does not have to be the case in another. This should not discourage the Extensionality of the dyadic obligation operator, but it demonstrates that we must be careful if we want the operator to be extensional. For example, one part of how to circumvent this problem is to use variable assignments that are dependent on the worlds.

Given the observation in the example above, we should discuss the evaluation of a formula $G(t)$ at a world with the interpretation of the term t bound to another world. When we keep the terms of a formula bound to the world we are evaluating the formula in, we can derive certain formulas which, at first sight seem counterintuitive but do indeed make sense if we keep the local nature of our interpretation in mind.

Example 2.5.2 (Kind tyrant)**.** *Let us demonstrate what that means by considering the following three statements concerning a "kind tyrant":*

- *Every person has an obligation to be kind to others.*
- *Every person has the obligation not to be a tyrant.*
- *There exists a tyrant.*

Those three statements do not contradict each other and are logically independent. If we say the formula $T(x)$ represents "x is a tyrant", and the formula $K(x)$ represents "x is kind", we can formalise the three statements the following way:

- $\forall x \bigcirc (K(x)/\top)$
- $\forall x \bigcirc (\neg T(x)/\top)$
- $\exists y(y = \imath z T(z))$

*From the first and the third statement, we can conclude that: "The tyrant has an obligation to be kind". From the second and the third statement, we can imply the statement: "The tyrant has an obligation not to be a tyrant". We can write those two inferred statements as $\bigcirc(K(\imath z T(z))/\top)$ and $\bigcirc(\neg T(\imath z T(z))/\top)$ respectively. In propositional DDL, if a formula contains a non-conditional obligation, the semantics checks the interpretation of the formula, inside the scope of the obligation, in the best worlds of the model. The formula $\bigcirc(K(\imath z T(z))/\top)$ seems to state that in the best possible world, the tyrant is kind, but this interpretation is a little too unspecific. The statement we want to describe with this formula is revealed if we look at the tyrant independently of their title, we get the desired interpretation. This means the person we call a tyrant in our current world is obligated to be kind without worrying if they are a tyrant in the best world. This interpretation can be made clearer if we take a look at the formula $\bigcirc(\neg T(\imath z T(z))/\top)$, which, at first, seems like a contradictory statement. Although what $\bigcirc(\neg T(\imath z T(z))/\top)$ actually states is an obligation for our current tyrant not to be a tyrant. In other words, the person we call a tyrant is not a tyrant in the best possible world. Since the axiom **UI** is going to be valid in the semantics we are going to establish in this thesis the formula $\bigcirc(\neg T(\imath z T(z))/\top)$*

will be true at a world w if the formulas $\forall x \bigcirc(\neg T(x)/\top)$ *and* $\exists y(y = \imath z T(z))$ *are true at the world w.*[6] *The reason for this is the local interpretation of the definite description* $\imath z T(z)$ *in our semantics, which ties the interpretation to the world in the obligation gets evaluated in.*

We have now seen many arguments for the Extensionality of the conditional obligation operator $\bigcirc$. What about the Extensionality of the necessitation operator $\Box$? In this case, it depends on our interpretation of necessitation. Let us again say the formula $T(x)$ represents "x is a tyrant", and the formula $K(x)$ represents "x is kind" what do we mean by $\Box\neg K(\imath z T(z))$ being true at a world w? We could say that the unique tyrant from world w is necessarily not kind, which would mean that the tyrant from world w is not kind in any world. This would align with our local interpretation of our conditional obligation operator and make the $\Box$ operator extensional but not global. On the other hand, if we want necessitation to be stronger, we could interpret $\Box$ as a global operator. In that regard, we could say $\Box\neg K(\imath z T(z))$ is true at a world w means that the tyrant of every world is not kind in any world. The second interpretation would make the necessitation global but not extensional. Both interpretations seem to make sense in their own way. The main focus of this thesis is on the first interpretation, which aligns with the main research question proposed in Chapter 1. In Chapters 3 and 4, we will build the syntax, semantics and Hilbert axiomatisation of a first-order dyadic deontic logic system, in which every formula can be considered extensional, no matter if it contains a certain operator or not. In Chapter 5, we will consider the second interpretation of the $\Box$ operator and tackle the second research question proposed in the introduction.

[6]Translating the formulas into the first-order deontic logic system which Lou Goble has established in [9] then $(\forall x \bigcirc(\neg T(x)) \wedge \exists y(y = \imath z T(z)) \supset \bigcirc(\neg T(\imath z T(z)))$ does hold as well.

3 Syntax and Semantics

Here we introduce the syntax and semantics of the first-order extension of **F** which we call $\mathbf{F}_1^{\forall}$. At the end of this chapter, we revisit the paradoxes discussed in Chapter 2 and show that $\mathbf{F}_1^{\forall}$ entails their desired result.

3.1 Language and Formulas

The definitions in this thesis concerning first-order modal logic are based on the work done on first-order modal logic by M. Fitting, R. Mendelsohn, L. Goble and R. Thomason, see [4], [9] and [18]. We start by defining the language $\mathscr{L}$ for our first-order dyadic deontic logic containing identity and definite descriptions. It consists of two countable disjoint sets V and C, representing the variables and constants of our logic respectively, seven different logical symbols, and for each $n \in \mathbb{Z}^+$ a countable set of n-place predicate symbols which are pairwise disjoint and also disjoint from V and C.

Definition 3.1.1. *The language $\mathscr{L}$ contains:*

- *A countable and well-ordered set of variables* $V := \{x, y, z, \ldots\}$
- *A countable and well-ordered set of constants* $C := \{a, b, c, \ldots\}$
- *Two logical connectives* $\wedge, \neg$
- *Three first-order logic symbols* $\forall, \imath, =$
- *A binary obligation operator* $\bigcirc(./.)$
- *A unary modal operator* $\Box$
- *For each for $n \in \mathbb{Z}^+$ a countable set of n-place predicate symbols* $\mathbb{P} := \{A^n, B^n, \ldots\}$

We want to define all terms, formulas and the free and bound variables (for short, fv and bv, respectively) in those terms and formulas. The notions of free and bound will be helpful when we later define the replacement of one variable by another because we do not want to replace a variable bound by a quantifier or replace a free variable with a

variable which will be bound in the resulting formula.
Since we want our terms and formulas to be a finite sequence of symbols, in other words, we want them to be well-formed, we have to build them from the ground up. Therefore we are simultaneously defining the so-called depth (for short *de*) of our terms and formulas. The depth of a term or formula is a natural number which counts how many instances of definite descriptions have been used on top of each other to build the current term or formula. For example the term $\imath y(y = \imath x\varphi)$ has a depth of 2 if φ is a formula with the depth of 1. One natural language example of a definite description of depth 2 would be "The partner of the King of Norway". The definition of depth will help us later in this thesis when we want to use induction over our formula construction.
Starting with the base level of our syntax, we first define the so-called atomic terms and formulas.

Definition 3.1.2 (Atomic term). *If $t \in V \cup C$ is a symbol from the set of variables or the set of constants we call it an atomic term with $fv(t) := \{t\} \cap V$ and $bv(t) := \emptyset$.*

Definition 3.1.3 (Atomic formula). *If $A^n \in \mathbb{P}$ is a n-place predicate symbol and $t_1,...,t_n \in V \cup C$ are atomic terms then $A^n(t_1,...,t_n)$ is an atomic formula with $fv(A^n(t_1,...,t_n)) := \{t_1,...,t_n\} \cap V$ and $bv(A^n(t_1,...,t_n)) := \emptyset$.*

Using this base, we can now inductively define our well-formed terms and well-formed formulas used in our logic.

Definition 3.1.4 (Terms and Formulas).

- *Every atomic term is a term of depth* 0
- *Every atomic formula is a formula of depth* 0
- *If t_1 and t_2 are terms then $t_1 = t_2$ is a formula with* $de(t_1 = t_2) := \max\{de(t_1), de(t_2)\}$, $fv(t_1 = t_2) := fv(t_1) \cup fv(t_2)$ *and* $bv(t_1 = t_2) := bv(t_1) \cup bv(t_2)$
- *If φ is a formula then $\neg\varphi$ is a formula with* $de(\neg\varphi) := de(\varphi)$, $fv(\neg\varphi) := fv(\varphi)$ *and* $bv(\neg\varphi) := bv(\varphi)$
- *If φ is a formula then $\Box\varphi$ is a formula read as "the necessity for φ" with* $de(\Box\varphi) := de(\varphi)$, $fv(\Box\varphi) := fv(\varphi)$ *and* $bv(\Box\varphi) := bv(\varphi)$
- *If φ and ψ are formulas then $\varphi \wedge \psi$ is a formula with* $de(\varphi \wedge \psi) := \max\{de(\varphi), de(\psi)\}$, $fv(\varphi \wedge \psi) := fv(\varphi) \cup fv(\psi)$ *and* $bv(\varphi \wedge \psi) := bv(\varphi) \cup bv(\psi)$

- *If φ and ψ are formulas then $\bigcirc(\varphi/\psi)$ is a formula read as "the obligation for φ under the condition ψ" with* $de(\bigcirc(\varphi/\psi)) := \max\{de(\varphi), de(\psi)\}$, $fv(\bigcirc(\varphi/\psi)) := fv(\varphi) \cup fv(\psi)$ *and* $bv(\bigcirc(\varphi/\psi)) := bv(\varphi) \cup bv(\psi)$

- *If φ is a formula and $x \in V$ then $\forall x\varphi$ is a formula with* $de(\forall x\varphi) := de(\varphi)$, $fv(\forall x\varphi) := fv(\varphi) \setminus \{x\}$ *and* $bv(\forall x\varphi) := bv(\varphi) \cup \{x\}$

- *If φ is a formula and $x \in V$ then $\imath x\varphi$ is a term with* $de(\imath x\varphi) := de(\varphi) + 1$, $fv(\imath x\varphi) := fv(\varphi) \setminus \{x\}$ *and* $bv(\imath x\varphi) := bv(\varphi) \cup \{x\}$

- *If $R^n \in \mathbb{P}$ is a n-place predicate symbol and $t_1, \ldots, t_n$ are terms then $R^n(t_1, \ldots, t_n)$ is a formula with* $de(R^n(t_1, \ldots, t_n)) := \max\{t_1, \ldots, t_n\}$, $fv(R^n(t_1, \ldots, t_n)) := \bigcup_{i=1,\ldots,n} fv(t_i)$ *and* $bv(R^n(t_1, \ldots, t_n)) := \bigcup_{i=1,\ldots,n} bv(t_i)$

We call the set of all (well-formed) formulas WF.

Definition 3.1.5 (Derived connectives)**.** *The symbols $\vee, \bot, \top, \supset, \leftrightarrow, \exists, \exists!, \Diamond$ and P are defined the following way:*

Let φ and ψ be formulas in WF, t a term and $x \in V$ a variable:

- $\varphi \vee \psi := \neg(\neg\varphi \wedge \neg\psi)$
- $\bot := \neg\varphi \wedge \varphi$ *(for any φ)*
- $\top := \neg\bot$
- $\varphi \supset \psi := \neg(\varphi \wedge \neg\psi)$
- $\varphi \leftrightarrow \psi := (\varphi \supset \psi) \wedge (\psi \supset \varphi)$
- $\exists x\varphi := \neg\forall x\neg\varphi$
- $\exists! x\varphi := \exists y\forall x(x = y \leftrightarrow \varphi)$, *$y$ is the first element of V such that $y \notin fv(\varphi) \cup bv(\varphi)$*
- $\Diamond\varphi := \neg\Box\neg\varphi$
- $P(\varphi/\psi) := \neg\bigcirc(\neg\varphi/\psi)$
- $\bigcirc(\varphi) := \bigcirc(\varphi/\top)$
- $P(\varphi) := \neg\bigcirc(\neg\varphi)$

- $E(t) := \exists x(x = t)$, *x is the first element of V such that* $x \notin fv(t) \cup bv(t)$ [1]

Given two term t_1 *and* t_2 *we also define* $t_1 \neq t_2 := \neg(t_1 = t_2)$.

Next, we formally define formulas that are built by replacing a term t inside of a given formula φ with a term s. In the following definition, we have to be careful not to replace a term which contains a variable bound by a quantifier in φ, or replace it with a term which contains a variable which a quantifier binds in φ.

Definition 3.1.6. *Given a formula* φ *and two terms t and s with* $fv(t) \cap bv(\varphi) = \emptyset$ *and* $fv(s) \cap bv(\varphi) = \emptyset$ *we define* $\varphi_{t \hookrightarrow s}$ *as the formula in which zero up to all free* [2] *occurrences of t in* φ *have been replaced by s and* $\varphi_{t \Rightarrow s}$ *as the formula in which ALL free occurrences of t in* φ *have been replaced by s.*

Given Definition 3.1.4 of the formula and term construction, we can see that term replacement in Definition 3.1.6 distributes over the symbols $\neg, \wedge, \Box, \bigcirc$ and $\forall$ with respect to the terms which have been replaced.

Fact 3.1.7. *Given two formula* φ *and* ψ*, a variable* $x \in V$ *and two terms t and s with* $fv(t) \cap bv(\varphi) = \emptyset$, $fv(s) \cap bv(\varphi) = \emptyset$, $fv(t) \cap bv(\psi) = \emptyset$ *and* $fv(s) \cap bv(\psi) = \emptyset$ *then:*

- $(\neg\varphi)_{t \Rightarrow s} = \neg(\varphi_{t \Rightarrow s})$
- $(\varphi \wedge \psi)_{t \Rightarrow s} = \varphi_{t \Rightarrow s} \wedge \psi_{t \hookrightarrow s}$
- $(\Box\varphi)_{t \Rightarrow s} = \Box(\varphi_{t \Rightarrow s})$
- $\bigcirc(\varphi/\psi)_{t \Rightarrow s} = \bigcirc(\varphi_{t \Rightarrow s}/\psi_{t \Rightarrow s})$
- $(\forall x\varphi)_{t \Rightarrow s} = \forall x(\varphi_{t \Rightarrow s})$ [3]

Similarly if $\varphi_{t \hookrightarrow s}$ *is used instead of* $\varphi_{t \Rightarrow s}$.

Because of this fact, we can drop the brackets when talking about replacement. Using the definition of term replacement we can give the following definition in which all constants and definite description of a formula get replaced by new free variables. This definition will later be used to define a rule in our proof system.

Definition 3.1.8. *Given a formula* φ*, we define* φ^* *as the formula in which all terms* $t_1, \ldots, t_n$*, which are not variables and are occurring in the formula* φ*, have been replaced by* $x_1, \ldots, x_n \in V$ *respectively. The variables* $x_1, \ldots, x_n$ *are the first, pairwise different, elements of V such that* $x_1, \ldots, x_n \notin fv(\varphi) \cup bv(\varphi)$.

[1] This definition is taken from [18] and guarantees us that the new x does not bind an already existing variable in t by the quantifier $\exists$.

[2] Here free means that the term does not contain a variable bound by a quantifier.

[3] For this equality we also need $x \notin fv(t) \cup fv(s)$ to hold.

Example 3.1.9. *Let A, B and C be predicate symbols, $x, y, z \in V$ the first three variables of V, $c \in C$ a constant and $\varphi \in WF$ a well-formed formula:*

- $A(\imath y\varphi, c)^* = A(x,z)$
- $\forall xA(\imath yB(y,d),x)^* = \forall xA(z,x)$
- $A(\imath yB(\imath xC(x,y)),y)^* = A(z,y)$
- $A(y,y)^* = A(y,y)$

3.2 Frames, Models and Validity

In this section, we introduce the semantics for $\mathbf{F}_1^{\forall}$ the first-order extension of system **F** containing equality and definite descriptions, which uses the previously established syntax. The models our first-order logic, which evaluate the well-formed formulas, are based on the preference models of the propositional semantics of system **F**. We start by defining the frames on which our models are built on.

Definition 3.2.1 (Frame). $\mathscr{F} = \langle W, \succeq, D \rangle$ *is called a frame, where*

- $W \neq \emptyset$ *is a set of worlds*
- $\succeq \subseteq W \times W$ *is a binary relation on W called the betterness relation*
- D *is a function which maps every world $w \in W$ to a non-empty set D_w, D is called the domain function, and D_w is called the domain of the world w*

We say that a world w is at least as good as a world v iff $w \succeq v$.

Definition 3.2.2. *Given a frame $\mathscr{F} = \langle W, \succeq, D \rangle$, we call $\mathbb{D} := \bigcup_{w \in W} D_w$ the existing domain of $\mathscr{F}$ and $\mathbb{D}^+ := \mathbb{D} \cup \{\mathbb{D}\}$ the (whole) domain of $\mathscr{F}$.*

A frame is the underlying structure of a model. A single world $w \in W$ is only one possible world as part of the set of all worlds in $\mathscr{F}$. Different statements can be true at each world, also properties of certain objects or persons can change from world to world.
The worlds are ranked by the betterness relation, which does not yet have any properties assigned to it. Since "w is at least as good as a world v" is meant in the sense of obligations fulfilled, we have to choose the properties of $\succeq$ accordingly, but we need more definitions before we can define them properly.
Different objects or persons can exist at different worlds as indicated by the domain D_w of a world w. Including the element $\mathbb{D}$ in the whole domain $\mathbb{D}^+$ of a frame, gives us the

extra element $\mathbb{D}$, which is not contained in the domain of any world. [4] This element of the whole domain will be used to interpret definite descriptions that do not designate. Such a definite description is either a description of something that does not exist or is not uniquely defined through the description. A natural language example for the first one would be "The King of France". An example of the second one would be "The person with long hair" if this description is used at a metal concert.

Definition 3.2.3 (Model)**.** *$\mathscr{M} = \langle W, \succeq, D, I\rangle$ is called a model (on the frame $\mathscr{F} = \langle W, \succeq, D\rangle$), where I is a function (called interpretation function) such that:*

- *for $c \in C$ and $w \in W$: $I(c,w) \in \mathbb{D}^+$*
- *for $R^n \in \mathbb{P}$ and $w \in W$: $I(R^n,w) \subseteq (\mathbb{D}^+)^n$*

The intuitive understanding for, an element a of the domain $\mathbb{D}^+$ to be an element of $I(A,w)$, is that a has the property A at w.
Our formulas can contain free and bound variables, therefore, we need a way to define what happens to them in our models. Consequently we define variable assignments that assign each variable-world pair (x,w) to an element of the whole domain.

Definition 3.2.4 (Variable assignment)**.** *Given a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$ we call a function $g : V \times W \mapsto \mathbb{D}^+$ a variable assignment (of $\mathscr{M}$).* [5]

The assignment $g(x,w) = a$ can be understood as "everyone at the world w sees x as the element a". Notice that $g(x,w)$ does not have to be an element of the domain of w [6]. The individual domains are used to define all objects which are addressed by the $\forall$ quantifier at a world. To capture this notion we define *an x-variant of a variable assignment g at a world w* as the variable assignment h which replaces all assignments of a variable x, at every world, with an element of the domain of w.

Definition 3.2.5 (x-variant)**.** *Given a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g of $\mathscr{M}$, a variable $x \in V$ and a world $w \in W$ we call a variable assignment h an x-variant of g at w if for every $(y,v) \in (V \setminus \{x\}) \times W$ we have $g(y,v) = h(y,v)$ and for every $v, v' \in W$ we have $h(x,v) = h(x,v') \in D_w$.*

The following definition will be used for some proofs later in this thesis:

Definition 3.2.6. *Given a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g of $\mathscr{M}$ and an element of the whole domain $d \in \mathbb{D}^+$. We write $g_{x \Rightarrow d}$ for the variable assignment, which replaces every assignment of the variable x at any world with the element d:*

[4] $\mathbb{D} \notin \mathbb{D}$
[5] The output of a variable assignment is dependent on the worlds.
[6] The element a does not even have to be contained in the existing domain.

$$g_{x\Rightarrow d}(z,v) := \begin{cases} d & \text{if } (z,v)\in\{x\}\times W \\ g(z,v) & \text{otherwise} \end{cases}$$

Given a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g, a world $w \in W$ and a formula $\varphi \in WF$ we want to define what it means that φ is true in $\mathscr{M}$ at w under g. We call the set of all worlds v for which φ is true in $\mathscr{M}$ at w under g the *φ-worlds of $\mathscr{M}$ under g according to w* and write this set as $||\varphi||^{\mathscr{M}}_{g,w}$. An intuitive meaning for a world $v \in ||\varphi||^{\mathscr{M}}_{g,w}$ is that the formula φ holds true at v under g if looked at from the point of view of a person living at the world w. It is important to mention that $v \in ||\varphi||^{\mathscr{M}}_{g,w}$ does not convey a truth value for the formula φ per se, but it is used to define the truth value for φ per inductive definition. Alternatively one could write $\mathscr{M}, v \models^{w}_{g} \varphi$ instead of $v \in ||\varphi||^{\mathscr{M}}_{g,w}$ and define the sets afterwards accordingly, which is a more common way to do it. However, since we want to emphasise that $v \in ||\varphi||^{\mathscr{M}}_{g,w}$ does not define a truth value directly, we are sticking with the first one. Furthermore, for the proofs in this thesis, we will mostly work with the sets directly, making the proofs easier to read and explain.
Given a set $||\varphi||^{\mathscr{M}}_{g,w}$ we can define a set of worlds called the *best φ-worlds of $\mathscr{M}$ under g according to w*. A set $best(||\varphi||^{\mathscr{M}}_{g,w})$ contains all φ-worlds v of $\mathscr{M}$ under g according to w for which there is no "better" alternative, although there could be φ-worlds of $\mathscr{M}$ under g according to w to which v is as good as the world itself. Since a world in $best(||\varphi||^{\mathscr{M}}_{g,w})$ is therefore at least as good as every other φ-world v of $\mathscr{M}$ under g according to w in $\mathbf{F}^{\forall}_{1}$ we also call a world in $best(||\varphi||^{\mathscr{M}}_{g,w})$ an *optimal φ-world of $\mathscr{M}$ under g according to w*.

Definition 3.2.7. *Given a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$ and a set of the worlds $W' \subseteq W$ we call*

$$best(W') := \{v \in W' : \forall v' \in W(v' \in W' \Rightarrow v \succeq v')\}$$

the best W' worlds in $\mathscr{M}$.

We will now simultaneously define the above described sets $||\varphi||^{\mathscr{M}}_{g,w}$ and the following function $(I * g)_w$ recursively, such that for every term t, we get $(I * g)_w(t) \in \mathbb{D} \cup \{\mathbb{D}\}$. The motivation for $(I * g)_w(t) = p$ is that a person living at the world w interprets the term t as the object p.

Definition 3.2.8. *Let $\mathscr{M} = \langle W, \succeq, D, I\rangle$ be a model, g a variable assignment of $\mathscr{M}$, t a term, φ a well-formed formula and $w \in W$ a world:*

- *If $t = x \in V$ then $(I * g)_w(t) := g(x, w)$*
- *If $t = c \in C$ then $(I * g)_w(t) := I(c, w)$*

- *If $R^n \in \mathbb{P}$ and $t_1, ..., t_n$ are terms, then*

$$||R^n(t_1,...,t_n)||^{\mathscr{M}}_{g,w} := \{v \in W : \langle (I*g)_w(t_1), ..., (I*g)_w(t_n)\rangle \in I(R^n, v)\}$$

- *If t_1 and t_2 are terms, then*

$$||t_1 = t_2||^{\mathscr{M}}_{g,w} := \{v \in W : (I*g)_w(t_1) = (I*g)_w(t_2)\}$$

- *If $\varphi = \neg\psi$, then*

$$||\varphi||^{\mathscr{M}}_{g,w} := W \setminus ||\psi||^{\mathscr{M}}_{g,w}$$

- *If $\varphi = \psi \wedge \chi$, then*

$$||\varphi||^{\mathscr{M}}_{g,w} := ||\psi||^{\mathscr{M}}_{g,w} \cap ||\chi||^{\mathscr{M}}_{g,w}$$

- *If $\varphi = \forall x \psi$, then*

$$||\varphi||^{\mathscr{M}}_{g,w} := \{v \in W : v \in ||\psi||^{\mathscr{M}}_{h,w} \text{ for all } x\text{-variants } h \text{ of } g \text{ at } v\}$$

- *If $\varphi = \Box\psi$, then*

$$||\varphi||^{\mathscr{M}}_{g,w} := \{v \in W : ||\psi||^{\mathscr{M}}_{g,w} = W\}$$

- *If $\varphi = \bigcirc(\psi/\chi)$, then*

$$||\varphi||^{\mathscr{M}}_{g,w} := \{v \in W : best(||\chi||^{\mathscr{M}}_{g,w}) \subseteq ||\psi||^{\mathscr{M}}_{g,w}\}$$

- *If $t = \iota x \varphi$, then*

$$(I*g)_w(t) := \begin{cases} h(x,w) & \text{if } h \text{ is the } \textbf{unique } x\text{-variant of } g \text{ at } w \\ & \text{such that } w \in ||\varphi||^{\mathscr{M}}_{h,w} \\ \mathbb{D} & \text{otherwise} \end{cases}$$

Definition 3.2.9 (Truth in $\mathbf{F}_1^{\forall}$)**.** *Given a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g, a formula φ and a world w we define what it means that φ is true in $\mathscr{M}$ in $\mathbf{F}_1^{\forall}$ at w under g (in symbols: $\mathscr{M}, w \models^1_g \varphi$) as*

$$\mathscr{M}, w \models^1_g \varphi :\Leftrightarrow w \in ||\varphi||^{\mathscr{M}}_{g,w}$$

Remark 3.2.10. *From the definition above it follows that for any model $\mathscr{M} = \langle W, \succeq, D, I\rangle$, variable assignment g, and world w given that a formula $\varphi \in WF$*

is of the form $\Box\psi, \bigcirc(\psi/\chi)$ *or* $t_1 = t_2$ *then the set* $||\varphi||^{\mathscr{M}}_{g,w}$ *is either equal to* W *or to* $\emptyset$.

It is important to note that $||\varphi||^{\mathscr{M}}_{g,w} = W$ in general does NOT imply $\mathscr{M}, v \models^1_g \varphi$ for any $v \in W$ other than w. The equality $||\varphi||^{\mathscr{M}}_{g,w} = W$ can be seen as a φ being necessarily true at the world w, which means that φ is true at every world for someone who refers to every object the way they are defined at the world w. Therefore $||t_1 = t_2||^{\mathscr{M}}_{g,w} = W$ should not be confused with t_1 and t_2 having the same interpretation at every world and rather understood as t_1 and t_2 being equal for someone who refers to every object the way they are defined at world w, which does not depend on any world different from w. The meaning of $\mathscr{M}, w \models^1_g \Box\varphi$ is that φ holds true under g at every world from the point of view of someone living at w.
Now we can define the properties we want our models to have and use them to define a whole class of models:

Definition 3.2.11. *We say that a model* $\mathscr{M} = \langle W, \succeq, D, I\rangle$ *is reflexive if*

$$\forall w \in W : w \succeq w.$$

We say that a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$ *fulfils the limitedness property (of* $\boldsymbol{F}^{\forall}_1$*) if for every formula* φ*, variable assignment* g *and world* $w \in W$ *we have*

$$||\varphi||^{\mathscr{M}}_{g,w} \neq \emptyset \Rightarrow best(||\varphi||^{\mathscr{M}}_{g,w}) \neq \emptyset.$$

We define $\mathscr{U}_1$ *as the class of all models which are reflexive and fulfil the limitedness property of* $\boldsymbol{F}^{\forall}_1$.

Let us start with the explanation of why we want our models to fulfil the limitedness property. Given a model $\mathscr{M}$, a world w, a formula φ and a variable assignment g, we want that if there exists a world v (not necessarily different from w) at which the formula φ is true under g for a person living at w than there should also exist an optimal φ-world of $\mathscr{M}$ under g according to w. This guarantees us that given a model $\mathscr{M}$, a world w, a formula φ and a variable assignment g, there is no infinite chain of increasingly better worlds at which the formula φ is true under g for a person living at w. Also, in the propositional case, the limitedness property is the one that is added to the models to make the axiom **D*** of the system **F** sound, without it, this is, in general not true.
In the case of reflexivity, let us take any model $\mathscr{M}$ and any world w, then it seems fitting to say that this world w is at least as good as w itself. Also, regarding the definition of our best worlds, we need the betterness relation to be reflexive. Otherwise, if a world w is not in relation with itself, then w can never be an element of a best set of worlds. More specific if $w \not\succeq w$ then for any set if worlds $W' \subseteq W$ we have $w \notin best(W')$. Given

a model $\mathscr{M}$ which does not fulfil reflexivity but does fulfil the limitedness property we could find a world w such that $w \not\succeq w$ and therefore $best(\{w\}) = \emptyset$. This would clash with the limitedness property, so to eliminate such redundant cases, we just add the property of reflexivity to all of our models.
We have defined what it means in $\mathbf{F}_1^{\forall}$ for a formula φ to be true in a model $\mathscr{M}$ at a world w under a variable assignment g, and now we want to scale this definition up to the entire class of models $\mathscr{U}_1$.

Definition 3.2.12 (Validity in $\mathbf{F}_1^{\forall}$)**.** *A formula $\varphi \in WF$ is called valid in a model $\mathscr{M}$ in $\mathbf{F}_1^{\forall}$ at a world w if for every variable assignment g we have that φ is true in $\mathscr{M}$ at w under g, in symbols:*

$$\mathscr{M}, w \models^1 \varphi$$

A formula $\varphi \in WF$is called valid in $\mathscr{M}$ in $\mathbf{F}_1^{\forall}$ if for every world w we have φ is true in $\mathscr{M}$ at w, in symbols:

$$\mathscr{M} \models^1 \varphi$$

A formula $\varphi \in WF$ is called valid in a class of models $\mathbb{M}$ in $\mathbf{F}_1^{\forall}$. If for every model $\mathscr{M} \in \mathbb{M}$ we have φ is valid in $\mathscr{M}$ in $\mathbf{F}_1^{\forall}$, in symbols:

$$\mathbb{M} \models^1 \varphi$$

A formula $\varphi \in WF$ is called valid in $\mathbf{F}_1^{\forall}$ if φ is valid in the class $\mathscr{U}_1$ in $\mathbf{F}_1^{\forall}$ from Definition 3.2.11, in symbols:

$$\models^1 \varphi$$

Using the definition of validity we can give a even more specific definition, the one of *semantic entailment*. This definition can be used to describe what it means for one or more formulas to semantically entail another formula in $\mathbf{F}_1^{\forall}$.

Definition 3.2.13. *Given a set of formulas $\Gamma \subseteq WF$ and a formula $\varphi \in WF$ then we say Γ semantically entails φ in $\mathbf{F}_1^{\forall}$ (in symbols: $\Gamma \models^1 \varphi$) if for every model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, every variable assignment g and every world $w \in W$ we have*

$$(\forall \psi \in \Gamma : \mathscr{M}, w \models^1_g \psi) \Rightarrow \mathscr{M}, w \models^1_g \varphi.$$

Remark 3.2.14. $\models^1 \varphi \Leftrightarrow \emptyset \models^1 \varphi$

We should mention what happens to non-denoting terms in our semantics. More precisely, what is the interpretation of a definite description that does not exist or is not unique. An example of such term is $\imath x\bot$. Another would $\imath x\exists y(x=y)$ if the domain of the world at which this formula is evaluated contains more than one element. By definition a non-denoting definite description will always be assigned to the element $\mathbb{D}$. This element is not contained in the existing domain or in other words, not contained in the domain of any world [7]. Therefore in $\mathbf{F}_1^{\forall}$, the formula $\exists y(y=\imath x\varphi)$ will never be true at any model and any world at which $\imath x\varphi$ does not denote. Although formulas containing a non-denoting definite description can be true at a world or even be valid, for example, the formula $\imath x\bot = \imath x\bot$ is valid. The formula $A(\imath x\bot)$ can be true at a world depending on the interpretation of the predicate symbol A. A non-denoting definite description can have certain properties but can neither exist nor be equal to something existing.

3.3 Logical Symbols and Derived Connectives

In this section we show that the truth definition at a world, and therefore the validity definition in $\mathbf{F}_1^{\forall}$, behave the usual way. This means that the logical connectives and the quantifiers have identical or similar interpretations as in classical first-order logic. We also show how the operators $\Box$ and $\bigcirc$ as well as all the derived connectives, given in Definition 3.1.5, behave in $\mathbf{F}_1^{\forall}$.

We start with formulas of the form $\neg\varphi$ and $\varphi\wedge\psi$:

Lemma 3.3.1. *Given any model $\mathscr{M}=\langle W,\succeq,D,I\rangle$, variable assignment g, world $w\in W$ and any two formulas $\varphi,\psi\in WF$, then*

$$\begin{aligned}\mathscr{M},w\models_g^1\neg\varphi &\quad\Leftrightarrow\quad \mathscr{M},w\not\models_g^1\varphi\\ \mathscr{M},w\models_g^1\varphi\wedge\psi &\quad\Leftrightarrow\quad \mathscr{M},w\models_g^1\varphi \text{ and } \mathscr{M},w\models_g^1\psi.\end{aligned}$$

Proof. By definition $||\neg\varphi||_{g,w}^{\mathscr{M}}=W\backslash||\varphi||_{g,w}^{\mathscr{M}}$. This implies that $w\in||\neg\varphi||_{g,w}^{\mathscr{M}}$ is equivalent to $w\notin||\varphi||_{g,w}^{\mathscr{M}}$. Also by definition $||\varphi\wedge\psi||_{g,w}^{\mathscr{M}}=||\varphi||_{g,w}^{\mathscr{M}}\cap||\psi||_{g,w}^{\mathscr{M}}$. This implies that $w\in||\varphi\wedge\psi||_{g,w}^{\mathscr{M}}$ is equivalent to w being an element of $||\varphi||_{g,w}^{\mathscr{M}}$ and $||\psi||_{g,w}^{\mathscr{M}}$.

□

Now we take a closer look at the derived connectives $\vee,\supset$ and $\leftrightarrow$:

[7] The idea is taken from Goble's paper [9].

Lemma 3.3.2. *Given any model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, variable assignment g, world $w \in W$ and any two formulas $\varphi, \psi \in WF$, then*

$$\begin{aligned}
\mathscr{M}, w \models_g^1 \varphi \vee \psi &\quad\Leftrightarrow\quad \mathscr{M}, w \models_g^1 \varphi \text{ or } \mathscr{M}, w \models_g^1 \psi \\
\mathscr{M}, w \models_g^1 \varphi \supset \psi &\quad\Leftrightarrow\quad \mathscr{M}, w \not\models_g^1 \varphi \text{ or } \mathscr{M}, w \models_g^1 \psi \\
\mathscr{M}, w \models_g^1 \varphi \leftrightarrow \psi &\quad\Leftrightarrow\quad \mathscr{M}, w \models_g^1 \varphi \text{ iff } \mathscr{M}, w \models_g^1 \psi.
\end{aligned}$$

Proof. The equality $||\varphi \vee \psi||_{g,w}^{\mathscr{M}} = ||\varphi||_{g,w}^{\mathscr{M}} \cup ||\psi||_{g,w}^{\mathscr{M}}$ [8] implies that $w \in ||\varphi \vee \psi||_{g,w}^{\mathscr{M}}$ is equivalent to $w \in ||\varphi||_{g,w}^{\mathscr{M}}$ or $w \in ||\psi||_{g,w}^{\mathscr{M}}$.
The equality $||\varphi \supset \psi||_{g,w}^{\mathscr{M}} = W \backslash (||\varphi||_{g,w}^{\mathscr{M}}) \cup ||\psi||_{g,w}^{\mathscr{M}}$ [9] implies that $w \in ||\varphi \supset \psi||_{g,w}^{\mathscr{M}}$ is equivalent to $w \notin ||\varphi||_{g,w}^{\mathscr{M}}$ or $w \in ||\psi||_{g,w}^{\mathscr{M}}$.
The last statement follows directly from the following chain of equivalences:

$$\begin{aligned}
w \in ||\varphi \leftrightarrow \psi||_{g,w}^{\mathscr{M}} &\Leftrightarrow \\
w \in ||\neg(\varphi \wedge \neg\psi) \wedge \neg(\neg\varphi \wedge \psi)||_{g,w}^{\mathscr{M}} &\Leftrightarrow \\
w \in (W \backslash ||\varphi||_{g,w}^{\mathscr{M}} \cup ||\psi||_{g,w}^{\mathscr{M}}) \cap (||\varphi||_{g,w}^{\mathscr{M}} \cup W \backslash ||\psi||_{g,w}^{\mathscr{M}}) &\Leftrightarrow \\
w \in (||\varphi||_{g,w}^{\mathscr{M}} \cap ||\psi||_{g,w}^{\mathscr{M}}) \cup (W \backslash (||\varphi||_{g,w}^{\mathscr{M}} \cup ||\psi||_{g,w}^{\mathscr{M}})) &\Leftrightarrow \\
w \in ||\varphi||_{g,w}^{\mathscr{M}} \cap ||\psi||_{g,w}^{\mathscr{M}} \quad \text{or} \quad w \notin ||\varphi||_{g,w}^{\mathscr{M}} \cup ||\psi||_{g,w}^{\mathscr{M}} &
\end{aligned}$$

□

Remark 3.3.3. *For the symbols $\perp$ and $\top$ it follows from the lemmas above that for any model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, variable assignment g, and world $w \in W$ the equalities $||\perp||_{g,w}^{\mathscr{M}} = \emptyset$ and $||\top||_{g,w}^{\mathscr{M}} = W$ hold.*

After establishing the interpretation of all propositional connectives contained in classical logic, we can show that $\mathbf{F}_1^\forall$ contains classical propositional logic as a sublogic. This means that any propositional tautology is valid in $\mathbf{F}_1^\forall$. First, we must properly define what "being a propositional tautology" means in our first-order dyadic deontic logic.

Definition 3.3.4. *A well-formed formula $\varphi \in WF$ is called a propositional tautology if there exists a tautology τ in classical propositional logic such that φ can be obtained by uniformly subsidising propositional variables with well-formed formulas of WF.*

[8] $||\varphi \vee \psi||_{g,w}^{\mathscr{M}} = W \backslash (||\neg\varphi \wedge \neg\psi||_{g,w}^{\mathscr{M}}) = W \backslash ((W \backslash ||\varphi||_{g,w}^{\mathscr{M}}) \cap (W \backslash ||\psi||_{g,w}^{\mathscr{M}})) = ||\varphi||_{g,w}^{\mathscr{M}} \cup ||\psi||_{g,w}^{\mathscr{M}}$
[9] $||\varphi \supset \psi||_{g,w}^{\mathscr{M}} = ||\neg(\varphi \wedge \neg\psi)||_{g,w}^{\mathscr{M}} = W \backslash (||\varphi||_{g,w}^{\mathscr{M}} \cap (W \backslash ||\psi||_{g,w}^{\mathscr{M}})) = W \backslash (||\varphi||_{g,w}^{\mathscr{M}}) \cup ||\psi||_{g,w}^{\mathscr{M}}$

Example 3.3.5. *Some examples for propositional tautologies of our logic are:*

- $\bigcirc(\varphi/\psi) \vee \neg\bigcirc(\varphi/\psi)$
- $\forall x\varphi \supset \forall x\varphi$
- $\neg(\Box\varphi \wedge \neg\Box\varphi)$

We can use the lemmas above to show that any propositional tautology is valid in $\mathbf{F}_1^{\forall}$. More precisely:

Theorem 3.3.6. *Given a propositional tautology φ, a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, variable assignment g of $\mathcal{M}$, and world $w \in W$ then $\mathcal{M}, w \models_g^1 \varphi$.*

Proof. This follows directly from Lemma 3.3.1 and Lemma 3.3.2. □

For the final part of this section we are going to take a closer look the interpretations of all the other operators and connectives, not yet discussed in this section. The following lemmas are going to be used in proofs throughout this thesis.

Lemma 3.3.7. *Given any model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, variable assignment g, world $w \in W$ and any two formulas $\varphi, \psi \in WF$, then*

$$\begin{aligned} \mathcal{M}, w \models_g^1 \Box\varphi \quad &\Leftrightarrow \quad W = ||\varphi||_{g,w}^{\mathcal{M}}. \\ \mathcal{M}, w \models_g^1 \bigcirc(\varphi/\psi) \quad &\Leftrightarrow \quad best(||\psi||_{g,w}^{\mathcal{M}}) \subseteq ||\varphi||_{g,w}^{\mathcal{M}}. \end{aligned}$$

Proof. Both claims follow directly from Definition 3.2.8. □

Lemma 3.3.8. *Given any model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, variable assignment g, world $w \in W$ and any two formulas $\varphi, \psi \in WF$, then*

$$\begin{aligned} \mathcal{M}, w \models_g^1 \Diamond\varphi \quad &\Leftrightarrow \quad ||\varphi||_{g,w}^{\mathcal{M}} \neq \emptyset \\ \mathcal{M}, w \models_g^1 P(\varphi/\psi) \quad &\Leftrightarrow \quad best(||\psi||_{g,w}^{\mathcal{M}}) \cap ||\varphi||_{g,w}^{\mathcal{M}} \neq \emptyset \end{aligned}$$

Proof. The first claim follows from the following chain of equivalences:

$$\begin{aligned} \mathcal{M}, w \models_g^1 \Diamond\varphi \Leftrightarrow w \in ||\Diamond\varphi||_{g,w}^{\mathcal{M}} \Leftrightarrow w \notin ||\Box\neg\varphi||_{g,w}^{\mathcal{M}} \Leftrightarrow ||\Box\neg\varphi||_{g,w}^{\mathcal{M}} = \emptyset \\ \Leftrightarrow \exists v \in W : v \notin ||\neg\varphi||_{g,w}^{\mathcal{M}} \Leftrightarrow \exists v \in W : v \in ||\varphi||_{g,w}^{\mathcal{M}} \Leftrightarrow ||\varphi||_{g,w}^{\mathcal{M}} \neq \emptyset \end{aligned}$$

The second claim follows from:

$$\begin{aligned}\mathcal{M},w \models^1_g P(\varphi/\psi) &\Leftrightarrow w \in ||P(\varphi/\psi)||^{\mathcal{M}}_{g,w} \Leftrightarrow w \notin ||O(\neg\varphi/\psi)||^{\mathcal{M}}_{g,w}\\ &\Leftrightarrow \{v \in W : best(||\psi||^{\mathcal{M}}_{g,w}) \subseteq ||\neg\varphi||^{\mathcal{M}}_{g,w}\} = \emptyset\\ &\Leftrightarrow best(||\psi||^{\mathcal{M}}_{g,w}) \not\subseteq ||\neg\varphi||^{\mathcal{M}}_{g,w}\\ &\Leftrightarrow best(||\psi||^{\mathcal{M}}_{g,w}) \not\subseteq W \setminus ||\varphi||^{\mathcal{M}}_{g,w}\\ &\Leftrightarrow \exists v \in W : v \in best(||\psi||^{\mathcal{M}}_{g,w}) \text{ and } v \in ||\varphi||^{\mathcal{M}}_{g,w}\\ &\Leftrightarrow best(||\psi||^{\mathcal{M}}_{g,w}) \cap ||\varphi||^{\mathcal{M}}_{g,w} \neq \emptyset\end{aligned}$$

□

Lemma 3.3.9. *Given any model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, variable $x \in V$ variable assignment g, world $w \in W$ and formula $\varphi \in WF$, then*

$\mathcal{M},w \models^1_g \forall x\varphi$ holds iff $\mathcal{M},w \models^1_h \varphi$ holds for every x-variant h of g at w.

$\mathcal{M},w \models^1_g \exists x\varphi$ holds iff $\mathcal{M},w \models^1_h \varphi$ holds for an x-variant of g at w.

*$\mathcal{M},w \models^1_g \exists! x\varphi$ holds iff $\mathcal{M},w \models^1_h \varphi$ holds for **exactly one** x-variant h of g at w.*

Proof. The first claim follows directly from Definitions 3.2.8 and 3.2.9.
The second claim follows from the fact that by definition $w \in ||\exists x\varphi||^{\mathcal{M}}_{g,w}$ is equivalent to $w \notin ||\forall x \neg\varphi||^{\mathcal{M}}_{g,w}$. This is equivalent to the existence of an x-variant h of g at w such that $w \notin ||\neg\varphi||^{\mathcal{M}}_{h,w}$, which is furthermore equivalent to $w \in ||\varphi||^{\mathcal{M}}_{h,w}$.
The last claim can be shown the following way. By definition $w \in ||\exists! x\varphi||^{\mathcal{M}}_{g,w}$ is equivalent to $w \in ||\exists y \forall x(x = y \leftrightarrow \varphi)||^{\mathcal{M}}_{g,w}$. This is equivalent to the existence of a y-variant j of g at w such that $w \in ||\forall x(x = y \leftrightarrow \varphi)||^{\mathcal{M}}_{j,w}$. Furthermore this is equivalent to the existence of a y-variant j of g at w such that for all x-variants h of j at w it holds that $w \in ||x = y \leftrightarrow \varphi||^{\mathcal{M}}_{h,w}$. This lets us conclude that there exists a unique element $d \in D_w$ such that $w \in ||\varphi||^{\mathcal{M}}_{h,w}$ holds exactly for the x-variant h at w with $h = g_{x \Rightarrow d}$.

□

Lemma 3.3.10. *Given any model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, variable assignment g, world $w \in W$ and any two formula $\varphi, \psi \in WF$ then*

$$\mathcal{M},w \models^1_g \Box(\varphi \leftrightarrow \psi) \quad \Leftrightarrow \quad ||\varphi||^{\mathcal{M}}_{g,w} = ||\psi||^{\mathcal{M}}_{g,w}.$$

Proof. This follows directly from the proof of Lemma 3.3.2 and the interpretation of the $\Box$ operator in $\mathbf{F}^{\forall}_1$.

□

Lemma 3.3.11. *Given any model* $\mathscr{M} = \langle W, \succeq, D, I \rangle$*, variable assignment* g*, world* $w \in W$ *and any two terms* t *and* s*, then*

$$\mathscr{M}, w \models^1_g t = s \quad \Leftrightarrow \quad (I * g)_w(t) = (I * g)_w(s)$$
$$\mathscr{M}, w \models^1_g E(t) \quad \Leftrightarrow \quad (I * g)_w(t) \in D_w.$$

Proof. The first claim follows directly from the definition since $w \in ||t = s||^{\mathscr{M}}_{g,w}$ is equivalent to $(I * g)_w(t) = (I * g)_w(s)$. The second claim follows from the following chain of equivalences:

$\mathscr{M}, w \models^1_g E(t)$ $\Leftrightarrow$

$w \in ||\exists(x = t)||^{\mathscr{M}}_{g,w}$ $\Leftrightarrow$

There exists an x-variant h of g at w such that $w \in ||(x = t)||^{\mathscr{M}}_{h,w}$ $\Leftrightarrow$

There exists an x-variant h of g at w such that $(I * h)_w(x) = (I * h)_w(t)$ $\Leftrightarrow$ [10]

There exists an x-variant h of g at w such that $h(x, w) = (I * g)_w(t)$ $\Leftrightarrow$

$(I * g)_w(t) \in D_w$

□

With the help of the above proven lemmas we can show that in $\mathbf{F}^{\forall}_1$, the dyadic deontic operator can be used to define the $\Box$ operator.

Theorem 3.3.12. $\Box\varphi \leftrightarrow \bigcirc(\bot/\neg\varphi)$ *is valid in* $\boldsymbol{F}^{\forall}_1$.

Proof. Given a model $\mathscr{M} \in \mathscr{U}_1$, a world $w \in W$, a variable assignment g and a formula φ then $v \in || \bigcirc (\bot/\neg\varphi)||^{\mathscr{M}}_{g,w}$ is equivalent to $best(W \backslash ||\varphi||^{\mathscr{M}}_{g,w}) \subseteq \emptyset$ for any world $v \in W$. Because of the limitedness property given to our models in Definition 3.2.11 the following equivalence holds $best(W \backslash ||\varphi||^{\mathscr{M}}_{g,w}) = \emptyset \Leftrightarrow ||\varphi||^{\mathscr{M}}_{g,w} = W$. This lets us conclude $|| \bigcirc (\bot/\neg\varphi)||^{\mathscr{M}}_{g,w} = ||\Box\varphi||^{\mathscr{M}}_{g,w}$.

□

3.4 Paradoxes revisited

Before moving on to the next chapter we show that the paradoxes discussed in Chapter 2 are resolved in $\mathbf{F}^{\forall}_1$. We start with the "Switching seats" paradox, see Example 2.5.1.

[10]This equivalence holds since x does not appear in t as a free variable.

Proposition 3.4.1. *Let* $\Gamma := \{\bigcirc(G(\imath yL(\imath xC(x),y))/G(\imath xC(x))), x_1 = \imath xC(x), x_2 = \imath yL(\imath xC(x),y)\}$ *then*

$$\Gamma \models^1 \bigcirc(G(x_2)/G(x_1)).$$

Proof. Let $\mathscr{M} = \langle W, \succeq, D, I\rangle$ be a model in $\mathscr{U}_1$. Given any variable assignment g of $\mathscr{M}$, if at a world $w \in W$ the formulas $\bigcirc(G(\imath y\chi(\imath xC(x),y))/G(\imath xC(x))), x_1 = \imath xC(x)$ and $x_2 = \imath yL(\imath xC(x),y)$ are true under g then we can infer:

$$best(||G(\imath xC(x))||^{\mathscr{M}}_{g,w}) \subseteq ||G(\imath yL(\imath xC(x),y))||^{\mathscr{M}}_{g,w}$$
$$g(x_1,w) = (I*g)_w(\imath xC(x))$$
$$g(x_2,w) = (I*g)_w(\imath yL(\imath xC(x),y))$$

This implies that for every world $v \in W$ $(I*g)_w(x_1) \in I(G,v)$ and $(I*g)_w(\imath xC(x)) \in I(G,v)$ are equivalent as well as $(I*g)_w(x_2) \in I(G,v)$ and $\imath yL(\imath xC(x),y) \in I(G,v)$. Hence the equalities $||G(x_1)||^{\mathscr{M}}_{g,w} = ||G(\imath xC(x))||^{\mathscr{M}}_{g,w}$ and $||G(x_2)||^{\mathscr{M}}_{g,w} = ||G(\imath yL(\imath xC(x),y))||^{\mathscr{M}}_{g,w}$ hold. We can conclude that $best(||G(x_1)||^{\mathscr{M}}_{g,w}) \subseteq ||G(x_2)||^{\mathscr{M}}_{g,w}$, which means

$$\mathscr{M}, w \models^1_g \bigcirc(G(x_2)/G(x_1)).$$ [11]

□

The "Kind tyrant" paradox also yields the desired semantic entailment, see Example 2.5.2.

Proposition 3.4.2. *Let* $\Gamma := \{\forall x \bigcirc(K(x)), \forall x \bigcirc(\neg T(x)), \exists y(y = \imath zT(z))\}$ *then*

$$\Gamma \models^1 \bigcirc(K(\imath zT(z))) \wedge \bigcirc(\neg T(\imath zT(z))).$$

Proof. Let $\mathscr{M} = \langle W, \succeq, D, I\rangle$ be a model in $\mathscr{U}_1$. Given any variable assignment g of $\mathscr{M}$, if at a world $w \in W$ the formulas $\forall x \bigcirc(K(x)), \forall x \bigcirc(\neg T(x))$ and $\exists y(y = \imath zT(z))$ are true, then we can infer that, for every x-variant h of g at w we have:

$$\mathscr{M}, w \models^1_h \bigcirc(K(x)), \quad \mathscr{M}, w \models^1_h \bigcirc(\neg T(x)) \quad \text{and} \quad (I*g)_w(\imath zT(z)) \in D_w$$

Therefore there exists an x-variant h of g at w with $h(x,w) = (I*g)_w(\imath zT(z))$ and since x does not appear as a free variable in $K(\imath zT(z))$ and $\neg T(\imath zT(z))$ this implies

$$\mathscr{M}, w \models^1_g \bigcirc(K(\imath zT(z))) \quad \text{and} \quad \mathscr{M}, w \models^1_g \bigcirc(\neg T(\imath zT(z))).$$

□

[11] This obligation does not have to be true at a world different from w.

An example of a model fulfilling the "Kind tyrant" paradox would be:

Example 3.4.3. *Let $\mathcal{M} := \langle W, \succeq, D, I\rangle$ be a model with*

$$W := \{w, v\}$$
$$\succeq := RC(\{(v,w)\})$$
$$D_w := \{a,b\}, \quad D_v := \{c\}$$
$$I(K,w) := \{b\}, \ I(K,v) := \{a,b,c\}$$
$$I(T,w) := \{a\}, \ I(T,v) := \{\}$$

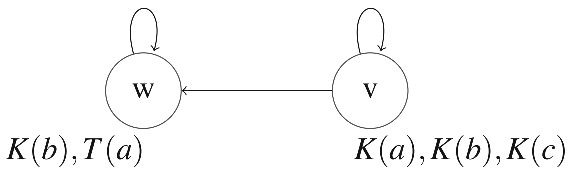

Figure 3.1: Kind tyrant model

Let g be any variable assignment of $\mathcal{M}$. If h is an x-variant of g at w then by definition $h(x,w) \in D_w = \{a,b\}$. Therefore $h(x,w) \in I(K,v)$ which means $v \in ||K(x)||^{\mathcal{M}}_{h,w}$. Since this implies $best(||\top||^{\mathcal{M}}_{h,w}) = \{v\} \subseteq ||K(x)||^{\mathcal{M}}_{h,w}$ we can conclude $\mathcal{M}, w \models^1_h \bigcirc K(x)$. This implies $\mathcal{M}, w \models^1_g \forall x \bigcirc K(x)$. If h is an x-variant of g at w then $h(x,w) \notin I(T,v)$ which means $v \in ||\neg T(x)||^{\mathcal{M}}_{h,w}$. Since this implies $best(||\top||^{\mathcal{M}}_{h,w}) = \{v\} \subseteq ||\neg T(x)||^{\mathcal{M}}_{h,w}$ we can conclude $\mathcal{M}, w \models^1_h \bigcirc \neg T(x)$. This implies $\mathcal{M}, w \models^1_g \forall x \bigcirc \neg T(x)$. Finally because $(I * g)_w(\imath z T(z)) = a \in D_w$ we have $\mathcal{M}, w \models^1_g \exists y (y = \imath z T(z))$.
We can see that $\mathcal{M}, w \models^1_g \bigcirc \neg T(\imath z T(z))$ must hold by the proposition above. $\mathcal{M}, w \models^1_g \bigcirc \neg T(\imath z T(z))$ can also be followed from the fact that $best(||\top||^{\mathcal{M}}_{g,w}) = \{v\} \subseteq \{v\} = ||\neg T(\imath z T(z))||^{\mathcal{M}}_{g,w}$.

4 Hilbert Axiomatisation

Here we introduce a Hilbert axiomatisation for $\mathbf{F}_1^{\forall}$. The axiomatisation, which we refer to as $\mathbf{HF}_1^{\forall}$ is shown to be (strongly) sound in $\mathbf{F}_1^{\forall}$, that is if every formula provable (from assumptions) in $\mathbf{HF}_1^{\forall}$ is true at every world under every variable assignment (at which the assumptions are true) in $\mathbf{F}_1^{\forall}$.

4.1 Provability in $\mathrm{F}_1^{\forall}$

The Hilbert system $\mathbf{HF}_1^{\forall}$ consists of the following axiom schemas and inference rules:

Axioms:

All axioms of system **F**	(**F**)
$t = s \supset (\varphi \leftrightarrow \varphi_{t \hookrightarrow s})$	(**I**)
$E(t) \supset (\forall x\varphi \supset \varphi_{x \Rightarrow t})$	(**UI**)
$\exists x \exists y (x = y)$	(**Ex**)
$t = t$	(**E1**)
$t \neq s \supset \Box t \neq s$	(**E2**)
$\forall y((\forall x(\varphi \leftrightarrow x = y)) \supset y = \iota x\varphi)$	(**D1**)
$E(\iota x\varphi) \supset \exists! x\varphi$	(**D2**)
$\forall x(E(x) \supset \varphi) \supset \forall x\varphi$	(**UQ**)
$(\forall x\varphi \wedge \forall x\psi) \leftrightarrow \forall x(\varphi \wedge \psi)$	(**QD**)

Rules:

If $\vdash \varphi \supset t \neq x$ then $\vdash \neg\varphi$ where $x \notin free(\varphi)$	(**R1**)
If $\vdash \varphi$ and $\vdash \varphi \supset \chi$ then $\vdash \chi$	(**MP**)
If $\vdash \varphi^*$ then $\vdash \Box\varphi$	(**N***)
If $\vdash \varphi \supset \psi$ then $\vdash \varphi \supset \forall x\psi$ where $x \notin free(\varphi)$	(**IU**)
If $\vdash \varphi \supset \Box\psi$ then $\vdash \varphi \supset \Box\forall x\psi$ where $x \notin free(\varphi)$	(**IUB**)

Remark 4.1.1. *Axioms **I** and **UI** have the usual restrictions concerning the bound variables in φ (as mentioned in Definition 3.1.6). This means in **I** it must hold that $bv(\varphi) \cap fv(t) = \emptyset$ and $bv(\varphi) \cap fv(s) = \emptyset$ and in **UI** it must hold that $bv(\varphi) \cap fv(t) = \emptyset$.*

$\mathbf{HF}_1^{\forall}$ includes every axiom of **F**, the Axiom of Replacement, and Universal Instantiation with no restrictions on which operators appear in the formulas. It also contains the rule **MP** and a variation of the old rule **N**, called **N***, which is discussed in detail in Section 4.2. Furthermore $\mathbf{HF}_1^{\forall}$ contains the axiom **Ex** which shows that there is always an existing object. This relates to the non-emptiness of every world domain. The axioms **E1** and **E2** define the equality symbol and its relation to the necessitation operator. **D1** and **D2** are the axioms discussed in Chapter 2 to define definite description in our system. The axioms **UQ** and **QD** and the rules **IU** and **IUB** are here for introducing the universal quantifier and its relation with $\Box$, $E(x)$, $\wedge$ and free and bound variables. The rule **R1** states that every element in the whole domain can be assigned to any free variable in our system.[1]

Given $\mathbf{HF}_1^{\forall}$ above, we can define what it means for a well-formed formula to be provable. We call **R1**, **MP**, **IU** and **IUB** the binary inference rules of $\mathbf{HF}_1^{\forall}$.

Definition 4.1.2 (Provability in $\mathbf{HF}_1^{\forall}$)**.** *Let $\varphi \in WF$ be a well-formed formula. We say a sequence $\varphi_1, \ldots, \varphi_n$ of well-formed formulas is a (formal) $\mathbf{HF}_1^{\forall}$-proof of φ if $\varphi_n = \varphi$ and for all $i = 1, \ldots, n$ one of the following holds:*

- *φ_i is an instance of an axiom schema of $\mathbf{HF}_1^{\forall}$.*
- *There exist $j, k < i$ such that φ_i is the result of an application of a binary inference rule of $\mathbf{HF}_1^{\forall}$ to φ_j and φ_k.*
- *There exists $j < i$ such that $\varphi_i = \Box\psi$ and $\varphi_j = \psi^*$.*

We say a well-formed formula $\varphi \in WF$ is provable in $\mathbf{HF}_1^{\forall}$ (in symbols: $\vdash^1 \varphi$) if there exists a $\mathbf{HF}_1^{\forall}$-proof of φ.

An example of a provable formula schema is:

Example 4.1.3. *The formula schema $t = s \supset \Box t = s$ is provable in $\mathbf{HF}_1^{\forall}$. A formal $\mathbf{HF}_1^{\forall}$-proof of $t = s \supset \Box t = s$ is*

[1]The axioms and rules introducing the universal quantifier are taken from [18] and [9].

(a) $x = x$	*(**E1**)*
(b) $\Box t = t$	*(a + **N***)*
(c) $t = s \supset (\Box t = t \leftrightarrow \Box t = s)$	*(**I**)*
(d) $\Box t = t \supset [(t = s \supset (\Box t = t \leftrightarrow \Box t = s)) \supset (t = s \supset \Box t = s)]$	*(Tautology)*
(e) $(t = s \supset (\Box t = t \leftrightarrow \Box t = s)) \supset (t = s \supset \Box t = s)$	*(b + d + **MP**)*
(f) $t = s \supset \Box t = s$	*(c + e + **MP**)*

Definition 4.1.4 (Derivability from assumptions in $\mathbf{HF}_1^{\forall}$)**.** *Let $\varphi \in WF$ be a well-formed formula and $\Gamma \subseteq WF$ a set of well-formed formulas, which we call assumptions. We say φ is derivable from Γ in $\mathbf{HF}_1^{\forall}$ (in symbols: $\Gamma \vdash^1 \varphi$) if there exists a sequence $\varphi_1, ..., \varphi_n$ of well-formed formulas with $\varphi_n = \varphi$ and for all $i = 1, ..., n$ one of the following holds:*

- $\vdash^1 \varphi_i$.
- $\varphi_i \in \Gamma$
- *There exist $j, k < i$ such that φ_i is the result of an application of modus ponens to φ_j and φ_k.*

Remark 4.1.5. $\vdash^1 \varphi \Leftrightarrow \emptyset \vdash^1 \varphi$. *This means the definition of derivability from assumptions in $\mathbf{HF}_1^{\forall}$ is more general than the definition of derivability from assumptions in $\mathbf{HF}_1^{\forall}$. Therefore we also say "φ is provable from the assumptions Γ" if $\Gamma \vdash^1 \varphi$ holds.*

Given the definition of derivability in $\mathbf{HF}_1^{\forall}$, we can show that the "Kind tyrant" paradox from Chapter 2 gives the desired result in $\mathbf{HF}_1^{\forall}$. In other words, the formulas $T(\imath z T(z)))$ $\bigcirc(K(\imath z T(z)))$ are derivable from the set $\{\forall x \bigcirc (K(x)), \forall x \bigcirc (\neg T(x)), \exists y(y = \imath z T(z))\}$. Let us consider the following example:

Example 4.1.6. $\bigcirc(\neg T(\imath z T(z)))$ *is derivable from* $\forall x \bigcirc (\neg T(x))$ *and* $\exists y(y = \imath z T(z))$.

$$\{\forall x \bigcirc (\neg T(x)), \exists y(y = \imath z T(z))\} \vdash^1 \bigcirc(\neg T(\imath z T(z)))$$

(a) $\exists y(y = \imath z T(z)) \supset (\forall x \bigcirc (\neg T(x)) \supset \bigcirc(\neg T(\imath z T(z)))$	*(**UI**)*
(b) $\exists y(y = \imath z T(z))$	*(Assumption)*
(c) $(\forall x \bigcirc (\neg T(x))) \supset \bigcirc(\neg T(\imath z T(z))$	*(a + b + **MP**)*
(d) $\forall x \bigcirc (\neg T(x))$	*(Assumption)*
(e) $\neg T(\imath z T(z)$	*(c + d + **MP**)*

Before we show the connection between the semantics defined in Chapter 3 and the Hilbert axiomatisation defined in this chapter we need a definition which captures this connection.

Definition 4.1.7 (Soundness)**.** *We say* $\mathbf{HF}_1^{\forall}$ *is (weakly) sound in* $\mathbf{F}_1^{\forall}$ *if for every well-formed formula* $\varphi \in WF$ *the implication*

$$\vdash^1 \varphi \Rightarrow \models^1 \varphi$$

holds. We say $\mathbf{HF}_1^{\forall}$ *is strongly sound in* $\mathbf{F}_1^{\forall}$ *if for every well-formed formula* $\varphi \in WF$ *and every set of well-formed formulas* $\Gamma \subseteq WF$ *the implication*

$$\Gamma \vdash^1 \varphi \Rightarrow \Gamma \models^1 \varphi$$

holds.

The weak soundness proof for $\mathbf{HF}_1^{\forall}$ can be accomplished by showing that every instance of an axiom of $\mathbf{HF}_1^{\forall}$ is valid in $\mathbf{F}_1^{\forall}$ and by showing that the inference rules of $\mathbf{HF}_1^{\forall}$ preserve validity. In those cases, we call the axiom schemas and the inference rules sound in $\mathbf{F}_1^{\forall}$. At the end of this chapter we will see that strong soundness follows from weak soundness.

4.2 Inclusion of F

In this section, we show that all rules and axioms of **F**, except of **N**, are sound in $\mathbf{F}_1^{\forall}$. This shows us that the semantics of $\mathbf{F}_1^{\forall}$ does indeed extend the propositional system **F**, except of **N**, into first-order. We also discuss the new rule **N*** in detail.

Theorem 4.2.1. *All axioms of* **F** *and the rule modus ponens are sound in* $\mathbf{F}_1^{\forall}$*.*

Proof. Let in the following φ, ψ and χ be well-formed formulas, $\mathcal{M} = \langle W, \succeq, D, I \rangle$ a model, $w \in W$ a world and g a variable assignment of $\mathcal{M}$:

- φ is a truth-functional tautology.

  The validity of φ follows from Theorem 3.3.6

- $\Box(\varphi \supset \chi) \supset (\Box\varphi \supset \Box\chi)$

  $\mathcal{M}, w \models_g^1 \Box(\varphi \supset \chi)$ is equivalent to $W \setminus ||\varphi||_{g,w}^{\mathcal{M}} \cup ||\chi||_{g,w}^{\mathcal{M}} = W$ [2]. Now if $w \in ||\Box\varphi||_{g,w}^{\mathcal{M}}$ then $||\varphi||_{g,w}^{\mathcal{M}} = W$, therefore together with $\mathcal{M}, w \models_g^1 \Box(\varphi \supset \chi)$ it follows that $||\chi||_{g,w}^{\mathcal{M}} = W$, which lets us conclude $\mathcal{M}, w \models_g^1 \Box\chi$.

- $\Box\varphi \supset \varphi$

  Since $\mathcal{M}, w \models_g^1 \Box\varphi$ holds iff $||\varphi||_{g,w}^{\mathcal{M}} = W$ it follows that φ holds in w under g according to w, which means $\mathcal{M}, w \models_g^1 \varphi$.

[2] $\mathcal{M}, w \models_g^1 \Box(\varphi \supset \chi) \Leftrightarrow w \in ||\Box(\varphi \supset \chi)||_{g,w}^{\mathcal{M}} \Leftrightarrow ||\varphi \supset \chi||_{g,w}^{\mathcal{M}} = W \Leftrightarrow W \setminus ||\varphi||_{g,w}^{\mathcal{M}} \cup ||\chi||_{g,w}^{\mathcal{M}} = W$

- $\Diamond\varphi \supset \Box\Diamond\varphi$

  If $\mathscr{M}, w \models^1_g \Diamond\varphi$ then $w \in ||\neg\Box\neg\varphi||^{\mathscr{M}}_{g,w}$ which implies $||\neg\Box\neg\varphi||^{\mathscr{M}}_{g,w} = W$, therefore $||\Box\neg\Box\neg\varphi||^{\mathscr{M}}_{g,w} = W$, hence $w \in ||\Box\Diamond\varphi||^{\mathscr{M}}_{g,w}$.

- $\bigcirc(\varphi \supset \chi/\psi) \supset (\bigcirc(\varphi/\psi) \supset \bigcirc(\chi/\psi))$

  This follows from the following three equivalences:
  $\mathscr{M}, w \models^1_g \bigcirc(\varphi \supset \chi/\psi) \Leftrightarrow best(||\psi||^{\mathscr{M}}_{g,w}) \subseteq W \backslash ||\varphi||^{\mathscr{M}}_{g,w} \cup ||\chi||^{\mathscr{M}}_{g,w}$,
  $\mathscr{M}, w \models^1_g \bigcirc(\varphi/\psi) \Leftrightarrow best(||\psi||^{\mathscr{M}}_{g,w}) \subseteq ||\varphi||^{\mathscr{M}}_{g,w}$ and
  $\mathscr{M}, w \models^1_g \bigcirc(\chi/\psi) \Leftrightarrow best(||\psi||^{\mathscr{M}}_{g,w}) \subseteq ||\chi||^{\mathscr{M}}_{g,w}$.

- $\bigcirc(\varphi/\psi) \supset \Box\bigcirc(\varphi/\psi)$

  $\mathscr{M}, w \models^1_g \bigcirc(\varphi/\psi)$ is equivalent to $\{v \in W : best(||\psi||^{\mathscr{M}}_{g,w}) \subseteq ||\varphi||^{\mathscr{M}}_{g,w}\} = W$ and therefore to $||\bigcirc(\varphi/\psi)||^{\mathscr{M}}_{g,w} = W$, which means $w \in ||\Box\bigcirc(\varphi/\psi)||^{\mathscr{M}}_{g,w}$.

- $\Box\varphi \supset \bigcirc(\varphi/\psi)$

  Since $\mathscr{M}, w \models^1_g \Box\varphi$ is equivalent to $||\varphi||^{\mathscr{M}}_{g,w} = W$, it follows that $w \in ||\Box\varphi||^{\mathscr{M}}_{g,w}$ implies that for any formula $\psi \in WF$ we have $best(||\psi||^{\mathscr{M}}_{g,w}) \subseteq W = ||\varphi||^{\mathscr{M}}_{g,w}$.

- $\Box(\varphi \leftrightarrow \psi) \supset (\bigcirc(\chi/\varphi) \leftrightarrow \bigcirc(\chi/\psi))$

  $\mathscr{M}, w \models^1_g \Box(\varphi \leftrightarrow \psi)$ holds iff $||\varphi||^{\mathscr{M}}_{g,w} = ||\psi||^{\mathscr{M}}_{g,w}$ (see Lemma 3.3.10), therefore it follows that $best(||\varphi||^{\mathscr{M}}_{g,w}) = best(||\psi||^{\mathscr{M}}_{g,w})$. This implies that $best(||\varphi||^{\mathscr{M}}_{g,w}) \subseteq ||\chi||^{\mathscr{M}}_{g,w}$ is equivalent to $best^{\mathscr{M}}_{g,w}(\psi) \subseteq ||\chi||^{\mathscr{M}}_{g,w}$.

- $\bigcirc(\varphi/\varphi)$

  This follows from $best(||\varphi||^{\mathscr{M}}_{g,w}) \subseteq ||\varphi||^{\mathscr{M}}_{g,w}$ which holds by definition of the best worlds.

- $\bigcirc(\varphi/\psi \wedge \chi) \supset \bigcirc(\chi \supset \varphi/\psi)$

  Let $v \in best(||\psi||^{\mathscr{M}}_{g,w})$ be an arbitrary but fixed optimized φ-world of $\mathscr{M}$ under g according to w and let us assume that $\mathscr{M}, w \models^1_g \bigcirc(\varphi/\psi \wedge \chi)$ holds, then $best(||\psi \wedge \chi||^{\mathscr{M}}_{g,w}) \subseteq ||\varphi||^{\mathscr{M}}_{g,w}$. We now differ between two cases:
  First if $v \in ||\chi||^{\mathscr{M}}_{g,w}$ then $v \in ||\psi||^{\mathscr{M}}_{g,w} \cap ||\chi||^{\mathscr{M}}_{g,w}$. As a result for all v' with $v' \in ||\psi||^{\mathscr{M}}_{g,w} \cap ||\chi||^{\mathscr{M}}_{g,w} = ||\psi \wedge \chi||^{\mathscr{M}}_{g,w}$ we have $v \succeq v'$ which means that $v \in best(||\psi \wedge \chi||^{\mathscr{M}}_{g,w}) \subseteq ||\varphi||^{\mathscr{M}}_{g,w} \subseteq ||\chi \supset \varphi||^{\mathscr{M}}_{g,w}$. Since $v \in best(||\psi||^{\mathscr{M}}_{g,w})$ was arbitrary we can conclude $\mathscr{M}, w \models^1_g \bigcirc(\chi \supset \varphi/\psi)$.
  In the second case $v \notin ||\chi||^{\mathscr{M}}_{g,w}$ then $v \in W \backslash ||\chi||^{\mathscr{M}}_{g,w} \cup ||\varphi||^{\mathscr{M}}_{g,w} = ||\chi \supset \varphi||^{\mathscr{M}}_{g,w}$. Hence again since $v \in best(||\psi||^{\mathscr{M}}_{g,w})$ was arbitrary we can conclude $\mathscr{M}, w \models^1_g \bigcirc(\chi \supset \varphi/\psi)$.

- $\Diamond\psi \supset (\bigcirc(\varphi/\psi) \supset P(\varphi/\psi))$

  $\mathscr{M}, w \models^1_g \Diamond\psi$ is equivalent to the existence of a world $v \in W$ such that $v \in ||\varphi||^{\mathscr{M}}_{g,w}$ and $\mathscr{M}, w \models^1_g \bigcirc(\varphi/\psi)$ is equivalent to $best(||\psi||^{\mathscr{M}}_{g,w}) \subseteq ||\varphi||^{\mathscr{M}}_{g,w}$. Furthermore using the limitedness property of our betterness relation $\succeq$ and $v \in ||\varphi||^{\mathscr{M}}_{g,w}$ we can find a world $v' \in W$ such that $v' \in best(||\psi||^{\mathscr{M}}_{g,w})$. Now if $\mathscr{M}, w \models^1_g \Diamond\psi$ and $\mathscr{M}, w \models^1_g \bigcirc(\varphi/\psi)$ hold we can find a world v' with $v' \in best(||\psi||^{\mathscr{M}}_{g,w}) \subseteq ||\varphi||^{\mathscr{M}}_{g,w}$. This lets us conclude $best(||\psi||^{\mathscr{M}}_{g,w}) \cap ||\varphi||^{\mathscr{M}}_{g,w} \neq \emptyset$, which is equivalent to $\mathscr{M}, w \models^1_g P(\varphi/\psi)$.

- The rule **MP**

  Modus ponens follows directly from Lemma 3.3.2.

□

The only part of the System **F** which is not sound in $\mathbf{F}^{\forall}_1$ is the rule **N**. This can be formalised as the following proposition:

Proposition 4.2.2. *The implication*

$$\models^1 \psi \Rightarrow \models^1 \Box\psi$$

*does **not** hold for every formula $\psi \in WF$.*

This proposition can be proven by exhibiting a counterexample. We start by showing that the formula $\psi := \exists y(y = \imath x R(x)) \supset R(\imath x R(x))$ is a valid in $\mathbf{F}^{\forall}_1$. Afterwards we define a model $\mathscr{M}$ in which $\Box\exists y(y = \imath x R(x)) \supset R(\imath x R(x))$ is not valid.

Proposition 4.2.3. *Let $\mathscr{M} := \langle W, \succeq, D, I\rangle$ be a model, g a variable assignment, $w \in W$ a world, two variables $x, y \in V$ and a 1-place predicate symbol R, then:*

$$\mathscr{M}, w \models^1_g \exists y(y = \imath x R(x)) \supset R(\imath x R(x))$$

In particular:

$$\models^1 \exists y(y = \imath x R(x)) \supset R(\imath x R(x))$$

Proof. We fix a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g, and a world $w \in W$. In the case that $\mathscr{M}, w \not\models^1_g \exists y(y = \imath x R(x))$ the implication is true at w by default (see Lemma 3.3.2). In the other case, there exists a y-variant h of g at w such that $(I * h)_w(y) = (I * h)_w(\imath x R(x))$, which implies that $h(y, w) = e$ for an $e \in D_w$. By definition, e is the unique element with $e \in I(R, w)$. As a result it follows that $w \in ||R(\imath x R(x))||^{\mathscr{M}}_{h,w}$. Since

y does not appear in the formula $R(\iota x R(x))$ we have $||R(\iota x R(x))||^{\mathscr{M}}_{h,w} = ||R(\iota x R(x))||^{\mathscr{M}}_{g,w}$ which implies $w \in ||R(\iota x R(x))||^{\mathscr{M}}_{g,w}$. Hence, the formula $\exists y(y = \iota x R(x)) \supset R(\iota x R(x))$ is true in $\mathscr{M}$ at w under g. Since the model, the variable assignment and the world were arbitrary, it follows that $\models^1 \exists y(y = \iota x R(x)) \supset R(\iota x R(x))$.

□

Even if the formula ψ is valid, we can show that $\Box\psi$ is not. This can be demonstrated with the help of the following counterexample:

Example 4.2.4. *Given a variable a model $\mathscr{M} := \langle W, \succeq, D, I\rangle$ with*

$$W := \{w, v\}$$
$$\succeq := \{(w,w),(v,w),(v,v)\}$$
$$D_w := \{a,b\}, \quad D_v := \{a,b\}$$
$$I(R,w) := \{a\}, \;\; I(R,v) := \{b\}$$

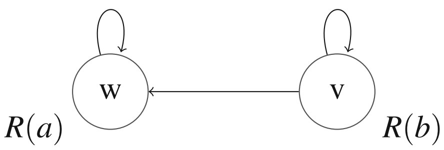

Figure 4.1: Counter-model rule **N**

Proof of Proposition 4.2.2. Let ψ be the formula $\exists y(y = \iota x R(x)) \supset R(\iota x R(x))$ from Proposition 4.2.3 and $\mathscr{M}$ as in Example 4.2.4. Although ψ is true at every world of $\mathscr{M}$ we are going to show that $w \notin ||\Box\psi||^{\mathscr{M}}_{g,w}$ for any variable assignment g.
Given a fixed variable assignment g we can define the following y-variant h of g at v:

$$h(x,v') := \begin{cases} a & \text{if } \; x = y \\ g(x,v') & \text{otherwise} \end{cases}$$

For this variable assignment h, it holds that $(I * h)_w(y) = a = (I * h)_w(\iota x R(x))$. Therefore $v \in ||\exists y(y = \iota x R(x))||^{\mathscr{M}}_{g,w}$ and since $(I * g)_w(\iota x R(x)) = a \notin I(R,v)$ it follows that $v \notin ||R(\iota x R(x))||^{\mathscr{M}}_{g,w}$, which lets us conclude $v \notin ||\exists y(y = \iota x R(x)) \supset R(\iota x R(x))||^{\mathscr{M}}_{g,w}$. In total we have $w \notin ||\Box\psi||^{\mathscr{M}}_{g,w}$ which means

$$\mathscr{M}, w \not\models^1_g \Box[(\exists y(y = \iota x R(x)) \supset R(\iota x R(x)))].$$

Since $\mathscr{M} \in \mathscr{U}_1$ the formula $\Box\psi$ is not valid in $\mathbf{F}_1^{\forall}$.

□

The reason for **N** not being a sound rule in $\mathbf{F}_1^{\forall}$ is that if a formula contains a definite description or a constant, it could be the case that this formula holds at every world as seen from the point of view of that world, but it does not have to be the case that the formula holds at another world from the point of view which is not the same world.
Let us use a natural language example of this phenomenon to give a better understanding to our interpretation of necessity. The importance is to keep the local interpretation of our terms in mind. If we interpret the formula $R(x)$ in the example above as "x is wearing a pink hat", then the validity of the formula $\exists y(y = \imath x R(x)) \supset R(\imath x R(x))$ in $\mathbf{F}_1^{\forall}$ can be interpreted as: "If the person wearing a pink hat exists, then they are wearing a pink hat", which is a statement whose validity is not to question. Nonetheless, the formula $\Box[\exists y(y = \imath x R(x)) \supset R(\imath x R(x))]$ represents the statement: It is necessary that if the person wearing a pink hat (fixed from the point of view of the world we are referencing from) exists, this person (fixed from the point of view of the world we are referencing from) is wearing a pink hat. This does not hold if the person changes hats from world to world.

We have just seen that in general the assumption of the validity of a formula is too weak to imply the validity of the necessity of this formula in $\mathbf{F}_1^{\forall}$. A strengthening of the antecedent is needed to create a sound rule in $\mathbf{F}_1^{\forall}$. The idea for the new inference rule **N*** is taken from Goble's paper [9] and it is defined as:

Definition 4.2.5. *If* $\vdash^1 \varphi^*$ *then* $\vdash^1 \Box\varphi$*, where* φ^* *is the result of replacing all constants and definite descriptions in* φ *with free variables not occurring in* φ*, see Definition 3.1.8.*

We want to prove that the rule **N*** does preserve validity in $\mathbf{F}_1^{\forall}$. In other words, we want to show that the rule **N*** is sound:

Theorem 4.2.6. *Given a well-formed formula* $\varphi \in WF$ *and a model* $\mathscr{M}$ *then*

$$\mathscr{M} \models^1 \varphi^* \quad \textit{implies} \quad \mathscr{M} \models^1 \Box\varphi.$$

This theorem can be proven with the help of two lemmas. The first lemma states that the validity of a formula of the form φ^* in a model implies the validity of $\Box\varphi^*$ in that model.[3] The second lemma states that the validity of a formula φ^* in a model implies the validity of φ in that model. The proofs of both lemmas are built on the fact that the variable assignments of a model $\mathscr{M}$ range over all possible elements of the whole domain of $\mathscr{M}$, see Definition 3.2.4.

[3]This means that given formula which does not contain constants or definite descriptions, the rule **N** does indeed preserve validity.

Lemma 4.2.7. *Given a well-formed formula $\varphi \in WF$ and a model $\mathcal{M}$ then*

$$\mathcal{M} \models^1 \varphi^* \Rightarrow \mathcal{M} \models^1 \Box\varphi^*.$$

Proof. Let φ be any fixed, well-formed formula and $\mathcal{M} = \langle W, \succeq, D, I\rangle$ a model. If for every world $w \in W$ and every variable assignment g of $\mathcal{M}$ it holds that $\mathcal{M}, w \models^1_g \varphi^*$, it follows that $w \in ||\varphi^*||^{\mathcal{M}}_{g,w}$ for every world $w \in W$ and every variable assignment g of $\mathcal{M}$. Now let us take two arbitrary but fixed worlds $v, w \in W$ and an arbitrary but fixed variable assignment g and define a new variable assignment $h : V \times W \to \mathbb{D}^+$ of $\mathcal{M}$ as:

$$h(x, v') := \begin{cases} g(x, w) & \text{if } v' = v \\ g(x, v') & \text{otherwise} \end{cases}$$

Since h copies the variables how g sees them at w to v ($\forall x \in V : h(x, v) = g(x, w)$) we get the equality $||\varphi^*||^{\mathcal{M}}_{h,v} = ||\varphi^*||^{\mathcal{M}}_{g,w}$. This gives us $v \in ||\varphi^*||^{\mathcal{M}}_{h,v} = ||\varphi^*||^{\mathcal{M}}_{g,w}$. Since v was arbitrary we can conclude $||\varphi^*||^{\mathcal{M}}_{g,w} = W$, which by definition is equivalent to $w \in ||\Box\varphi^*||^{\mathcal{M}}_{g,w}$ and to $\mathcal{M}, w \models^1_g \Box\varphi^*$. Because w was an arbitrary world and g was an arbitrary variable assignment of $\mathcal{M}$, we can further conclude $\mathcal{M} \models^1 \Box\varphi^*$. □

Lemma 4.2.8. *Given a well-formed formula $\varphi \in WF$ and a model $\mathcal{M}$ then*

$$\mathcal{M} \models^1 \varphi^* \Rightarrow \mathcal{M} \models^1 \varphi.$$

Proof. We are going to prove the second lemma by contraposition. Let us assume that there exists a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, a world $w \in W$ and a variable assignment g such that $\mathcal{M}, w \not\models^1_g \varphi$. Let $t_1, \ldots, t_n$ be all terms in φ which are replaced by the corresponding variables $x_1, \ldots, x_n$ in φ^* then for the variable assignment

$$h(x, v) := \begin{cases} (I * g)_w(t_i) & \text{if } (x, v) = (x_i, w) \text{ for } i \in \{1, \ldots, n\} \\ g(x, v) & \text{otherwise} \end{cases}$$

we have $\mathcal{M}, w \not\models^1_h \varphi^*$.

□

Theorem 4.2.6 can now be proven by putting Lemma 4.2.7 and Lemma 4.2.8 together:

Proof of Theorem 4.2.6. $\mathcal{M} \models^1 \varphi^* \Rightarrow \mathcal{M} \models^1 \Box\varphi^* \Leftrightarrow \mathcal{M} \models^1 (\Box\varphi)^* \Rightarrow \mathcal{M} \models^1 \Box\varphi.$ □

Theorem 4.2.1 together with Theorem 4.2.6 show us that if we only take into account the terms and formulas without constants or definite descriptions then all the axioms and rules of **F** are sound in $\mathbf{F}_1^\forall$. Therefore the $\mathbf{F}_1^\forall$ without constants or definite descriptions are in fact first-order dyadic deontic logic semantics of an extension of the system **F**.

4.3 The Axiom I in $\mathsf{F}_1^\forall$

This section is dedicated to proving soundness of the axiom **I** in $\mathbf{F}_1^\forall$. This means that given a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g, a world $w \in W$, two terms t and s and a formula φ as in Definition 3.1.6 we want to show

$$\mathcal{M}, w \models_g^1 t = s \supset (\varphi \leftrightarrow \varphi_{t \hookrightarrow s}).$$

To prove the soundness of the Axiom of Replacement in $\mathbf{F}_1^\forall$, we use the following theorem:

Theorem 4.3.1. *Given a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, two terms s and t, a variable assignment g, a world $w \in W$ and a formula $\varphi \in WF$ with $fv(t) \cap bv(\varphi) = \emptyset$ and $fv(s) \cap bv(\varphi) = \emptyset$ then $(I * g)_w(t) = (I * g)_w(s)$ implies:*

$$||\varphi_{t \hookrightarrow s}||_{g,w}^{\mathcal{M}} = ||\varphi||_{g,w}^{\mathcal{M}} \tag{4.1}$$

The proof of this theorem is given in three steps. First, we show that the set equivalence 4.1 holds for all atomic formulas, then for all formulas of depth zero and finally for all formulas. This is done by induction over the formula construction of all formulas of depth zero and then by induction over the depth of the formulas. We start by proving the set equivalence 4.1 for every atomic formula.

Lemma 4.3.2. *Given a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, two terms s and t, a variable assignment g, a world $w \in W$ and an atomic formula $A^n(t_1, \ldots, t_n)$, then $(I * g)_w(t) = (I * g)_w(s)$ implies*

$$||A^n(t_1, \ldots, t_n)_{t \hookrightarrow s}||_{g,w}^{\mathcal{M}} = ||A^n(t_1, \ldots, t_n)||_{g,w}^{\mathcal{M}}.$$

Proof. Since we are dealing with an atomic formula for every $i = 1, \ldots, n$ it holds that $t_i \in V \cup C$. Hence the only case where t can appear in $A^n(t_1, \ldots, t_n)$ is as a t_i, because all t_i do not contain further terms.
If for all $i = 1, \ldots, n$ it holds that $t_i \neq t$ then $A^n(t_1, \ldots, t_n)_{t \hookrightarrow s}$ and $A^n(t_1, \ldots, t_n)$ are the same formula and therefore $||A^n(t_1, \ldots, t_n)_{t \hookrightarrow s}||_{g,w}^{\mathcal{M}} = ||A^n(t_1, \ldots, t_n)||_{g,w}^{\mathcal{M}}$ by default.
In the other case there exists (at least one) $i = 1, \ldots, n$ such that $t_i = t$. Without loss of generality let us assume that $t_1 = t$, then $A^n(t_1, \ldots, t_n)_{t \hookrightarrow s} = A^n(s, \ldots, t_n)$. By definition $v \in ||A^n(t_1, \ldots, t_n)||_{g,w}^{\mathcal{M}}$ is equivalent to $\langle (I * g)_w(t_1), \ldots, (I * g)_w(t_n)\rangle \in I(A^n, v)$ and because of the equalities $(I * g)_w(t_1) = (I * g)_w(t) = (I * g)_w(s)$ this is also equivalent to $\langle (I * g)_w(s), (I * g)_w(t_2), \ldots, (I * g)_w(t_n)\rangle \in I(A^n, v)$. In conclusion we get $||A^n(t_1, \ldots, t_n)_{t \hookrightarrow s}||_{g,w}^{\mathcal{M}} = ||A^n(t_1, \ldots, t_n)||_{g,w}^{\mathcal{M}}$.

□

Next, we will prove the equivalence 4.1 for every formula of the form $t_1 = t_2$ of depth 0.

Lemma 4.3.3. *Given a model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, two terms s,t, a variable assignment g, a world $w \in W$, two terms $t_1, t_2 \in V \cup C$, then $(I * g)_w(t) = (I * g)_w(s)$ implies*

$$||(t_1 = t_2)_{t \hookrightarrow s}||^{\mathscr{M}}_{g,w} = ||t_1 = t_2||^{\mathscr{M}}_{g,w}.$$

Proof. Similarly as in the proof of the lemma above the only non-trivial case we have to consider is if either t_1 or t_2 is the term t, w.l.o.g $t = t_1$. Then the claim follows from the following equivalences: $v \in ||t_1 = t_2||^{\mathscr{M}}_{g,w} \Leftrightarrow (I * g)_w(t_1) = (I * g)_w(t_2) \Leftrightarrow (I * g)_w(s) = (I * g)_w(t_2) \Leftrightarrow v \in ||(t_1 = t_2)_{t \hookrightarrow s}||^{\mathscr{M}}_{g,w}$.

□

Remark 4.3.4. *Lemma 4.3.2 and 4.3.3 have been proven for any two terms, any variable assignment and any world.*

To finish the first part of the proof of Theorem 4.3.1, we will prove the equivalence 4.1 for every formula $\varphi \in WF$ with $de(\varphi) = 0$ by induction over the formula construction without definite descriptions.

Lemma 4.3.5. *Given a model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, two terms s and t, a variable assignment g, a world $w \in W$ and a formula $\varphi \in WF$ with $de(\varphi) = 0$, $fv(t) \cap bv(\varphi) = \emptyset$ and $fv(s) \cap bv(\varphi) = \emptyset$, then $(I * g)_w(t) = (I * g)_w(s)$ implies*

$$||\varphi_{t \hookrightarrow s}||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{g,w}.$$

Proof. If φ is an atomic formula or of the form $t_1 = t_2$, we have already seen a proof in Lemma 4.3.2 and Lemma 4.3.3. Now we work by induction over the formula construction of all formulas of depth zero, given in Definition 3.1.4.
Let us therefore assume that given two formulas ψ and χ of depth 0, the set equivalence 4.1 already holds for any variable assignment, world and any two terms fulfilling $fv(t) \cap bv(\psi) = \emptyset$, $fv(s) \cap bv(\psi) = \emptyset$, $fv(t) \cap bv(\chi) = \emptyset$ and $fv(s) \cap bv(\chi) = \emptyset$. This means if $w' \in W$ is any world and h is any variable assignment, then $(I * h)_{w'}(t) = (I * h)_{w'}(s)$ implies $||\psi_{t \hookrightarrow s}||^{\mathscr{M}}_{h,w'} = ||\psi||^{\mathscr{M}}_{h,w'}$ and $||\chi_{t \hookrightarrow s}||^{\mathscr{M}}_{h,w'} = ||\chi||^{\mathscr{M}}_{h,w'}$. Now we consider every case of how a well-formed φ of depth zero can be constructed from ψ and χ. We show that the set equivalence 4.1 holds for φ given the restrictions concerning the terms t and s:

$\varphi = \neg\psi$:

$$||\varphi_{t \hookrightarrow s}||^{\mathscr{M}}_{g,w} = W / ||\psi_{t \hookrightarrow s}||^{\mathscr{M}}_{g,w} = W / ||\psi||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{g,w}$$

$\varphi = \psi \wedge \chi$:

$$||\varphi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w} = ||\psi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w} \cap ||\chi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w} = ||\psi||^{\mathscr{M}}_{g,w} \cap ||\chi||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{g,w}$$

$\varphi = \Box\psi$:

$$\forall v \in W : (v \in ||\varphi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w} \Leftrightarrow W = ||\psi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w} = ||\psi||^{\mathscr{M}}_{g,w} \Leftrightarrow v \in ||\varphi||^{\mathscr{M}}_{g,w})$$

$\varphi = \forall x\psi$:
Since $fv(t) \cap bv(\varphi) = \emptyset = fv(s) \cap bv(\varphi)$ it follows that $x \notin fv(t)$ and $x \notin fv(s)$. Hence given an x-variant h of g (at any world v) we get $(I*g)_w(t) = (I*h)_w(t)$ and $(I*g)_w(s) = (I*h)_w(s)$ which implies $(I*h)_w(t) = (I*h)_w(s)$. A world v being an element of $||\varphi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w}$ is by definition equivalent to $v \in ||\psi_{t\hookrightarrow s}||^{\mathscr{M}}_{h,w}$ for all x-variants h of g at v. By use of the induction hypothesis we can conclude that this is equivalent to $v \in ||\psi||^{\mathscr{M}}_{h,w}$ for all x-variants h of g at v, which is the definition for $v \in ||\varphi||^{\mathscr{M}}_{g,w}$.

$\varphi = \bigcirc(\psi/\chi)$:
It follows directly from the definition of the best worlds that $||\chi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w} = ||\chi||^{\mathscr{M}}_{g,w}$ implies $best(||\chi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w}) = best(||\chi||^{\mathscr{M}}_{g,w})$.[4] Hence we can conclude that $best(||\chi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w}) \subseteq ||\psi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w}$ and $best(||\chi||^{\mathscr{M}}_{g,w}) \subseteq ||\psi||^{\mathscr{M}}_{g,w}$ are equivalent, which leads us to $||\varphi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{g,w}$.

□

Now finally, we want to prove Theorem 4.3.1. This will be achieved by induction over the depth of the formulas.

Proof of Theorem 4.3.1. The start of this induction proof (for all formulas $\varphi \in WF$ with $de(\varphi) = 0$) has already been done in Lemma 4.3.5.
Given an arbitrary but fixed $m \in \mathbb{N}$, we assume the set equivalence 4.1 holds for every formula of depth lower or equal to m. This means if $w' \in W$ is a world, h a variable assignment and ψ a formula with $de(\psi) \leq m$ then $(I*h)_{w'}(t) = (I*h)_{w'}(s)$ implies $||\psi_{t\hookrightarrow s}||^{\mathscr{M}}_{h,w'} = ||\psi||^{\mathscr{M}}_{h,w'}$. We are now using this assumption to prove the set equivalence 4.1 for any formula φ of depth lower or equal to $m+1$.
Let us start with formulas of the form $\varphi = R^n(t_1,...,t_n)$, where R^n is an n-place predicate symbol. If for a fixed $i = 1,...,n$ the term t_i is of the form $\imath x\psi$, for a variable x and a formula ψ, then by definition $de(\psi) \leq m$. Since $x \notin fv(t) \cup fv(s)$ the equivalence $(I*g)_w(t) = (I*g)_w(t)$ implies $(I*h)_w(t) = (I*h)_w(t)$ for every x-variant h

[4] $v \in best(||\chi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w}) \Leftrightarrow v \in ||\chi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w} \wedge [\forall v' \in W(v' \in ||\chi_{t\hookrightarrow s}||^{\mathscr{M}}_{g,w} \Rightarrow v \succeq v')]$
$\Leftrightarrow v \in |\chi||^{\mathscr{M}}_{g,w} \wedge [\forall v' \in W(v' \in ||\chi||^{\mathscr{M}}_{g,w} \Rightarrow v \succeq v')] \Leftrightarrow v \in best(||\chi||^{\mathscr{M}}_{g,w})$

of g (at any world). Hence by the induction hypothesis and because $bv(\psi) \subseteq bv(\varphi)$ we have $||\psi||^{\mathcal{M}}_{h,w} = ||\psi_{t \hookrightarrow s}||^{\mathcal{M}}_{h,w}$ for every x-variant h of g (at any world). Using this equivalence and the definition of the evaluation of a definite description $\imath x \psi$ we can conclude $(I * g)_w(\imath x \psi) = (I * g)_w(\imath x \psi_{t \hookrightarrow s})$. On the other hand if for a $i = 1, \ldots, n$ the term t_i is not of the form $\imath x \psi$ then $t_i \in V \cup C$. For this type of term we already know that $(I * g)_w(t_i) = (I * g)_w((t_i)_{t \hookrightarrow s})$ holds (see proof of Lemma 4.3.2). We can conclude that for all $i = 1, \ldots, n$, it is true that $(I * g)_w(t_i) = (I * g)_w((t_i)_{t \hookrightarrow s})$. Hence we arrive at $||R^n(t_1, \ldots, t_n)_{t \hookrightarrow s}||^{\mathcal{M}}_{g,w} = ||R^n(t_1, \ldots, t_n)||^{\mathcal{M}}_{g,w}$.
With the same argument we can also infer that for any two terms t_1, t_2 with $de(t_1) \leq m+1$ and $de(t_2) \leq m+1$ the equivalence $||(t_1 = t_2)_{t \hookrightarrow s}||^{\mathcal{M}}_{g,w} = ||t_1 = t_2||^{\mathcal{M}}_{g,w}$ holds. Constructing well-formed formulas with $\neg, \wedge, \forall, \Box$ and $\bigcirc$ out of formulas with depth lower or equal to $m+1$ does result in a formula of depth lower or equal to $m+1$. For that reason and because of the fact that we did not use the property $de(\varphi) = 0$ in the induction proof of Lemma 4.3.5 we can conclude that $(I * g)_w(t) = (I * g)_w(s)$ implies $||\varphi_{t \hookrightarrow s}||^{\mathcal{M}}_{g,w} = ||\varphi||^{\mathcal{M}}_{g,w}$ for any formula with $de(\varphi) \leq m+1$.

□

With the help of Theorem 4.3.1 the following corollary follows directly:

Corollary 4.3.6. *Given a model $\mathcal{M} = \langle W, \succeq, D, I \rangle$, a variable assignment g, a world $w \in W$, two terms t, s and a formula $\varphi \in WF$ with $bv(\varphi) \cap fv(t) = \emptyset$ and $bv(\varphi) \cap fv(s) = \emptyset$ then*

$$\mathcal{M}, w \models^1_g t = s \supset (\varphi \leftrightarrow \varphi_{t \hookrightarrow s}).$$

In particular: Given two terms t, s and a formula $\varphi \in WF$ with $bv(\varphi) \cap fv(t) = \emptyset$ and $bv(\varphi) \cap fv(s) = \emptyset$ then

$$\models^1 t = s \supset (\varphi \leftrightarrow \varphi_{t \hookrightarrow s}).$$

4.4 The Axiom UI in $\mathbf{F}^{\forall}_1$

Here we show the axiom **UI** is sound in $\mathbf{F}^{\forall}_1$. This means that given a model $\mathcal{M} = \langle W, \succeq, D, I \rangle$, a variable assignment g, a world $w \in W$, two terms t and s and a formula φ as in Definition 3.1.6 we want to show

$$\mathcal{M}, w \models^1_g E(t) \supset (\forall x \varphi \supset \varphi_{x \Rightarrow t}).$$

We are going to show the soundness of **UI** in $\mathbf{F}^{\forall}_1$ with the help of the following theorem:

Theorem 4.4.1. *Given a formula well-formed $\varphi \in WF$, a model $\mathscr{M}$, a variable assignment g, a world $w \in W$, a term t with $bv(\varphi) \cap fv(t) = \emptyset$ and $d := (I * g)_w(t)$ then the following equation holds:*

$$||\varphi_{x \Rightarrow t}||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{g_{x \Rightarrow d},w} \tag{4.2}$$

The proof of this theorem is again done by induction over the formula construction. We start by proving this theorem for every atomic formula.

Lemma 4.4.2. *Given a atomic formula $A^n(t_1,...,t_n)$, a model $\mathscr{M}$, a term t and $d := (I * g)_w(t)$ then the following equation holds*

$$||A^n(t_1,...,t_n)_{x \Rightarrow t}||^{\mathscr{M}}_{g,w} = ||A^n(t_1,...,t_n)||^{\mathscr{M}}_{g_{x \Rightarrow d},w}$$

Proof. If $v \in ||A^n(t_1,...,t_n)||^{\mathscr{M}}_{g,w}$ then by definition we have $\langle (I * g)_w(t_1),...,(I * g)_w(t_n) \rangle \in I(A^n, v)$. Since we are dealing with an atomic formula for every $i = 1,...n$ we have $t_i \in V \cup C$. Hence the only case where x can appear in $A^n(t_1,...,t_n)$ is as a t_i, because all t_i do not contain further terms.
If for all $i = 1,...,n$ it holds that $t_i \neq x$ then $A^n(t_1,...,t_n)_{x \Rightarrow t}$ and $A^n(t_1,...,t_n)$ are the same formula and the different evaluation of x does not affect the sets, therefore $||A^n(t_1,...,t_n)_{x \Rightarrow t}||^{\mathscr{M}}_{g,w} = ||A^n(t_1,...,t_n)||^{\mathscr{M}}_{g_{x \Rightarrow t},w}$. In the other case that there exists (at least one) $i = 1,...,n$ such that $t_i = x$. In this case $v' \in ||A^n(t_1,...,t_n)_{x \Rightarrow t}||^{\mathscr{M}}_{g,w}$ holds if and only if $\langle (I * g)_w(t_1),...,(I * g)_w(t_n) \rangle \in I(A^n, v')$ holds, where all t_i with $t_i = x$ have been replaced by t. Now since $(I * g)_w(t) = (I * g_{x \Rightarrow d})_w(x)$ this is equivalent to $v' \in ||A^n(t_1,...,t_n)||^{\mathscr{M}}_{g_{x \Rightarrow d},w}$.
□

Next, we will prove Theorem 4.4.1 for every formula of the form $t_1 = t_2$ of depth 0.

Lemma 4.4.3. *Given a model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, a term t, a variable assignment g, a world $w \in W$, two terms $t_1, t_2 \in V \cup C$ and $d := (I * g)_w(t)$ then*

$$||(t_1 = t_2)_{x \Rightarrow t}||^{\mathscr{M}}_{g,w} = ||t_1 = t_2||^{\mathscr{M}}_{g_{x \Rightarrow d},w}.$$

Proof. As in the proof of the lemma above the only non-trivial case we have to consider is if either t_1 or t_2 is the variable x, w.l.o.g $x = t_1$. In this case for all $v \in W$ the following equivalences hold $v \in ||(t_1 = t_2)_{x \Rightarrow t}||^{\mathscr{M}}_{g,w} \Leftrightarrow (I * g)_w(t) = (I * g)_w(t_2) \Leftrightarrow v \in ||t_1 = t_2||^{\mathscr{M}}_{g_{x \Rightarrow d},w}$, which proves the lemma. □

Remark 4.4.4. *Lemma 4.4.2 and 4.4.3 have been proven for any term, any variable assignment and any world.*

To finish the first part of the proof of Theorem 4.4.1, we will prove the set equivalence 4.2 for every formula $\varphi \in WF$ with $de(\varphi) = 0$ by induction over the formula construction of all formulas of depth zero.

Lemma 4.4.5. *Given a model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, a term t, a variable assignment g, a world $w \in W$ and a formula $\varphi \in WF$ with $de(\varphi) = 0$, $fv(t) \cap bv(\varphi) = \emptyset$ and $d := (I * g)_w(t)$ then*

$$||\varphi_{x\Rightarrow t}||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{g_{x\Rightarrow d},w}.$$

Proof. If φ is an atomic formula or of the form $t_1 = t_2$ of depth 0, we have already seen a proof in Lemmas 4.4.2 and 4.4.3. Now we work by induction over the formula construction given in Definition 3.1.4.
Let us therefore assume that given two formulas ψ and χ of depth 0, the set equivalence 4.2 already holds for any variable assignment, world and any term t fulfilling $fv(t) \cap bv(\psi) = \emptyset$, and $fv(t) \cap bv(\chi) = \emptyset$. This means if $w' \in W$ is a world and h a variable assignment, then $||\psi_{x\Rightarrow t}||^{\mathscr{M}}_{h,w'} = ||\psi||^{\mathscr{M}}_{h_{x\Rightarrow d},w'}$ and $||\chi_{x\Rightarrow t}||^{\mathscr{M}}_{h,w'} = ||\chi||^{\mathscr{M}}_{h_{x\Rightarrow d},w'}$.
Now we consider every case of how a well-formed φ of depth zero can be built. The cases $\varphi = \neg\psi$, $\varphi = \psi \wedge \chi$, $\varphi = \Box\psi$ and $\varphi = \bigcirc(\psi/\chi)$ can be proven in the same way as in Lemma 4.3.5. The only step that needs some extra explanation is when $\varphi = \forall y \psi$. Starting with an arbitrary world $v \in W$ such that $v \in ||\varphi_{x\Rightarrow t}||^{\mathscr{M}}_{g,w}$. This is equivalent to $v \in ||\psi_{x\Rightarrow t}||^{\mathscr{M}}_{g',w}$ for all y-variants g' of g at v by Definition 3.2.9. By use of the induction assumption together with the fact that $y \notin bv(t)$, we can conclude that this is equivalent to $v \in ||\psi||^{\mathscr{M}}_{g',w}$ for all y-variants g' of $g_{x\Rightarrow d}$ at v. Therefore we arrive at $v \in ||\forall y \psi||^{\mathscr{M}}_{g_{x\Rightarrow d},w}$.

□

Now finally, we want to prove Theorem 4.4.1. We will do this by induction over the depth of φ.

Proof of Theorem 4.4.1. Given a number $m \in \mathbb{N}$ we assume that Theorem 4.4.1 holds for every formula φ with $de(\varphi) \leq m$. Given a formula ψ of depth lower or equal to $m+1$ and with $bv(\psi) \cap fv(t) = \emptyset$, if $\imath y \varphi$ appears as a term in the formula ψ it follows that $de(\varphi) \leq n$ and $y \notin fv(t)$. The induction assumption then lets us derive that $(I * g)_w(\imath y \varphi_{x\Rightarrow t}) = (I * g_{x\Rightarrow t})_w(\imath y \varphi)$. By the same induction steps as in the proof of Theorem 4.3.1 we can conclude that the equivalence $||\varphi_{x\Rightarrow t}||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{g_{x\Rightarrow d},w}$ holds.

□

As a corollary of Theorem 4.4.1 follows the soundness of **UI** in $\mathbf{F}_1^{\forall}$:

Corollary 4.4.6. *Given a model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, a variable assignment g, a world $w \in W$, a variable $x \in V$, a term t, a formula $\varphi \in WF$ with $bv(\varphi) \cap fv(t) = \emptyset$ then*

$$\mathcal{M}, w \models_g^1 E(t) \supset (\forall x\varphi \supset \varphi_{x\Rightarrow t}).$$

In particular: Given a term a variable $x \in V$, a term t and a formula $\varphi \in WF$ with $bv(\varphi) \cap fv(t) = \emptyset$ then

$$\models^1 E(t) \supset (\forall x\varphi \supset \varphi_{x\Rightarrow t}).$$

Proof. Since $w \in ||E(t)||_{g,w}^{\mathcal{M}}$ is equivalent to $(I * g)_w(t) \in D_w$ the assumption $\mathcal{M}, w \models_g^1 E(t)$ makes the variable assignment $g_{x\Rightarrow d}$ an x-variant of g at w. This fact, together with Theorem 4.4.1 proves the corollary.

□

4.5 Soundness

We conclude this chapter by showing that the Hilbert axiomatisation $\mathbf{HF}_1^\forall$ is strongly sound in $\mathbf{F}_1^\forall$. We start by showing that it is weakly sound.

Theorem 4.5.1. *Given a well-formed formula $\varphi \in WF$, a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, a world $w \in W$ and a variable assignment g of $\mathcal{M}$:*

$$\vdash^1 \varphi \Rightarrow \mathcal{M}, w \models_g^1 \varphi$$

More specific:

$$\vdash^1 \varphi \Rightarrow \models^1 \varphi$$

Proof. As mentioned in the beginning of this chapter, a soundness proof is achieved by showing that every instance of axioms are valid and that every rule preserves validity. We start with the axioms of $\mathbf{HF}_1^\forall$:

- All axioms of system **F**.

  This has been shown in Theorem 4.2.1.

- $t = s \supset (\varphi \leftrightarrow \varphi_{t\hookrightarrow s})$

  This has been shown in Corollary 4.3.6.

- $E(t) \supset (\forall x\varphi \supset \varphi_{x\Rightarrow t})$

  This has been shown in Corollary 4.4.6.

- $\exists x \exists y(x = y)$
  This follows from a property in Definition 3.2.3 of our models. Namely, for every world w of every model $\mathscr{M}$ we have, the domain D_w is not empty.

- $t = t$
  This follows directly from Definition 3.2.9.

- $t \neq s \supset \Box t \neq s$
  This follows from the following equivalences: $w \in ||t \neq s||^{\mathscr{M}}_{g,w} \Leftrightarrow$ $w \in W \backslash ||t = s||^{\mathscr{M}}_{g,w} \Leftrightarrow ||t = s||^{\mathscr{M}}_{g,w} = \emptyset \Leftrightarrow ||t \neq s||^{\mathscr{M}}_{g,w} = W \Leftrightarrow w \in ||\Box t \neq s||^{\mathscr{M}}_{g,w}$.

- $\forall y((\forall x(\varphi \leftrightarrow x = y)) \supset y = \imath x\varphi)$
  Let h be a y-variant of g at w then $w \in ||\forall x(\varphi \leftrightarrow x = y)||^{\mathscr{M}}_{h,w}$ is equivalent to $w \in ||\varphi||^{\mathscr{M}}_{j,w} \Leftrightarrow w \in ||x = y||^{\mathscr{M}}_{j,w}$ for all x-variants j of h at w. This is furthermore equivalent to $w \in ||\varphi||^{\mathscr{M}}_{j,w} \Leftrightarrow j(x,w) = j(y,w) = h(y,w)$ for all x-variants j of h at w. This means that $w \in ||\varphi||^{\mathscr{M}}_{j,w}$ holds for exactly one x-variant j of h at w and for this j the equivalence $j(x,w) = h(y,w)$ holds. In other words this means that $w \in ||y = \imath x\varphi||^{\mathscr{M}}_{h,w}$. Since h was an arbitrary y-variant of g at w we can conclude $\mathscr{M}, w \models^1_g \forall y((\forall x(\varphi \leftrightarrow x = y)) \supset y = \imath x\varphi)$.

- $E(\imath x\varphi) \supset \exists! x\varphi$
  $\mathscr{M}, w \models^1_g E(\imath x\varphi)$ is equivalent to $(I * g)_w(\imath x\varphi) \in D_w$. As a result $(I * g)_w(\imath x\varphi) \neq \{\mathbb{D}\}$ which means that there is a unique x-variant h of g at w such that $w \in ||\varphi||^{\mathscr{M}}_{h,w}$.

- $\forall x(E(x) \supset \varphi) \supset \forall x\varphi$
  For $w \in ||\forall x(E(x) \supset \varphi)||^{\mathscr{M}}_{g,w}$ to be true we need for all x variants h of g at w for either $w \notin ||E(x)||^{\mathscr{M}}_{h,w}$ or $w \in ||\varphi||^{\mathscr{M}}_{h,w}$ to hold. Since $w \notin ||E(x)||^{\mathscr{M}}_{h,w}$ is equivalent to $h(x,w) \notin D_w$ the first case cannot be true. We can conclude that $w \in ||\forall x(E(x) \supset \varphi)||^{\mathscr{M}}_{g,w}$ implies that for all x variants h of g at w we have $w \in ||\varphi||^{\mathscr{M}}_{h,w}$.

- $(\forall x\varphi \wedge \forall x\psi) \leftrightarrow \forall x(\varphi \wedge \psi)$
  $v \in ||\forall x\varphi \wedge \forall x\psi||^{\mathscr{M}}_{g,w}$ means that for every x-variant h of g at v it holds that $v \in ||\varphi||^{\mathscr{M}}_{h,w}$ and that for every x-variant h of g at v we have $v \in ||\psi||^{\mathscr{M}}_{h,w}$. This is the same as saying that for every x-variant h of g at v we have $v \in ||\varphi||^{\mathscr{M}}_{h,w} \cap ||\psi||^{\mathscr{M}}_{h,w}$ which is equivalent to $v \in ||\forall x(\varphi \wedge \psi)||^{\mathscr{M}}_{g,w}$.

Now we show that every rule of $\mathbf{HF}^{\forall}_1$ is sound in $\mathbf{F}^{\forall}_1$:

- The rule **R1**

  We are going to prove this by contraposition. Suppose there exists a model $\mathscr{M}$, a world w, a well-formed formula φ and a variable assignment g such that $\mathscr{M}, w \not\models^1_g \neg\varphi$. Then $w \notin ||\neg\varphi||^{\mathscr{M}}_{g,w}$ which is equivalent to $w \in ||\varphi||^{\mathscr{M}}_{g,w}$. Using the term t in the formula $\varphi \supset t \neq x$, we can define the following variable assignment:

$$h(z,v) := \begin{cases} (I * g)_w(t) & \text{if } (z,v) = (x,w) \\ g(z,v) & \text{otherwise} \end{cases}$$

  By using the fact that $x \notin free(\varphi)$ it follows that $||\varphi||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{h,w}$ and $w \in ||t = x||^{\mathscr{M}}_{g,w}$. In total this gives us $w \notin ||\varphi \supset t \neq x||^{\mathscr{M}}_{g,w}$.

- The rule **MP**

  This has been shown in Theorem 4.2.1.

- The rule **N***

  This has been shown in Theorem 4.2.6.

- The rule **IU**

  By contraposition. Suppose there exists a model $\mathscr{M}$, a world w, two well-formed formulas φ and ψ and a variable assignment g such that $\mathscr{M}, w \not\models^1_g \varphi \supset \forall x\psi$. Then $w \notin ||\varphi \supset \forall x\psi||^{\mathscr{M}}_{g,w}$ which is equivalent to $w \in ||\varphi||^{\mathscr{M}}_{g,w}$ and $w \notin ||\forall x\psi||^{\mathscr{M}}_{g,w}$. Hence we can find an x-variant of g at w such that $w \notin ||\psi||^{\mathscr{M}}_{h,w}$. If $x \notin free(\varphi)$, it holds that $||\varphi||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{g,h}$. We have therefore found a world w in and a variable assignment h of $\mathscr{M}$ such that $w \notin ||\varphi \supset \psi||^{\mathscr{M}}_{h,w}$ which means $\mathscr{M} \not\models^1_g \varphi \supset \psi$.

- The rule **IUB**

  By contraposition. Suppose there exists a model $\mathscr{M}$, a world w, two well-formed formulas φ and ψ and a variable assignment g such that $\mathscr{M}, w \not\models^1_g \varphi \supset \Box\forall x\psi$. Then $w \notin ||\varphi \supset \Box\forall x\psi||^{\mathscr{M}}_{g,w}$ which is equivalent to $w \in ||\varphi||^{\mathscr{M}}_{g,w}$ and $W \neq ||\forall x\psi||^{\mathscr{M}}_{g,w}$. Hence we can find a world v and an x-variant of g at v such that $v \notin ||\psi||^{\mathscr{M}}_{h,w}$ which implies $w \notin ||\Box\psi||^{\mathscr{M}}_{h,w}$. If $x \notin free(\varphi)$, it holds that $||\varphi||^{\mathscr{M}}_{g,w} = ||\varphi||^{\mathscr{M}}_{h,w}$. We have therefore found a world w and a variable assignment h of $\mathscr{M}$ such that $w \notin ||\varphi \supset \Box\psi||^{\mathscr{M}}_{h,w}$ which means $\mathscr{M} \not\models^1_h \varphi \supset \Box\psi$.

□

Theorem 4.5.2 (Strong soundness)**.** $\boldsymbol{HF}_1^{\forall}$ *is strongly sound in* $\boldsymbol{F}_1^{\forall}$.

Proof. Let us fix a well-formed formula $\varphi \in WF$ and a set of well-formed formulas $\Gamma \subseteq WF$. We show that $\Gamma \vdash^1 \varphi$ implies $\Gamma \models^1 \varphi$. Let $\mathscr{M} = \langle W, \succeq, D, I \rangle$ be a model, $w \in W$ a world and g a variable assignment of $\mathscr{M}$ such that $\forall \psi \in \Gamma : \mathscr{M}, w \models_g^1 \psi$. Let us assume that $\Gamma \vdash^1 \varphi$ holds. This means there exists a sequence of formulas $\varphi_1, \ldots, \varphi_n$ as in Definition 4.1.4. We show that for every formula φ_i of this sequence $\mathscr{M}, w \models_g^1 \varphi_i$ holds by induction over $i \in \{1, \ldots, n\}$. For φ_1 it must either hold that $\vdash^1 \varphi_1$ or $\varphi_1 \in \Gamma$. In the first case $\mathscr{M}, w \models_g^1 \varphi_1$ follows from weak soundness. In the second case $\mathscr{M}, w \models_g^1 \varphi_1$ follows from the assumption. Now let φ_i be a fixed element of the sequence such that for all $j < i$ $\mathscr{M}, w \models_g^1 \varphi_j$ holds. If $\vdash^1 \varphi_i$ or $\varphi_i \in \Gamma$ then $\mathscr{M}, w \models_g^1 \varphi_i$ follows again from weak soundness and assumption respectively. If $\varphi_i \in \Gamma$ is derived by modus pones, then we can find two formulas φ_j and φ_k in the sequence with $j, k < i$ such that $\varphi_j = (\varphi_k \supset \varphi_i)$. Since $\mathscr{M}, w \models_g^1 \varphi_k \supset \varphi_i$ and $\mathscr{M}, w \models_g^1 \varphi_k$ hold by the induction assumption we can conclude $\mathscr{M}, w \models_g^1 \varphi_i$. Finally, since $\varphi_n = \varphi$ we get $\mathscr{M}, w \models_g^1 \varphi$. □

5 The System $\mathsf{F}_2^{\forall}$

Here we explore the other possible interpretation of the $\Box$ operator mentioned at the end of Chapter 2. Therefore we want $\Box$ to now be a global operator, which means $\Box\varphi$ being true at one world under a variable assignment implies $\Box\varphi$ being true at all worlds under that variable assignment. The main goal of this chapter is to show that it is possible to create a first-order dyadic deontic logic system with an extensional deontic operator and a global non-extensional modal operator, in which most, but not all, of the axioms of **F** are sound. To achieve this, we define semantics of a second system, which we call $\mathbf{F}_2^{\forall}$. We also revisit the "Switching seats" paradox and show that in $\mathbf{F}_2^{\forall}$, we can use the modal operator to express the difference between a local and global obligation. At the end of this chapter we introduce a Hilbert axiomatisation of $\mathbf{F}_2^{\forall}$, which we call $\mathbf{HF}_2^{\forall}$ and give a soundness proof.

5.1 Truth and Validity

We start by defining sets of the form $[\varphi]_{g,w}^{\mathscr{M}}$ for $\mathbf{F}_2^{\forall}$. Those sets will serve the same function as the sets $||\varphi||_{g,w}^{\mathscr{M}}$ of $\mathbf{F}_1^{\forall}$, see Definition 3.2.8. The sets $||\varphi||_{g,w}^{\mathscr{M}}$ and $[\varphi]_{g,w}^{\mathscr{M}}$ are equal if φ does not contain the modal operator $\Box$. Therefore the semantics of $\mathbf{F}_2^{\forall}$ will only differ from the semantics of $\mathbf{F}_1^{\forall}$ in the evaluation of the modal operator $\Box$. In $\mathbf{F}_2^{\forall}$, the modal operator $\Box$ is a global and non-extensional operator. Therefore in this system, the axioms **I** and **UI** will only be sound for every well-formed formula not containing the operator $\Box$. We have the same intuitive understanding of $v \in [\varphi]_{g,w}^{\mathscr{M}}$ as for $v \in ||\varphi||_{g,w}^{\mathscr{M}}$, namely that φ holds true at v for a person living at w. Hence in $\mathbf{F}_2^{\forall}$, $[\varphi]_{g,w}^{\mathscr{M}} = W$, can also be seen as a φ being true at every world for someone who refers to every object the way they are defined at world w. The important difference is that in $\mathbf{F}_2^{\forall}$ we define φ being necessarily true at a world w as $[\varphi]_{g,v}^{\mathscr{M}} = W$ for all worlds $v \in W$. This means that in $\mathbf{F}_2^{\forall}$ the interpretation of the $\Box$ operator is stronger in $\mathbf{F}_1^{\forall}$.
We are again going to simultaneously define a function I_w^g recursively, such that for every term t we get $I_w^g(t) \in \mathbb{D} \cup \{\mathbb{D}\}$. The motivation for $I_w^g(t) = p$ is that a person living at the world w interprets the term t as the object p, but in this case, in $\mathbf{F}_2^{\forall}$.

Definition 5.1.1. *Let $\mathcal{M} = \langle W, \succeq, D, I \rangle$ be a model, g a variable assignment of $\mathcal{M}$, t a term, φ a well-formed formula and $w \in W$ a world:*

- *If $t = x \in V$ then $I_w^g(t) := g(x,w)$*
- *If $t = c \in C$ then $I_w^g(t) := I(c,w)$*
- *If $R^n \in \mathbb{P}$ and $t_1,\ldots,t_n$ are terms, then*

$$[R^n(t_1,\ldots,t_n)]_{g,w}^{\mathcal{M}} := \{v \in W : \langle I_w^g(t_1),\ldots,I_w^g(t_n)\rangle \in I(R^n,v)\}$$

- *If t_1 and t_2 are terms, then*

$$[t_1 = t_2]_{g,w}^{\mathcal{M}} := \{v \in W : I_w^g(t_1) = I_w^g(t_2)\}$$

- *If $\varphi = \neg\psi$, then*

$$[\varphi]_{g,w}^{\mathcal{M}} := W \setminus [\psi]_{g,w}^{\mathcal{M}}$$

- *If $\varphi = \psi \wedge \chi$, then*

$$[\varphi]_{g,w}^{\mathcal{M}} := [\psi]_{g,w}^{\mathcal{M}} \cap [\chi]_{g,w}^{\mathcal{M}}$$

- *If $\varphi = \forall x\psi$, then*

$$[\varphi]_{g,w}^{\mathcal{M}} := \{v \in W : v \in [\psi]_{h,w}^{\mathcal{M}} \text{ for all } x\text{-variants } h \text{ of } g \text{ at } v\}$$

- *If $\varphi = \Box\psi$, then*

$$[\varphi]_{g,w}^{\mathcal{M}} := \{v \in W : \forall k \in W \;\; [\psi]_{g,k}^{\mathcal{M}} = W\}$$

- *If $\varphi = \bigcirc(\psi/\chi)$, then*

$$[\varphi]_{g,w}^{\mathcal{M}} := \{v \in W : best([\chi]_{g,w}^{\mathcal{M}}) \subseteq [\psi]_{g,w}^{\mathcal{M}}\}$$

- *If $t = \iota x\varphi$, then*

$$I_w^g(t) := \begin{cases} h(x,w) & \text{if } h \text{ is the } \textbf{unique } x\text{-variant of } g \text{ at } w \\ & \text{such that } w \in [\varphi]_{h,w}^{\mathcal{M}} \\ \mathbb{D} & \text{otherwise} \end{cases}$$

Definition 5.1.2. *We say that a model $\mathcal{M} = \langle W, \succeq, D, I \rangle$ fulfils the limitedness property of $\boldsymbol{F}_2^{\forall}$ if for every formula φ, variable assignment g and world $w \in W$ we have*

$$[\varphi]^{\mathcal{M}}_{g,w} \neq \emptyset \Rightarrow best([\varphi]^{\mathcal{M}}_{g,w}) \neq \emptyset.$$

We define $\mathcal{U}_2$ as the class of all models which are reflexive and fulfil the limitedness property of $\boldsymbol{F}_2^{\forall}$.

The definitions for truth, validity and semantic entailment in $\mathbf{F}_2^{\forall}$ (in symbols: $\models^2$) are as in $\mathbf{F}_1^{\forall}$, by switching the sets of the form $||\varphi||^{\mathcal{M}}_{g,w}$ with sets of the form $[\varphi]^{\mathcal{M}}_{g,w}$.

The first thing to mention is that every well-formed formula $\varphi \in WF$, which does not contain the modal operator $\Box$, has the same truth sets in both systems. More precisely:

Fact 5.1.3. *Given a model $\mathcal{M} = \langle W, \succeq, D, I \rangle$, a variable assignment g, a world w and formula φ which does not contain the modal operator $\Box$ then:*

$$||\varphi||^{\mathcal{M}}_{g,w} = [\varphi]^{\mathcal{M}}_{g,w}$$

This fact leads to the following corollary:

Corollary 5.1.4. *Given a model $\mathcal{M} = \langle W, \succeq, D, I \rangle$, a variable assignment g, a world $w \in W$ and a formula $\varphi \in WF$ which does not contain the modal operator $\Box$ then:*

$$\mathcal{M}, w \models^1_g \varphi \Leftrightarrow \mathcal{M}, w \models^2_g \varphi$$

In particular:

$$\mathcal{M} \models^1 \varphi \Leftrightarrow \mathcal{M} \models^2 \varphi$$

We can see that the logical connectives $\neg, \wedge$, the quantifier $\forall$, the equality symbol $=$, the dyadic deontic operator $\bigcirc$ as well as the derived connectives $\top, \bot, \vee, \supset, \leftrightarrow, \exists, \exists!$ and P have the same interpretations as in $\mathbf{F}_1^{\forall}$ see Lemmas 3.3.1, 3.3.2, 3.3.7, 3.3.8 and 3.3.9.

Let us move on to the differences between Systems 1 and 2, in other words, let us take a look at the interpretation of the $\Box$ operator in $\mathbf{F}_2^{\forall}$. The following lemmas show that in $\mathbf{F}_2^{\forall}$ a formula of the form $\Box\varphi$ is not dependent on the world it is evaluated at. The lemmas will also be used in the soundness proof at the end of this chapter.

Lemma 5.1.5. *Given a model $\mathcal{M} = \langle W, \succeq, D, I \rangle$, variable $x \in V$, variable assignment g, world $w \in W$ and a formula $\varphi \in WF$, then*

$$\mathcal{M}, w \models_g^2 \Box\varphi \text{ holds iff } \forall k \in W : [\varphi]_{g,k}^{\mathcal{M}} = W$$
$$\mathcal{M}, w \models_g^2 \Diamond\varphi \text{ holds iff } \exists k \in W : [\varphi]_{g,k}^{\mathcal{M}} \neq \emptyset.$$

Proof. The first claim follows directly from the definition.
The second claim can seen by the following string of equivalences:
$\mathcal{M}, w \models_g^2 \Diamond\varphi \Leftrightarrow w \in [\Diamond\varphi]_{g,w}^{\mathcal{M}} \Leftrightarrow w \notin [\Box\neg\varphi]_{g,w}^{\mathcal{M}} \Leftrightarrow [\Box\neg\varphi]_{g,w}^{\mathcal{M}} = \emptyset \Leftrightarrow \exists v, k \in W : v \notin [\neg\varphi]_{g,k}^{\mathcal{M}} \Leftrightarrow \exists v, k \in W : v \in [\varphi]_{g,k}^{\mathcal{M}} \Leftrightarrow \exists k \in W : [\varphi]_{g,k}^{\mathcal{M}} \neq \emptyset$

□

Lemma 5.1.6. *Given a model $\mathcal{M} = \langle W, \succeq, D, I \rangle$, variable $x \in V$, variable assignment g, world $w \in W$ and a formula $\varphi \in WF$, then*

$$\mathcal{M}, w \models_g^2 \Box(\varphi \leftrightarrow \psi) \text{ is equivalent to } \forall v \in W : [\varphi]_{g,v}^{\mathcal{M}} = [\psi]_{g,v}^{\mathcal{M}}.$$

Proof. This follows directly from the proof of Lemma 3.3.2 and the interpretation of $\mathbf{F}_2^{\forall}$'s $\Box$ operator.

□

Not so surprisingly, in $\mathbf{F}_2^{\forall}$, the deontic operator $\bigcirc$ cannot be used to define $\Box$, since one of them is a global operator and the other is not:

Theorem 5.1.7. *$\Box\varphi \leftrightarrow \bigcirc(\bot/\neg\varphi)$ is NOT valid in $\mathbf{F}_2^{\forall}$.*

Proof. Given a variable $x \in V$, two 1-place predicate symbol R and Q and a model $\mathcal{M} := \langle W, \succeq, D, I \rangle$ with

$$W := \{w, v\}$$
$$\succeq := RC(\{(v, w)\})$$
$$D_w := \{a, b\}, \ D_v = \{a, b\}$$
$$I(R, w) := \{a\}, \ I(R, v) := \{a\}$$
$$I(Q, w) := \{a\}, \ I(Q, v) := \{b\}$$

Given any variable assignment g then $I_v^g(\imath x Q(x)) = b \notin I(R, w)$, therefore $[R(\imath x Q(x))]_{g,v}^{\mathcal{M}} \neq W$, which means $\mathcal{M}, w \not\models_g^2 \Box R(\imath x Q(x))$. On the other hand since $I_w^g(\imath x Q(x)) = a \in I(R, w)$ and $I_w^g(\imath x Q(x)) = a \in I(R, v)$ we have $[R(\imath x Q(x))]_{g,w}^{\mathcal{M}} = \{w, v\}$, which lets us imply $best(W\backslash[R(\imath x Q(x))]^{\mathcal{M}}_{g,w}) = best(\emptyset) \subseteq \emptyset$ and therefore $\mathcal{M}, w \models^2_g \bigcirc(\bot / \neg R(\imath x Q(x)))$.

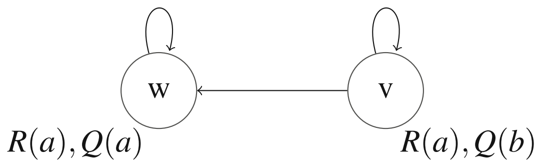

Figure 5.1: Counter-model $\Box - \bigcirc$ relation in $\mathbf{F}_2^\forall$

□

One important distinction between $\mathbf{F}_1^\forall$ and $\mathbf{F}_2^\forall$ is that $\mathbf{F}_2^\forall$ lets us highlight the differences between local and global obligations. This means if given a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$ and a world $w \in W$, $\mathbf{F}_2^\forall$ we can distinguish between $\mathcal{M}, w \models^2 \bigcirc(\varphi/\psi)$ and $\mathcal{M}, w \models^2 \Box\bigcirc(\varphi/\psi)$. The first one describes that the obligation holds true (maybe only) at world w and the second one means that the obligation holds true at every world. Let us demonstrate this distinction by revisiting the "Switching seats" paradox. Since in Example 2.5.1 we want the statement "the person to the right of the person sitting in the corner has the obligation to get up under the condition that the person in the corner gets up" to be true at every world we write it as $\Box\bigcirc(G(\imath y L(\imath x C(x), y))/G(\imath x C(x)))$.

Proposition 5.1.8. *Let* $\Gamma := \{\Box\bigcirc(G(\imath y L(\imath x C(x), y))/G(\imath x C(x))), x_1 = \imath x C(x), x_2 = \imath y L(\imath x C(x), y)\}$ *then*

$$\Gamma \models^2 \bigcirc(G(x_2)/G(x_1)) \text{ but } \Gamma \not\models^2 \Box\bigcirc(G(x_2)/G(x_1)).$$

Proof. Since $\mathcal{M}, w \models^2_g \Box\bigcirc(G(x_2)/G(x_1))$ implies $\mathcal{M}, w \models^2_g \bigcirc(G(x_2)/G(x_1))$ the proof for $\Gamma \models^2 \bigcirc(G(x_2)/G(x_1))$ is the same as the one for $\Gamma \models^1 \bigcirc(G(x_2)/G(x_1))$. We can show $\Gamma \not\models^2 \Box\bigcirc(G(x_2)/G(x_1))$ by the use of the following counterexample:

Example 5.1.9. *Given a variable $x \in V$, two 1-place predicate symbols C and G, a 2-place predicate symbol L, a model $\mathcal{M} := \langle W, \succeq, D, I\rangle$ with*

$$W := \{w, v\}$$
$$\succeq := RC(\{(v, w)\})$$
$$D_w := \{b, i, t\}, \; D_v := \{b, i, t\}$$
$$I(C, w) := \{b\}, \; I(C, v) := \{t\}$$
$$I(G, w) := \{\}, \; I(G, v) := \{b, i\}$$
$$I(L, w) := \{(b, i)\}, \; I(L, v) := \{\}$$

Given a variable assignment g with $g(x_1,w) = b$, $g(x_2,w) = i$, $g(x_1,v) = i$ and $g(x_2,v) = \mathbb{D}$ then we can see that at the world w the equalities in Γ are true in regards to g. Also since $best([G(\imath xC(x))]^{\mathcal{M}}_{g,w}) = best(\{v\}) = \{v\} \subseteq \{v\} = [G(\imath yL(\imath xC(x),y)))]^{\mathcal{M}}_{g,w}$ and because $best([G(\imath xC(x))]^{\mathcal{M}}_{g,v}) = best(\emptyset) = \emptyset \subseteq \emptyset = [G(\imath yL(\imath xC(x),y)))]^{\mathcal{M}}_{g,v}$ the formula $\Box\bigcirc(G(\imath yL(\imath xC(x),y))/G(\imath xC(x)))$ is true at the world w under g. On the other hand because $best([G(x_1)]^{\mathcal{M}}_{g,v}) = best(\{v\}) = \{v\} \not\subseteq \emptyset = [G(x_2)]^{\mathcal{M}}_{g,v}$ we do NOT have $\mathcal{M},w \models^2_g \Box\bigcirc(G(x_2)/G(x_1))$.

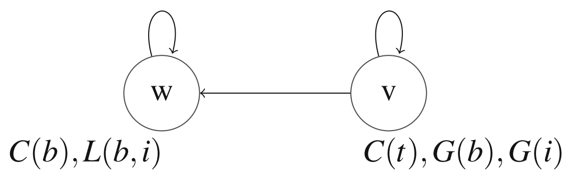

Figure 5.2: Switching seats model

□

5.2 Provability and Soundness

We now introduce a Hilbert axiomatisation for $\mathbf{F}_2^\forall$, which we call $\mathbf{HF}_2^\forall$. Given $\mathbf{HF}_2^\forall$, we define what it means for a formula $\varphi \in WF$ to be provable in $\mathbf{HF}_2^\forall$ (in symbols: $\vdash^2$) the same way as in $\mathbf{HF}_1^\forall$, but replacing the axiom schemas and rules of $\mathbf{HF}_1^\forall$ by the the axioms and rules of $\mathbf{HF}_2^\forall$. The same goes for the definition of weak and strong soundness of $\mathbf{HF}_2^\forall$ in $\mathbf{F}_2^\forall$.

As we have seen in the previous section, $\mathbf{F}_1^\forall$ and $\mathbf{F}_2^\forall$ are similar in many ways. Therefore we take $\mathbf{HF}_1^\forall$ as a basis for this second Hilbert axiomatisation and check which axioms and rules have to be altered to create a sound Hilbert System for $\mathbf{F}_2^\forall$. We are going to prove in this section that all rules of $\mathbf{HF}_1^\forall$ are also sound in $\mathbf{F}_2^\forall$. Let us, therefore, discuss which axioms of $\mathbf{HF}_1^\forall$ need to be altered to fit $\mathbf{F}_2^\forall$. Because of the different interpretation of the $\Box$ operator the formula schema $t \neq s \supset \Box t \neq s$ is not valid in $\mathbf{F}_2^\forall$ for every pair of terms t and s, hence the axiom **E2** is not going to be part of $\mathbf{HF}_2^\forall$. The axioms **Nec** and **Ext** must be elevated to new axioms **Nec2** and **Ext2** by strengthening their respective consequences since the original rules would be too weak for $\mathbf{F}_2^\forall$. The axioms **I** and **UI** have restrictions placed on them, making them applicable if the replaced term is not contained in the scope of the $\Box$ operator. Unsurprisingly, the axiom **Abs** is not sound in $\mathbf{F}_2^\forall$ because this would mean that $\bigcirc$ would be a global operator, which it is not. The axiom **D*** is also not sound

sound in $\mathbf{F}_2^{\forall}$ because of the different interpretation of the derived operator $\Diamond$. Those two axioms are replaced by their weakened counterparts, the rule **Abs E** and the axiom **D**, respectively. $\mathbf{HF}_2^{\forall}$ consists of the following axioms and rules:

Axioms:

All truth functional tautologies		(**PL**)
S5-schemata for $\Box$ and $\Diamond$		(**S5**)
$P(\varphi/\psi) \leftrightarrow \neg\bigcirc(\neg\varphi/\psi)$		(**DfP**)
$\bigcirc(\varphi \supset \chi/\psi) \supset (\bigcirc(\varphi/\psi) \supset \bigcirc(\chi/\psi))$		(**COK**)
$\Box\varphi \supset \Box\bigcirc(\varphi/\psi)$		(**Nec2**)
$\Box(\varphi \leftrightarrow \psi) \supset \Box(\bigcirc(\chi/\varphi) \leftrightarrow \bigcirc(\chi/\psi))$		(**Ext2**)
$\bigcirc(\varphi/\varphi)$		(**Id**)
$\bigcirc(\varphi/\psi \wedge \chi) \supset \bigcirc(\chi \supset \varphi/\psi)$		(**Sh**)
$t = s \supset (\varphi \leftrightarrow \varphi_{t \hookrightarrow s})$	if t is not in the scope of the $\Box$ operator	(**I**)
$E(t) \supset (\forall x\varphi \supset \varphi_{x \Rightarrow t})$	if x is not in the scope of the $\Box$ operator	(**UI**)
$\exists x \exists y(x = y)$		(**Ex**)
$t = t$		(**E1**)
$\forall y((\forall x(\varphi \leftrightarrow x = y)) \supset y = \iota x\varphi)$		(**D1**)
$E(\iota x\varphi) \supset \exists! x\varphi$		(**D2**)
$\forall x(E(x) \supset \varphi) \supset \forall x\varphi$		(**UQ**)
$(\forall x\varphi \wedge \forall x\psi) \leftrightarrow \forall x(\varphi \wedge \psi)$		(**QD**)
$\bigcirc\varphi \supset P\varphi$		(**D**)

Rules:

If $\vdash \bigcirc(\varphi/\psi)$ then $\vdash \Box\bigcirc(\varphi/\psi)$		(**Abs E**)
If $\vdash \varphi \supset t \neq x$ then $\vdash \neg\varphi$	where $x \notin \text{free}(\varphi)$	(**R1**)
If $\vdash \varphi$ and $\vdash \varphi \supset \chi$ then $\vdash \chi$		(**MP**)
If $\vdash \varphi^*$ then $\vdash \Box\varphi$		(**N***)
If $\vdash \varphi \supset \psi$ then $\vdash \varphi \supset \forall x\psi$	where $x \notin \text{free}(\varphi)$	(**IU**)
If $\vdash \varphi \supset \Box\psi$ then $\vdash \varphi \supset \Box\forall x\psi$	where $x \notin \text{free}(\varphi)$	(**IUB**)

In the same way, as for $\mathbf{F}_1^{\forall}$, the soundness proof for $\mathbf{F}_2^{\forall}$ can be accomplished by showing that every instance of every axiom of $\mathbf{HF}_2^\forall$ is valid and that the inference rules preserve validity in $\mathbf{F}_2^\forall$. Hence for the final part of this thesis, we show that $\mathbf{HF}_2^\forall$ is strongly sound in $\mathbf{F}_2^\forall$ by first showing that is weakly sound in $\mathbf{F}_2^\forall$:

Theorem 5.2.1. *Given a well-formed formula $\varphi \in WF$, a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$, a world $w \in W$ and a variable assignment g of $\mathscr{M}$ the implication*

$$\vdash^2 \varphi \Rightarrow \mathscr{M}, w \models^2_g \varphi$$

holds. In general, the implication

$$\vdash^2 \varphi \Rightarrow \models^2 \varphi$$

holds.

Let us start with the axioms of the propositional system **F**. As we have already seen in Example 5.1.9 **Abs** is not sound in $\mathbf{F}_2^\forall$. Nevertheless, most of the axioms of **F** are sound in $\mathbf{F}_2^\forall$. More specific, all rules and axioms except for **N**, **Abs** and **D*** are sound in $\mathbf{F}_2^\forall$. In the soundness proof of the axioms of **F** in $\mathbf{F}_1^\forall$, we did not use the interpretation of $\Box$ when talking about the axiom schemas not containing the modal operator $\Box$ in their formula schema. We can therefore see that the following theorem holds.

Theorem 5.2.2. *Every axiom of system **F** not containing the modal operator $\Box$ in its formula schema is sound in $\mathbf{F}_2^\forall$.*

Proof. Replace $||\varphi||^{\mathscr{M}}_{g,w}$ with $[\varphi]^{\mathscr{M}}_{g,w}$ in the soundness proofs of the axioms **COK**, **Id**, **Sh** and the rule **MP**.

$\Box$

Even though the interpretation of the $\Box$ operator is different in $\mathbf{F}_2^\forall$ than in $\mathbf{F}_1^\forall$, the axioms **Nec** and **Ext**, [1] as well as the **S5**-schemata for $\Box$ and $\Diamond$ are sound in $\mathbf{F}_2^\forall$.

Theorem 5.2.3. *The axioms of **S5** and the axioms **Nec2** and **Ext2** are sound in $\mathbf{F}_2^\forall$.*

Proof.
- $\Box(\varphi \supset \chi) \supset (\Box\varphi \supset \Box\chi)$

  $\mathscr{M}, w \models^2_g \Box(\varphi \supset \chi)$ is equivalent to $\forall k \in W : W \backslash [\varphi]^{\mathscr{M}}_{g,k} \cup [\chi]^{\mathscr{M}}_{g,k} = W$. Now if $w \in [\Box\varphi]^{\mathscr{M}}_{g,w}$ then $\forall k \in W : [\varphi]^{\mathscr{M}}_{g,k} = W$, therefore together with $\mathscr{M}, w \models^2_g \Box(\varphi \supset \chi)$ we can infer that $\forall k \in W : [\chi]^{\mathscr{M}}_{g,k} = W$, which lets us conclude $\mathscr{M}, w \models^2_g \Box\chi$.

[1] They were replaced in $\mathbf{HF}_2^\forall$ by their stronger versions **Nec2** and **Ext2**.

- $\Box\varphi \supset \varphi$

  Since $\mathscr{M}, w \models^1_g \Box\varphi$ holds iff $\forall k \in W : [\varphi]^{\mathscr{M}}_{g,k} = W$ it follows that φ holds in w under g according to w, which means $\mathscr{M}, w \models^2_g \varphi$.

- $\Diamond\varphi \supset \Box\Diamond\varphi$

  If $\mathscr{M}, w \models^2_g \Diamond\varphi$ then $w \in [\neg\Box\neg\varphi]^{\mathscr{M}}_{g,w}$ which implies $\forall k \in W : [\neg\Box\neg\varphi]^{\mathscr{M}}_{g,k} = W$, therefore $[\Box\neg\Box\neg\varphi]^{\mathscr{M}}_{g,w} = W$, hence $w \in [\Box\Diamond\varphi]^{\mathscr{M}}_{g,w}$.

- $\Box\varphi \supset \Box\bigcirc(\varphi/\psi)$

  Given a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g and a world $w \in W$ then the following equivalences hold: $\mathscr{M}, w \models^2_g \Box\varphi \Leftrightarrow w \in \{v \in W : \forall k \in W [\varphi]^{\mathscr{M}}_{g,k} = W\} \Leftrightarrow W = \{v \in W : \forall k \in W [\varphi]^{\mathscr{M}}_{g,k} = W\} \Leftrightarrow \forall k \in W : [\varphi]^{\mathscr{M}}_{g,k} = W$. Hence given any well-formed formula $\psi \in WF$, we can conclude that for every world $k \in W$ we have $best([\psi]^{\mathscr{M}}_{g,k}) \subseteq W = [\varphi]^{\mathscr{M}}_{g,k}$ which is equivalent to $\mathscr{M}, w \models^2_g \Box\bigcirc(\varphi/\psi)$.

- $\Box(\varphi \leftrightarrow \psi) \supset \Box(\bigcirc(\chi/\varphi) \leftrightarrow \bigcirc(\chi/\psi))$

  $\mathscr{M}, w \models^2_g \Box(\varphi \leftrightarrow \psi)$ is equivalent to $\forall k \in W : [\varphi]^{\mathscr{M}}_{g,k} = [\psi]^{\mathscr{M}}_{g,k}$, see Lemma 5.1.6. This equivalence implies that from $\mathscr{M}, w \models^2_{g,w} \Box(\varphi \leftrightarrow \psi)$ it follows that for every world k the optimized φ worlds of $\mathscr{M}$ in the $\mathbf{F}_2^{\forall}$ under g according to k coincide with the optimized ψ worlds of $\mathscr{M}$ in the $\mathbf{F}_2^{\forall}$ under g according to k. This means $\forall k \in W : best([\varphi]^{\mathscr{M}}_{g,k}) = best([\psi]^{\mathscr{M}}_{g,k})$. This implication lets us conclude that $\mathscr{M}, w \models^2_g \Box(\varphi \leftrightarrow \psi) \supset \Box(\bigcirc(\chi/\psi) \leftrightarrow \bigcirc(\chi/\varphi))$.

$\Box$

I and **UI** are sound in $\mathbf{F}_2^{\forall}$ under certain restrictions: we cannot replace terms inside the scope of the $\Box$ operator without changing the formula's truth value, as already seen in Example 5.1.9. The soundness of **I** and **UI** can be proven by showing that Theorems 4.3.1 and 4.4.1 also hold for $\mathbf{F}_2^{\forall}$ if we do not allow φ to contain the modal operator $\Box$:

Theorem 5.2.4. *Given a model $\mathscr{M} = \langle W, \succeq, D, I\rangle$, two terms s and t, a variable assignment g, a world $w \in W$ and a formula $\varphi \in WF$ which does not contain the modal operator $\Box$ and with $fv(t) \cap bv(\varphi) = \emptyset$ and $fv(s) \cap bv(\varphi) = \emptyset$ then $I^g_w(t) = I^g_w(s)$ implies:*

$$[\varphi_{t \hookrightarrow s}]^{\mathscr{M}}_{g,w} = [\varphi]^{\mathscr{M}}_{g,w} \tag{5.1}$$

Proof. This is the same proof as for Theorem 4.3.1 except that we replace the connotations for $\mathbf{F}_1^{\forall}$ with those for $\mathbf{F}_2^{\forall}$ and skip the step containing the $\Box$ operator.

$\Box$

Theorem 5.2.5. *Given a formula well-formed $\varphi \in WF$ which does not contain the modal operator $\Box$, a model $\mathscr{M}$, a variable assignment g, a world $w \in W$, a term t with $bv(\varphi) \cap fv(t) = \emptyset$ and $d := I_w^g(t)$ then the following equation holds:*

$$[\varphi_{x \Rightarrow t}]_{g,w}^{\mathscr{M}} = [\varphi]_{g_{x \Rightarrow d},w}^{\mathscr{M}}$$

Proof. This is the same proof as for Theorem 4.4.1 except that we replace the connotations for $\mathbf{F}_1^{\forall}$ with those for $\mathbf{F}_2^{\forall}$ and skip the step containing the $\Box$ operator.

□

The theorems above imply the following corollaries in the same way as in Chapter 4:

Corollary 5.2.6. *Given a model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, a variable assignment g, a world $w \in W$, two terms t, s and a formula $\varphi \in WF$ which does not contain the modal operator $\Box$ and with $bv(\varphi) \cap fv(t) = \emptyset$ and $bv(\varphi) \cap fv(s) = \emptyset$ then*

$$\mathscr{M}, w \models_g^2 t = s \supset (\varphi \leftrightarrow \varphi_{t \hookrightarrow s}).$$

In particular: Given two terms t, s and a formula $\varphi \in WF$ which does not contain the modal operator $\Box$ and with $bv(\varphi) \cap fv(t) = \emptyset$ and $bv(\varphi) \cap fv(s) = \emptyset$ then

$$\models^2 t = s \supset (\varphi \leftrightarrow \varphi_{t \hookrightarrow s}).$$

Corollary 5.2.7. *Given a model $\mathscr{M} = \langle W, \succeq, D, I \rangle$, a variable assignment g, a world $w \in W$, a variable $x \in V$, a term t, a formula $\varphi \in WF$ which does not contain the modal operator $\Box$ and with $bv(\varphi) \cap fv(t) = \emptyset$ then*

$$\mathscr{M}, w \models_g^2 E(t) \supset (\forall x \varphi \supset \varphi_{x \Rightarrow t}).$$

In particular: Given a variable $x \in V$, a term t and a formula $\varphi \in WF$ which does not contain the modal operator $\Box$ and with $bv(\varphi) \cap fv(t) = \emptyset$ then

$$\models^2 E(t) \supset (\forall x \varphi \supset \varphi_{x \Rightarrow t}).$$

We are again using two lemmas to prove that the rule **N*** is sound, but now in $\mathbf{F}_2^{\forall}$.

Lemma 5.2.8. *Given a well-formed formula $\varphi \in WF$ and a model $\mathscr{M}$ then*

$$\mathscr{M} \models^2 \varphi^* \Rightarrow \mathscr{M} \models^2 \Box \varphi^*.$$

Proof. Let φ be any fixed, well-formed formula and $\mathscr{M} = \langle W, \succeq, D, I \rangle$ a model. If for every world $w \in W$ and every variable assignment g of $\mathscr{M}$ it holds that $\mathscr{M}, w \models_g^2 \varphi^*$, it follows that $w \in [\varphi^*]^{\mathcal{M}}_{g,w}$ for every world $w \in W$ and every variable assignment g of $\mathcal{M}$. Now let us take two arbitrary but fixed worlds $v, w \in W$ and an arbitrary but fixed variable assignment g and define a new variable assignment $h : V \times W \to \mathbb{D}^+$ of $\mathcal{M}$ as:

$$h(x,v') := \begin{cases} g(x,w) & \text{if } v' = v \\ g(x,v) & \text{if } v' = w \\ g(x,v') & \text{otherwise} \end{cases}$$

Since h and g only swap how they see the variables at w and v we get the equality $[\varphi^*]^{\mathcal{M}}_{h,v} = [\varphi^*]^{\mathcal{M}}_{g,w}$. It follows that $v \in [\varphi^*]^{\mathcal{M}}_{h,v} = [\varphi^*]^{\mathcal{M}}_{g,w}$. Since v was arbitrary we can conclude $[\varphi^*]^{\mathcal{M}}_{g,w} = W$. Since w was an arbitrary world and g was an arbitrary variable assignment of $\mathcal{M}$, we can further conclude $\forall w \in W : [\varphi^*]^{\mathcal{M}}_{g,w} = W$ for all variable assignments g of $\mathcal{M}$ which means $\mathcal{M} \models^2 \Box\varphi^*$. □

Lemma 5.2.9. *Given a well-formed formula $\varphi \in WF$ and a model $\mathcal{M}$ then*

$$\mathcal{M} \models^2 \varphi^* \Rightarrow \mathcal{M} \models^2 \varphi.$$

Proof. The following proof is done by contraposition. Let us assume that there exists a model $\mathcal{M} = \langle W, \succeq, D, I \rangle$, a world $w \in W$ and a variable assignment g such that $\mathcal{M}, w \not\models^2_g \varphi$. Let $t_1, \ldots, t_n$ be all terms in φ which are replaced by the corresponding variables $x_1, \ldots, x_n$ in φ^* then for the variable assignment

$$h(x,v) := \begin{cases} I^g_v(t_i) & \text{if } (x,v) \in \{x_i\} \times W \text{ where } i \in \{1,\ldots,n\} \\ g(x,v) & \text{otherwise} \end{cases}$$

we have $\mathcal{M}, w \not\models^2_h \varphi^*$.

□

Putting those two lemmas together, we can prove the soundness of rule **N***:

Theorem 5.2.10. *Given a well-formed formula $\varphi \in WF$ and a model $\mathcal{M}$ then*

$$\mathcal{M} \models^2 \varphi^* \quad \textit{implies} \quad \mathcal{M} \models^2 \Box\varphi.$$

Proof. $\mathcal{M} \models^2 \varphi^* \Rightarrow \mathcal{M} \models^2 \Box\varphi^* \Leftrightarrow \mathcal{M} \models^2 (\Box\varphi)^* \Rightarrow \mathcal{M} \models^2 \Box\varphi$. □

In $\mathbf{F}_2^{\forall}$ the soundness proofs for the axioms **Ex**, **E1**, **D1**, **D2**, **UQ** and **QD** and the rules **R1** and **IU** are the same as for $\mathbf{F}_1^{\forall}$, by just replacing sets of the form $||\varphi||^{\mathcal{M}}_{g,w}$ with $[\varphi]^{\mathcal{M}}_{g,w}$. Therefore, for this final proof, the only part left to prove is the soundness of the axiom **D** and of the rules **Abs E** and **IUB**:

Proof of Theorem 5.2.1.

- $\bigcirc\varphi \supset P\varphi$

  $\mathscr{M},w \models_g^2 \bigcirc\varphi$ is equivalent to $best([\top]_{g,w}^{\mathscr{M}}) \subseteq [\varphi]_{g,w}^{\mathscr{M}}$. Furthermore using the limitedness property of our betterness relation $\succeq$ we get $best([\top]_{g,w}^{\mathscr{M}}) \neq \emptyset$. This means we can find a world $v' \in W$ with $v' \in best([\top]_{g,w}^{\mathscr{M}}) \subseteq [\varphi]_{g,w}^{\mathscr{M}}$. This lets us conclude $best([\top]_{g,w}^{\mathscr{M}}) \cap [\varphi]_{g,w}^{\mathscr{M}} \neq \emptyset$, which is equivalent to $\mathscr{M},w \models_g^2 P\varphi$.

- The rule **Abs E**

  $\mathscr{M} \models^2 \bigcirc(\varphi/\psi)$ is equivalent to $\{v \in W : best([\psi]_{g,w}^{\mathscr{M}}) \subseteq [\varphi]_{g,w}^{\mathscr{M}}\} = W$ for all worlds w and all variable assignments g. This is the same as $\{v \in W : \forall w \in W\ best([\psi]_{g,w}^{\mathscr{M}}) \subseteq [\varphi]_{g,w}^{\mathscr{M}}\} = W$ for all variable assignments g which is equivalent to $\mathscr{M} \models^2 \Box\bigcirc(\varphi/\psi)$.

- The rule **IUB**

  By contraposition: Suppose there exists a model $\mathscr{M}$, a world w, two well-formed formulas φ and ψ and a variable assignment g such that $\mathscr{M},w \not\models_g^2 \varphi \supset \Box\forall x\psi$. Then $w \notin [\varphi \supset \Box\forall x\psi]_{g,w}^{\mathscr{M}}$ which is equivalent to $w \in [\varphi]_{g,w}^{\mathscr{M}}$ and $w \notin [\Box\forall x\psi]_{g,w}^{\mathscr{M}}$. Hence we can find two worlds v and v' and an x-variant of g at v such that $v \notin [\psi]_{h,v'}^{\mathscr{M}}$ which implies $w \notin [\Box\psi]_{h,v'}^{\mathscr{M}} = [\Box\psi]_{h,w}^{\mathscr{M}}$.[2] If $x \notin free(\varphi)$ it holds that $[\varphi]_{g,w}^{\mathscr{M}} = [\varphi]_{h,w}^{\mathscr{M}}$. We have therefore found a world w and a variable assignment h of $\mathscr{M}$ such that $w \notin [\varphi \supset \Box\psi]_{h,w}^{\mathscr{M}}$ which means $\mathscr{M} \not\models \varphi \supset \Box\psi$.

□

Remark 5.2.11. *Strong soundness of $\boldsymbol{HF}_2^{\forall}$ follows directly from weak soundness of $\boldsymbol{HF}_2^{\forall}$ and by the respective definition of derivability, similar to $\boldsymbol{HF}_1^{\forall}$.*

[2] A set of the form $[\Box\psi]_{h,v'}^{\mathscr{M}}$ does not depend on v'.

6 Conclusion

The main goal of this thesis was to create a first-order dyadic deontic logic system extending the propositional system **F**, that includes equality, definite description and an extensional conditional obligation operator.
We discussed which axioms such a dyadic deontic logic has to fulfil and based on the work done in [8], we have shown that adding those axioms to a first-order extension of **F** results in the collapse of the modal and deontic operators if no restrictions are placed on the rules of **F**. In particular a restriction on the rule **N** was needed to avoid the collapse of the operators.
We have established the semantics for a first-order dyadic deontic logic system called $\mathbf{F}_1^{\forall}$. In this system, the interpretation of every term gets bound to the world at which the formula is evaluated, even if the term is contained inside a modal or deontic operator.
The Hilbert axiomatisation $\mathbf{HF}_1^{\forall}$ includes every axiom of the propositional system **F**, a restricted version of the rule **N** and the axioms concerning extensional operators and definite descriptions. Finally, by showing that this Hilbert system is sound in the semantics of the previous chapter, we can answer the **main research question** of this thesis:

> Is it possible to create a first-order dyadic deontic logic system extending the propositional system **F**, with an extensional deontic operator $\bigcirc$, while avoiding its collapse?

Answer: Yes, it is possible, with the minor exception of putting a restriction on the rule **N**.

By defining and discussing the "Switching seats" paradox, we have seen that considering an operator to be extensional and global simultaneously results in contradictory statements. Hence, in this thesis, we only considered our dyadic deontic operator to have a local interpretation. In $\mathbf{F}_2^{\forall}$ we made the $\Box$ operator into a global operator by strengthening its interpretation. This made it a non-extensional operator while keeping the Extensionality of $\bigcirc$ unchanged. We defined a Hilbert axiomatisation $\mathbf{HF}_2^{\forall}$ for $\mathbf{F}_2^{\forall}$ and by showing that this Hilbert system is sound in $\mathbf{F}_2^{\forall}$, we have seen that it is possible to create a first-order dyadic deontic logic system with an extensional deontic operator $\bigcirc$ and a non-extensional modal operator $\Box$.

Unsurprisingly, we did not create a first-order system with an extensional deontic operator $\bigcirc$ and a non-extensional modal operator $\Box$ without losing at least one axiom of the propositional system **F**. $\mathbf{HF}_2^{\forall}$ does not contain the axioms **Abs** and **D***. However, those two axioms are replaced by a weaker rule and axiom, respectively. The reason is that for all axioms of **F** to be sound, the operators $\Box$ and $\bigcirc$ have to share a similar connection as they do in the propositional system **F**. Since in the system **F** the formula schema $\Box\varphi \leftrightarrow \bigcirc(\bot/\neg\varphi)$ is provable for every formula φ we cannot expect for all axioms of **F** to be sound when the properties of $\Box$ and $\bigcirc$ differ too much. Nevertheless, we have shown that it is possible to create a first-order dyadic deontic logic system with an extensional deontic operator $\bigcirc$ and a non-extensional modal operator $\Box$ with most of the axioms of **F** being sound. Therefore we can answer the **second research question**:

> Is it possible to create a first-order dyadic deontic logic system extending the propositional system **F**, with an extensional deontic operator $\bigcirc$ and a non-extensional modal operator $\Box$?

Answer: Yes, it is possible, with the exception of putting a restriction on the rule **N** and weakening the axioms **Abs** and **D***.

6.1 Further research

The biggest unanswered questions of this thesis are: "Is every valid formula of $\mathbf{F}_1^{\forall}$ provable in $\mathbf{HF}_1^{\forall}$?" and "Is every valid formula of $\mathbf{F}_2^{\forall}$ provable in $\mathbf{HF}_2^{\forall}$?". In other words, "Are $\mathbf{F}_1^{\forall}$ and $\mathbf{F}_2^{\forall}$ complete with regards to the introduced axiomatisations?". To achieve completeness it may be the case that missing axioms and rules have to be added to the proposed Hilbert systems in Chapters 4 and 5.

Another suggestion for future research is to check which properties in Definition 3.2.1 and Definition 3.2.3 can be changed without changing the axioms or rules of the $\mathbf{HF}_1^{\forall}$ or $\mathbf{HF}_2^{\forall}$ too much or at all. Examples of questions regarding such property changes are: "Is it possible for the betterness relation $\succeq$ to have different properties than reflexivity and limitedness?", "Which effect would it have if the best worlds are defined as the maximal worlds of a set instead of the optimal worlds of a set?" and "What happens if the domains of a model are increasing or decreasing regarding the betterness relation?"

Lastly, it would be interesting to check if combining $\mathbf{F}_1^{\forall}$ and $\mathbf{F}_2^{\forall}$ into one system is possible. "Is it possible to define a system with two different necessitation operators, one being an extensional and local operator like in $\mathbf{F}_1^{\forall}$ and the other being a non-extensional global operator, like in $\mathbf{F}_2^{\forall}$?" This new system would cover the expressiveness of both systems established in this thesis and could be used to express even more detailed statements than $\mathbf{F}_1^{\forall}$ or $\mathbf{F}_2^{\forall}$ alone.

Bibliography

[1] L. Åqvist. An Introduction to Deontic logic and the Theory of Normative Systems. *Naples: Bibliopolis*, 1987.

[2] R. M. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis*, 24(2):33–36, 1963.

[3] S. Danielsson. Preference and obligation. *Filosofiska Färeningen, Uppsala*, 1968.

[4] M. Fitting and R. L. Mendelsohn. *First-order modal logic*, volume 277. Springer Science & Business Media, 2012.

[5] J. W. Forrester. Gentle murder, or the adverbial Samaritan. *The Journal of Philosophy*, 81(4):193–197, 1984.

[6] D. Gabbay, J. Horty, X. Parent, R. van der Meyden, and L. van der Torre. *Handbook of Deontic Logic and Normative Systems*, volume 1. College Publications, United Kingdom, 2013.

[7] D. Gabbay, J. Horty, X. Parent, R. van der Meyden, and L. van der Torre. *The Handbook of Deontic Logic and Normative Systems*, volume 2. College Publications, United Kingdom, 07 2021.

[8] L. Goble. Opacity and the Ought-to-be. *Noûs*, pages 407–412, 1973.

[9] L. Goble. Quantified deontic logic with definite descriptions. *Logique et Analyse*, 37(147/148):239–253, 1994.

[10] L. Goble. 'Ought' and Extensionality. *Noûs*, 30(3):330–355, 1996.

[11] B. Hansson. An analysis of some deontic logics. *Noûs*, pages 373–398, 1969.

[12] D. M. Hausman. Philosophy of Economics. `https://plato.stanford.edu/archives/win2021/entries/economics/`, 2021.

[13] C. I. Lewis, C. H. Langford, and P. Lamprecht. *Symbolic logic*, volume 170. Dover Publications New York, 1959.

[14] D. Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.

[15] X. Parent. Completeness of Åqvist's systems E and F. *The Review of Symbolic Logic*, 8(1):164–177, 2015.

[16] A. Prior. *Past, Present and Future*. Oxford University Press, 05 1967.

[17] W. Sinnott-Armstrong. A solution to Forrester's paradox of gentle murder. *The Journal of Philosophy*, 82(3):162–168, 1985.

[18] R. H. Thomason. Some completeness results for modal predicate calculi. In *Philosophical problems in logic*, pages 56–76. Springer, 1970.

[19] G. H. von Wright. Deontic logic. *Mind*, 60(237):1–15, 1951.

[20] G. H. von Wright. A note on deontic logic and derived obligation. *Mind*, 65(260):507–509, 1956.

List of Figures