How Knowledge Affects Obligations

Su, Xingchi

DOI:
10.33612/diss.243686372

Publication date:
2022

Xingchi Su
苏兴池

How Knowledge Affects Obligations:
A Study on the Logic of Knowledge-Based Obligations

Printing: Ridderprint B.V.

Colophon: This thesis was typeset with LaTeX, using Diego Puga's Pazo math fonts.

The research described in this dissertation has been carried out at the Faculty of Philosophy and the Faculty of Science and Engineering, University of Groningen, the Netherlands.

The research reported in this dissertation was supported by the China Scholarship Council (CSC), the Chinese Ministry of Education.

Cover design by Mingpu Ren

university of groningen

How Knowledge Affects Obligations:

A Study on the Logic of Knowledge-Based Obligations

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with the decision by the College of Deans.

This thesis will be defended in public on

Thursday 29 September at 12:45 hours

by

Xingchi Su

born on 20 September 1993
in Gansu, China

Supervisors
Dr. B.P. Kooi
Prof. L.C. Verbrugge
Prof. D. Grossi

Assessment committee
Prof. H.B. Verheij
Prof. J.M. Broersen
Prof. F. Liu

To my beloved family

Increased knowledge clearly implies increased responsibility.

—Nicolaas Bloembergen

Contents

Chapter 1

Introduction

Every member of society has their own obligations. These obligations arise from morality, ethical codes, laws or regulations. According to some norms or rules, a person is assigned several obligations. For example, a doctor's professional norms require them to treat their patients. Therefore, a doctor has the obligation to treat his/her patients. It is straightforward to see that our obligations are dependent on the norms that we are following.

Furthermore, norms give rise to obligations towards people in a *conditional* manner. It is reasonable to say that a doctor ought to treat a patient if the patient is in his/her practice. We can imagine that almost all obligations are conditional even if they are involved in some extreme cases. For instance, we ought not to kill people. But there is an exception when we are defending ourselves justifiably or when we are in war as soldiers.

A very natural question is: what is the type of the conditions embedded in conditional obligations? An answer is *fact*. If it is a fact that the patient is in the doctor's practice, then the doctor ought to treat the patient. However, in real-life scenarios, we are inclined to judge whether a person has an obligation to do something by assessing the person's *knowledge* in a certain situation. For example, it makes sense to say that if the doctor *knows* that the patient is in his/her practice, then the doctor ought to treat the patient. Relatedly, if the doctor does not know whether the patient is in his/her practice, it is confusing for us to tell whether the doctor has the obligation to treat the patient. For another example, if an employee knows that the documents should be sent by tonight, then the employee ought to send the documents by tonight. Considering similar cases, knowledge can also be a suitable candidate for the role of conditions in conditional obligations in the agent-involved cases. Accordingly, we call them *knowledge-based conditional obligations.*

Since conditional obligations can be based on knowledge, does knowledge change lead to obligation change? The dynamics of knowledge has been widely investigated and formalized in dynamic epistemic logic for decades, see Van Benthem (1997), Gerbrandy and Groeneveld (1997), van Ditmarsch et al. (2007), van Ditmarsch and Kooi (2008), Wang and Cao (2013). A person's knowledge can be updated by new information or actions by performing knowingly. Since there are growing interests in knowledge-based obligations in recent years (see Pacuit et al. (2006)), knowledge dynamic naturally gives rise to obligation dynamic. For example, before a patient is assigned to the doctor's hospital, the doctor does not know that the patient is ill and does not have the obligation to treat the patient. Once the patient arrives in the hospital and the doctor gets to know the situation, he/she has a new obligation to treat the patient. It is apparent that obligations can be updated by knowledge change.

Apart from the epistemic factor, norm updates can also change obligations. As mentioned above, norms fundamentally determine our obligations. Our obligations can be drastically changed as the norms that we are following are updated. For instance, an employee is asked to send an email by 9 o'clock. Then her boss asks her to finish it two hours earlier. Then the employee ought to send the email by 7 o'clock, rather than 9 o'clock. The original obligation is revoked and a new obligation takes its place as a result of the new norms given by the boss.

The relation between norms and obligations is widely known as the distinction between prescription and description in deontic logic. Norms are some rules that are to *prescribe* which states of affairs are better and which are worse under some conditions. They are not able to have truth values. In contrast, an obligation *describes* what state of affairs is good for some particular person. An obligation is normally formalized as a proposition which can have truth values. There are growing appeals for treating norms and obligations separately, such as Makinson and van der Torre (2007b), Hansen et al. (2007), van der Torre and Tan (1998), Yamada (2006, 2008), Aucher et al. (2009), etc., so as to resolve the issues due to the discrepancy between prescriptive and descriptive readings of normative sentences.

This thesis grew out of interests in providing a novel logic of knowledge-based conditional obligations and different dynamic logics of obligations with respect to epistemic change or norm updates. The logic of knowledge-based conditional obligations is a static formalization of agent-based obligations. It explicitly characterizes the scenarios where agents' knowledge is embedded in their conditional obligations. Then we present a dynamic logic of knowledge-based conditional obligation, which is able to capture agents' conditional obligation change due to the agents' epistemic change or factual change. Followed by norm change, a logic of relativized conditional obligations to normative systems captures the dynamic obligations as a result of norm change. In the end, we will jump out of the tradition of 'ought-to-be' and establish a new logic of knowledge-based 'ought-to-do', where dynamic epistemic logic plays a different role from what it does in epistemic and factual changes.

In the remaining part of this chapter, to better understand the background of this thesis, we give an overview of core ideas behind the related fields. We hope readers will be able to figure out the point of departure of our research.

1.1 Standard deontic logic

Deontic logic is a branch of philosophical logic that investigates the formal representation and reasoning of deontic notions, such as obligations, permission, prohibition etc. Obligation is normally considered as the most primitive notion in deontic logic. Von Wright (1951) launched the area and first put forward the obligation operator O which represents what action ought to be done. In the following decades, researchers provided various frameworks for deontic logic so as to conceptualize 'it ought to be' and 'someone ought to do'.

Standard deontic logic (SDL) is the first milestone in deontic logic. Partly following the idea of Von Wright (1951), standard deontic logic interprets the objects of obligations as propositions rather than actions, which also made a great impact on the subsequent research. Prior (1955) and Hintikka (1957) refined Von Wright's original system and Von Wright (1970) himself introduced a new SDL system where the objects of obligations are represented as propositions as well. In the new system, the term Op is read as 'one ought to see to the state of affairs p'. The semantics of the operator O is defined in the same way as the alethic operator $\Box$ over classical Kripke models, i.e., $M = \langle W, R, V \rangle$. The relation R in a model is used for representing the set of the most ideal states with respect to each state. Consequently, 'it ought to be ϕ' ($O\phi$) is conceptualized as 'ϕ is true over all the most ideal states'. See the below formalism:

$$M, s \models O\phi \text{ iff } M, t \models \phi \text{ for all } t \text{ such that } sRt.$$

The basic idea of this conceptualization has often been used in formalization of obligations applied to propositions afterwards, even though many other deontic logic systems throw away the notion of the most ideal states.

The well-known Kangerian-Andersonian reduction (KA-reduction) provides a profound understanding on the idea of defining obligations in standard deontic logic (see Anderson (1958), Kanger (1970)). The reduction shows that an SDL-style obligation can be expressed by an alethic formula with a propositional constant. Kanger (1970) claims that "Ought A is true in the universe of discourse if and only if A is entailed by each non-utopical, complete and true welfare program for this universe." In other words, Op is true if and only if all morally good states satisfy p. Let a propositional constant D_s represent that all normative demands accepted by the state s are met. Let U be the universal alethic operator. Accordingly, we have the following equivalence:

$$M, s \models O\phi \text{ iff } M, s \models U(D_s \rightarrow \phi).$$

This indicates that 'it is obligatory to achieve ϕ in the state s' can be understood as 'all ideal states with respect to the state s satisfy ϕ'.

The approach of Anderson (1958) is equivalent to Kanger's and he uses a constant S_s which represents a sanction with respect to the state s. So we can translate $O\phi$ based on Anderson's proposal as follows:

$$M, s \models O\phi \text{ iff } M, s \models U(\neg\phi \to S_s).$$

If we take $D_s =_{def} \neg S_s$, it is easily obtained that $U(D_s \to \phi)$ is equivalent to $U(\neg\phi \to S_s)$ in normal modal logic. Kangerian-Andersonian reduction plays an important role in our research as well. It inspires us to reduce the deontic operators to some existing modal operators and a constant representing the ideal situations in Chapter 4 and 5.

1.2 Various deontic logics with different focuses

Standard deontic logic established the most widely used framework for the field based on Kripke semantics, of which the underlying idea inspires follow-up studies to characterize a obligation with respect to propositions as 'all ideal situations satisfy the state of affairs (denoted by the proposition)'. However, this imitation of classical alethic modal logic is unsatisfactory because of leading to several notorious paradoxes, such as Ross's paradox (disjunction paradox), Weinberger's paradox (conjunction paradox), contrary-to-duty paradox, etc. It is worth noting that these so-called 'paradoxes' in deontic logic are not same as those classical paradoxes in the theory of truth or set theory. Haack (1978) described paradoxes in the classical sense as 'contradictory conclusions are followed by apparently unexceptionable reasoning from apparently unexceptionable premises.' For example, the sentence 'this sentence is false' is true if and only it is false, and the sentence is false if and only if it is true. This example is known as the Liar paradox in the theory of truth. The contradiction is obtained from the clear meaning of the sentence and the unexceptionable semantic rules. However, the paradoxes in deontic logic does not lead to any contradiction. They are called paradoxes since some deontic logic systems derive some valid but counter-intuitive formulas. These paradoxes in deontic logic will be investigated particularly in Section 6.5.

Besides the paradox issues, standard deontic logic captures obligations in an extremely general way since it does not take some crucial characteristics of obligations into account, such as agency, time and knowledge. For lack of capability in expressing these elements, deontic logic has been developed out of different problematic concerns.

Deontic stit logic Stit logic originated from Belnap and Perloff (1988) and was further developed by Belnap et al. (2001). It is famous for being a logic of agency since it aims for capturing agents' abilities in choosing some actions to perform. Comparing several possible English expressions which contain the meaning of agency[1], '*see to it that*' (stit) is selected to paraphrase sentences with action-like verbs. Stit logic formula '$[i\ stit : \phi]$' represents that agent i sees to it that ϕ.

A stit model is established based on a tree-like branching time model. Each moment is a set of histories whose pre-segments are the same at the moment, but whose post-segments are different. Moreover, for each particular agent, every moment is finitely partitioned into subsets of histories to simulate finitely many available actions for the agent. In other words, taking a specific action is modelled as choosing its corresponding subset of histories. In this way, an agent sees to it that ϕ if and only if there is an available action for the agent such that all the histories included in the action satisfy ϕ.

Deontic logic and stit logic integration was originally provided for representing the notion of 'ought-to-do', which should be distinguished from 'ought-to-be'. If obligations are defined with respect to propositions, we call this type of obligations 'ought-to-be' obligations. So standard deontic logic is one of the deontic logics characterizing 'ought-to-be' obligations. As shown in Von Wright (1951), actions are also appropriate to play the role of obligations since normative sentences normally express that agents ought to take some actions. Therefore, stit logic naturally shows the potential to define deontic logic in a more action-based way involving agency. Belnap and Perloff briefly mentioned that an obligation of the agent i could be reformulated as $O[i\ stit : \phi]$ which means that it is obliged that i sees to it that ϕ, so as to externalize the flavor of agency underlying obligations. Belnap (1991) also found that stit logic can enhance the expressive power of classical deontic logic. He analyzed some deontic expressions, like 'could have done otherwise' and 'refraining from'.

Horty (1996) investigated deontic stit logic systematically. Assigning values (natural numbers) to every history in stit models, Horty extends stit models to utilitarian stit models on which all histories are comparable with respect to their values. Although there are several possible definitions of obligations in deontic stit logic, their basic ideas are the same: the agent i ought to see to it that ϕ if and only if the optimal available action for i guarantees ϕ. Different ways to define optimal actions lead to different definitions of obligations. But deontic stit logic provides an original approach to embed agency and actions into the definitions of obligations.

Preference-based dyadic deontic logic Deontic stit logic focuses on the aspect of showing the agency behind the obligations. Besides that, as mentioned above, standard deontic logic is also challenged by many paradoxes, especially the contrary-to-duty paradox which influenced the development of the studies on conditional obligations.

Conditional obligations are defined to describe obligations in a conditional manner, which means that all obligations only come into force when they are

[1]For instances, bring it about, make it the case, be responsible for the fact that, allow it to be the case that, take steps in order that, see to it that.

'triggered' by some conditions. For example, a doctor ought to treat someone under the condition that the person is ill. In standard deontic logic, a superficial way to formalize a conditional obligation is resorting to a material implication where the condition is a fact and the consequence is an obligation, e.g., $\psi \to O\phi$ which represents that 'it ought to be ϕ given the condition ψ'. This formulation has been abandoned due to the *contrary-to-duty* paradox (see page 22 in van der Torre (1997), Greenspan (1975)). Let us first see an example of contrary-to-duty paradox:

1. We ought to keep a promise. (Op)
2. It ought to be the case that if we keep the promise, we do not apologize. ($O(p \to \neg q)$ or $p \to O\neg q$)
3. If we do not keep the promise, we ought to apologize. ($O(\neg p \to q)$ or $\neg p \to Oq$)
4. We do not keep the promise. ($\neg p$)

Sentence 1 is an obligation. Sentence 2 is called compatible-to-duty obligation. Sentence 3 is contrary-to-duty obligation and Sentence 4 is saying that the fact is a violation of the obligation. According to our intuition, the above four sentences should be consistent and are independent from each other. However, any selection of formulations will lead to conflicting obligations by standard deontic logic: $\bigcirc\neg q \wedge \bigcirc q$, or two non-independent formulas, i.e., $O(\neg p \to q)$ can be derived from Op.

The crucial weakness of SDL-style formalization is hidden in contrary-to-duty obligations. As mentioned in Section 1.1, the monadic obligation $O\phi$ in standard deontic logic represents that in all most ideal situations, the proposition ϕ is true. However, a contrary-to-duty obligation indicates the obligation under some *sub-ideal* situations. For example, sentence 3 says that we ought to apologize if the situation is not the most ideal (the promise has already not been kept), rather than saying that my apology is always performed in the most ideal situation where I keep my promise.

Considering the inappropriate formulation, the dyadic deontic operator comes forth. It normally represents some conditional obligation by formula $\bigcirc(\phi|\psi)$ which intuitively denotes that 'it ought to be ϕ given the condition ψ' (see Von Wright (1956), Van Fraassen (1973), Horty (1993), Kooi and Tamminga (2008)). One earliest and influential picture for the dyadic obligation is given by Hansson (1969) where the conditional obligations are interpreted over preference-based models. The semantics of $\bigcirc(\phi|\psi)$ is: the best ψ-states are ϕ-states.

Hansson realized that a simple dichotomy between ideal and non-ideal situations results in the failure of standard deontic logic in contrary-to-duty obligations. Preference-based models can be normally defined as a tuple $\langle W, \leqslant, V\rangle$ where $\leqslant$ is a partial order over the domain W. The relation $\leqslant$ is also called betterness relation since $s \leqslant t$ represents that the state t is at least as good as the state s. In this way, there is a betterness ordering over the realm of states and, consequently, we can find the most ideal, sub-ideal, sub-sub-ideal $\cdots$ states. The condition ψ in a dyadic obligation $O(\phi|\psi)$ restricts a model to these ψ-states and we only need to focus on the best states among this subset, instead of the whole domain. Therefore, the formula $O(\phi|\psi)$ successfully captures the sub-ideal situations by restrictions on the concerning subset.

There has been seen plenty of articles following Hansson's approach (see Lewis (1974), Prakken and Sergot (1997), van der Torre and Tan (1999), van Benthem et al. (2014)). A full-fledged study on Hansson's dyadic deontic obligation can be found in van der Torre (1997). He parsed dyadic obligations $O(\phi|\psi)$ as a preference of $(\psi \wedge \phi)$ over $(\psi \wedge \neg\phi)$. And then $O(\phi|\psi)$ can be defined by the classical modal operator for the binary relation $\leqslant$. In other words, the semantics of $O(\phi|\psi)$ is reduced to a formula only consisting of $\Box$ operators.

1.3 Interactions between obligations and knowledge

In the previous section, we introduced the deontic stit logic, which is originally designed for formalizing the agency in obligations and the dyadic deontic logic which aims to formalize conditional obligations in a proper way. Another important element related to obligations is knowledge. The possible combinations of knowledge and obligations have been studied for decades from various aspects. Apart from the notion of knowledge-based obligation which is to be investigated systematically in this thesis, there are many other notions involving both obligation and knowledge. We will briefly review them in this section, which is partly inspired by Başķent et al. (2012). We use symbols O and K to represent obligation and knowledge respectively in this chapter. The letter p denotes a proposition.

Ought to know The concept of 'ought to know' is usually known as *epistemic obligation* (OKp). The studies on 'ought to know' mainly focus on the solutions to the famous *epistemic obligation paradox*. We give an example of this paradox from one of the earliest works by Åqvist (1967):

1. It ought to be that Smith refrains from robbing Jones.

2. I ought to know that Smith robs Jones.

These two intuitively consistent sentences can be formalized as $O\neg p$ and OKp respectively. However, by standard deontic logic (to be introduced in the following chapter), they lead to a conclusion $Op \wedge O\neg p$ which means that it ought to be the case that Smith robs Jones and it is also ought to be the case that Smith refrains from robbing Jones.

Feldman (1990) gave a simple solution to this paradox by introducing temporal subscripts into obligation operators. Specifically, if we pick too late a time when Jones has already been robbed by Smith, there should be no possibility that Smith ought to refrain from robbing Jones. In contrast, if we pick too early a time before Smith robs Jones, I do not have obligation to know that Smith robs Jones. Hulstijn (2008) reformulated the paradox with logic of questions and replaced classical 'knowing that' operator with 'knowing-wh' operator. 'I ought to know that Smith robs Jones' (OKp) is reformulated as 'I ought to know whether Smith robs Jones' ($OK?p$). In his logic, the contradiction is no longer derived.

Besides these studies on the epistemic obligation paradox, Lomuscio and Sergot (2003) extended classical interpreted systems in computer science into *deontic interpreted systems*, where they introduced 'allowed' and 'disallowed' states to define the obligations. When knowledge is defined classically, epistemic obligation can be formalized as $OK_i\phi$ which means that it ought to be that i knows that ϕ.

In order to distinguish descriptive and prescriptive use of normative sentences, Aucher et al. (2009) defined an epistemic deontic logic where propositions and practitions (norms) are treated differently in syntax and semantics. Knowledge (or belief) in their language can be expressed in two ways where Bp in the propositional form means that 'the agent knows (believes) that p' whereas $B'p$ in practitional form means that 'to know (believe) that p'. And obligation operator O is only able to quantify norms. Thus, their language can express some formula like $OB'p$ which means that it is the obligation for the agent *to know* that p.

Know what obligation is Intuitively, the interaction in this category can be formalized as 'KOp' which normally means that the agent knows that p is an obligation.

Following Lomuscio and Sergot (2003) mentioned above, besides epistemic obligation, they also defined K_iO_jp with the normal meaning that i knows that j is obliged to achieve p.

Broersen (2008) discussed some interactions between 'obligation to do' and 'knowingly doing' in the framework of stit logic. He defined the agent i's obligation as 'i ensures that p is true over all next moments reached by the current moment'. Meanwhile, Broersen also defined knowledge in Stit frames in the classical way. Thus, we can also express 'i knows that him/herself is obliged to ensure p' in his language.

Aucher et al. (2009) used formula BOp to denote that the agent knows that it is obligatory for him to achieve p. Moreover, as their language introduced the knowledge operator both in a propositional and a practitional way, it can constitute two different formulas representing 'know what obligation is': BOp and $B'Op$. The former one denotes an indicative formula which represents that 'the agent knows (believes) that it is obligatory to achieve p'. The latter should be understood as an imperative which prescribes the agent *to know* that it is obligatory to achieve p.

In terms of speech acts, like commanding, requesting, committing, etc. would affect the agent's knowledge and obligation. Yamada (2006) focuses on *request* and constructed a dynamic logic of knowledge and obligation for characterizing the acts of requesting. In his multi-agent epistemic deontic logic, there is a type of formulas which characterizes that after the agent i commands (requests) the agent j that p, j will know that j has an obligation to i that p. In other words, i is the obligation addresser and j is the obligation addressee. This formula shows that after an act of requesting, the request addressee knows his/her new obligation.

Another paper on deontic logic which introduced dynamic information is given by van Benthem et al. (2014). After introducing public announcement operator with the same semantics as the classical $\mathbb{PAL}$, they can capture (make valid) one important principle in deontic logic: *factual detachment* by the way of conditional obligations. Intuitively, if it is obligatory for the agent i to achieve p under the condition q, then, after announcing q we will know that q is true and the agent also knows that it is obligatory to achieve p unconditionally.

Knowing-strategy-based responsibility De Lima et al. (2010) distinguished the notions of obligation and responsibility. The two concepts are not two sides of one coin anymore. They defined 'the agent i ought to achieve p' as 'over all outcomes (after arbitrary execution), if $\neg p$ is achieved, i will meet a violation'. It is apparent that no epistemic elements are introduced into this definition. However, they defined responsibility with knowledge in an inductive way. Briefly speaking, the inductive definition means that i is (forward-looking) responsible for p if and only if it is obligatory for i to keep the knowledge of how to ensure p. In other words, De Lima et al. (2010) declared that some agent should be responsible for some outcome, only if he/she knows explicitly about the strategy of achieving this outcome.

Obligation-based knowledge The concept of *obligation-based knowledge* is not investigated very broadly. The only paper (as far as I know) that mentions this concept is from Lomuscio and Sergot (2003). They defined a formula '$\widehat{K}_i^j p$' whose semantics is that the agent i knows that p on the assumption that the agent j is performing morally. This semantics is also equivalent to 'i knows that if j complies with obligation, then ϕ is true'.

In the following section, the background of the core notion of this thesis, knowledge-based obligations, is to be introduced.

1.4 Knowledge-based conditional obligation

Among various possible interactions between knowledge and obligation, this thesis focuses on the notion of knowledge-based obligation. It is a type of obligations in which knowledge is embedded. The idea of the notion comes from an observation that an agent's knowledge (information) may affect the agent's obligation. Pacuit et al. (2006) gave an example for knowledge-based obligation in our real life:

- Uma is a physician whose neighbor is ill. Uma does not know and has not been informed. Uma has no obligation to treat her neighbor. But once Uma is informed by her neighbor's families, Uma is obliged to treat her neighbor.

They defined a knowledge-based deontic logic on tree-like but *history-based* models. The histories in their models consist of events from a set $\{E_i \mid i \in G\}$, instead of moments that stit models consist of. For any two global histories H and H', if they have the same pre-segment at the moment t, then the agent i cannot distinguish the pair (H, t) and (H', t). What makes the approach different is that Pacuit and his colleagues introduced actions explicitly into the language and defined $G(a)$ as 'action a is a morally good action'. With the concept of 'good action', 'the agent i is obliged to perform action a' is interpreted as 'i can perform a and i knows that a is good'. Thus, the authors did not define obligation directly and make it reducible into other operators. In other words, if you bear some knowledge-based obligation to take an action, you must know that the action is morally good.

Broersen (2008) also dealt with some similar examples involving knowledge-based obligations. Broersen defined the notion based on stit logic following the idea of Kangerian-Andersonian reduction: $OK[i\ xstit]\phi =_{def} \Box(\neg K_i[i\ xstit]\phi \rightarrow [i\ xstit]V)$ (the constant V is taken for the same purpose as Anderson's constant S with the meaning of violation). From Broersen's point of view, it is (epistemically) obligated for the agent i to see to it that ϕ if and only if when the agent i does not know that he/she ensures ϕ in the current state (although he/she factually ensures ϕ), the agent i will bring about a violation. It should be noticed that, although Broersen expressed this knowledge-based obligation with '$OK[i\ xstit]$' which looks similar to the expression of 'epistemic obligation' ($\bigcirc K$), it substantially shows the meaning that the agent cannot bear some obligation unless he/she knows it. Broersen further proposed an even stronger knowledge-based obligation as: $K \bigcirc K[i\ xstit]\phi =_{def} \Box(\neg K_i[i\ xstit]\phi \rightarrow K[i\ xstit]V)$. In this definition, the agent i will knowingly bring about a violation if he/she does not comply with his/her obligation.

In the systematic study on the deontic stit logic, Horty (2019) defined a notion of epistemic oughts by extending utilitarian stit frames to *labeled stit semantics* where action types [2] are added. Intuitively, the agent i has an epistemic ought

[2] Horty and Pacuit (2017) specifically study the notion of action types in stit semantics.

to see to it that ϕ if and only if i knows that which action type is morally good and it always leads to ϕ over all epistemically indistinguishable moments.

In summary, Section 1.3 and this section review literature on possible interactions between knowledge and obligation. The following chart shows which paper investigates which types of interactions.

Interaction / Paper	$\bigcirc K$	$K\bigcirc$	K-based $\bigcirc$	KS-based $\bigcirc$	$\bigcirc$-based K
Åqvist (1967)	√				
Feldman (1990)	√				
Hulstijn (2008)	√				
Lomuscio and Sergot (2003)	√	√			√
Aucher et al. (2009)	√	√			
Broersen (2008)		√	√		
Yamada (2006)		√			
van Benthem et al. (2014)		√			
Pacuit et al. (2006)			√		
Horty (2019)			√		
De Lima et al. (2010)				√	

The column filled by gray color shows the three papers that we found on knowledge-based obligations. This thesis is also partly motivated by these works, but we follow a different approach. We will investigate the knowledge-based obligations in a conditional manner, which means that we will embed epistemic elements into conditional obligations. As illustrated in Section 1.2, Hansson's preference-based dyadic deontic logic established a satisfactory framework as basis for conditional obligations.

Let us review Hansson's preference-based models: $M = \langle W, \leqslant, V \rangle$. The partial order $\leqslant$ is a binary relation between states from W where $s \leqslant t$ represents that the state t is at least as good as the state s. Therefore, in the remaining part of this thesis, we also call Hansson's preference-based models betterness structures to fit in the deontic context. If we only consider the single-agent case, a betterness structure consists of a set of states for the agent, a valuation on each world and a betterness relation between worlds with respect to the agent's deontic rules. An obligation of the agent is a common proposition satisfied by all the best worlds with respect to $\leqslant$. Briefly speaking, we have the following two claims:

- (W, V) provides a set of possible situations.
- $\leqslant$ provides the criteria on the betterness between these situations.

If we want to proceed with our investigation on knowledge-based conditional obligation, how to combine epistemic elements with betterness structures? In classical epistemic logic, an epistemic model is a tuple $M_E = \langle W, \sim, V$

(single-agent case) where $\sim$ is a binary relation to show the epistemic indistinguishability of the agent between possible worlds. Similarly, we also have two claims about epistemic models:

- (W, V) provides a set of possible situations.
- $\sim$ shows the agent's epistemic indistinguishability between these situations.

Therefore, it seems very natural to put all these ingredients together and construct an epistemic betterness structure i.e., $M_{EB} = \langle W, \leqslant, \sim, V \rangle$ where our logic of knowledge-based conditional obligations will be built. The investigation on this new notion is to to be shown in Chapter 3.

However, an agent's obligations are not immutable. In terms of knowledge-based conditional obligation, the changes on knowledge may also bring about the changes on obligations. Since epistemic betterness structures essentially are extensions to epistemic models, dynamic epistemic logic should be the first alternative tool used for capturing epistemic change in the context of knowledge-based conditional obligations.

Dynamic epistemic logic is famous for action models which can be components of both syntax and semantics. On the semantic side, an action model is a mechanism changing the epistemic relation $\sim$ or updating the valuation V in an epistemic model. Changing epistemic relations captures the agent's information changes[3], whereas updating the valuation characterizes factual changes. Since the definition of the knowledge-based conditional obligations are decided by both epistemic relations and valuations, the two kinds of changes might lead to new obligations. Chapter 4 mainly discusses obligation change due to epistemic and factual changes.

Figure 1.1 visualizes the whole background of our study on the knowledge-based conditional obligations as a summary of the previous sections.

1.5 Prescription and norm change

So far we have introduced deontic logics in which only the descriptive aspects of obligations are concerned. These logics are merely about agents' obligations. The obligations belong to a special type of propositions which specially describe deontic facts. In contrast, it is inappropriate to interpret all normative sentences to propositions (obligations) since many evaluative sentences prescribe norms rather than describing deontic facts. See the sentence below:

Due to the COVID-19 pandemic, you ought to keep a good ventilation.

[3] In this thesis, the two terms 'epistemic change' and the term 'information change' refer to the same concepts. The term 'epistemic change' will be used more often in technical contexts and the term 'information change' will occur more frequently in informal contexts.

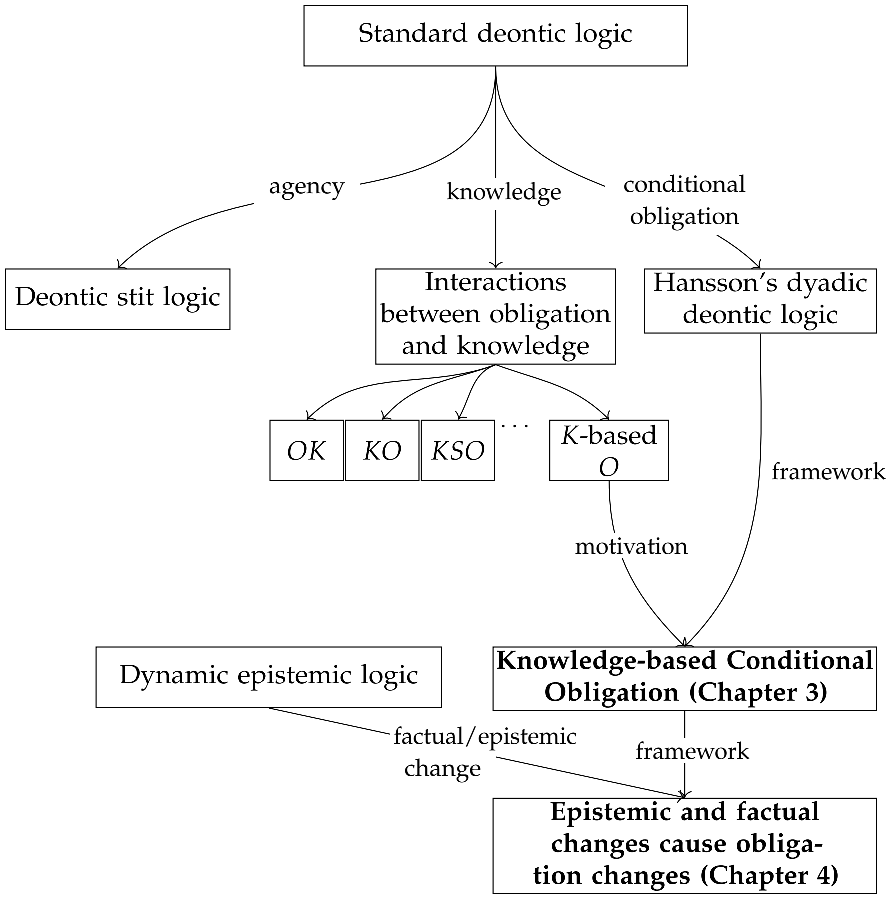


Figure 1.1: Background of knowledge-based obligations studied in this thesis (each label on arrows represents the factor that leads to the birth of the target logic)

The descriptive reading of the sentence implies that there are already some public rules requiring all people to keep a good ventilation and as a consequence, it is a fact that you ought to follow the rules. Whereas the prescriptive reading means that the sentence is announced by some authority (maybe a government health department) to prescribe a new rule that everyone ought to keep a good ventilation. It is obvious that the prescriptive reading translates the sentence to an evaluative sentence which is neither true nor false. If an evaluative sentence cannot be assigned truth values, is it possible to establish a logic of evaluative sentences? This issue is well-known as *Jørgensen's dilemma*. It actuates the differentiation between description and prescription in deontic logic. More precisely, they generally denote normative propositions (description) and norms (prescription) respectively.

As mentioned in Section 1.3, Aucher et al. (2009) defined an epistemic deontic logic where propositions and practitions (norms) are treated differently in syntax and semantics to distinguish description and prescription. A more influential logic is *input/output logic* which is designed specifically for dividing norms and normative propositions, the latter of which are actually obligations (see Makinson and Van Der Torre (2000) and Makinson and van der Torre (2007a)). Input/output logic actually is not an axiomatization based on some semantics in the classical sense. It instead provides a very general framework for all logic systems that treat norms and obligations separately. The input of input/output logic is a set of propositions which represents several facts. The core of the logic is a set of conditional norms which is just like an algorithm which can output obligations based on what facts are input. Therefore, the norms in input/output logic are not propositions. Rather, they form a transformation machine to produce obligations from facts.

Hansen (2008) argues that there is *no* logic of imperatives. But he claims that "imperatives still can be meaningfully used to determine what obligations arise in a certain situation", which leads us to a logic *about* imperatives. The concept of prioritized imperative structure is put forward by Hansen (2006) so as to show an ordered set of imperatives which do not have truth values but can provide norms for obligations. A prioritized imperative structure is a tuple $\langle I, f, < \rangle$ where I is a set of imperatives, f is a function mapping each imperative to their corresponding descriptive sentences and $<$ is a binary relation to show which imperative is more important. This structure inspires a key notion in this thesis – priority structures which was originally proposed by Liu (2008) and van Benthem et al. (2014). We will use it as the normative system in Chapter 4, 5 and 6. A normative system provides a criteria on which states of affairs are better or worse. Obligations of agents are dependent on the normative system.

Besides the epistemic change and the factual change mentioned in Section 1.4, norm change would affect obligations more substantially. Chapter 5 will show several different updates on priority structures and obligations are changed accordingly. A logic of conditional obligations relativized to normative systems is established.

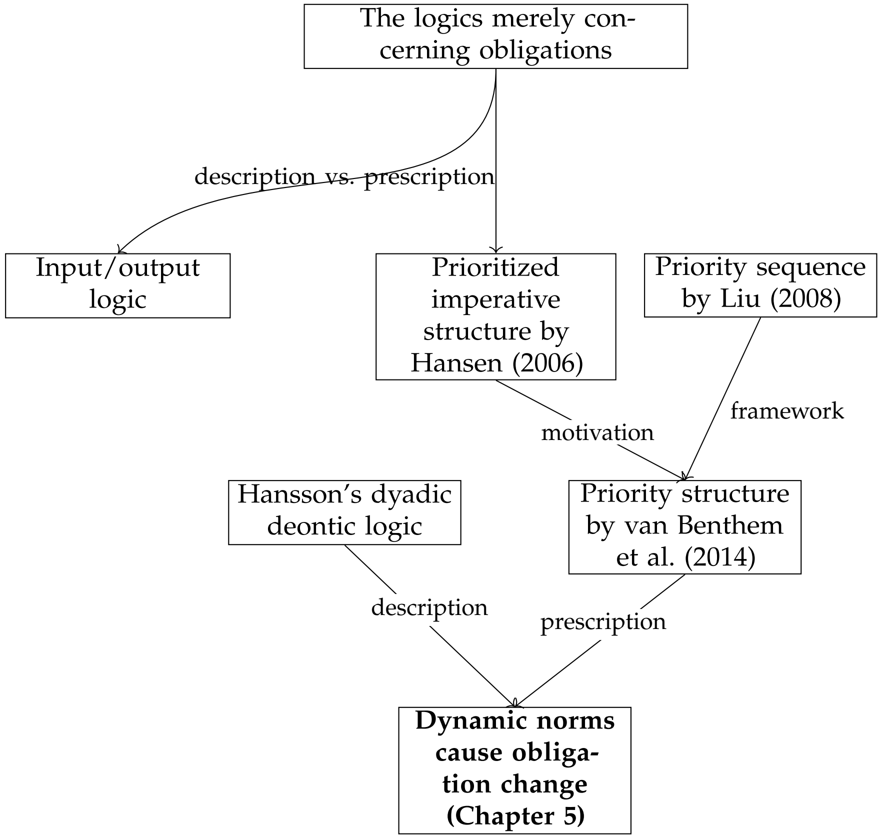


Figure 1.2: Background of Chapter 5

Figure 1.2 shows the background of our research on prescription and description to be investigated in Chapter 5.

1.6 Actions and 'ought-to-do'

It is widely agreed that obligations are descriptive since they merely describe agents' deontic states rather than prescribing some new norms. Obligations themselves are propositions. However, what a deontic operator should be applied to has been controversial since the beginning of deontic logic. As mentioned in Section 1.1, Von Wright (1951)'s first deontic logic takes actions as primary, which means that his deontic operator O is applied to actions instead of propositions. But standard deontic logic applies the deontic operator to propositions, which also makes a great impact on the following deontic logics. Until Meyer (1988) established a variant of deontic logic based on dynamic logic, researchers were considering the issue whether obligations should be proposition-based or action-based. Deontic stit logic was developed out of the idea that obligations should involve actions performed by agents. But it does not really make actions explicit in their language and their deontic formula is still read as 'some agent ought to see to it that some proposition'.

In the context of knowledge-based obligations, is it possible to construct a deontic logic based on actions? Dynamic epistemic logic immediately comes into our mind since action models which are designed specifically for modelling actions and the epistemic information is also combined. Moreover, action models in dynamic epistemic logic can be both parts of syntax and semantics, which means that it is possible to add them into our language to show which actions ought to be done.

But how to build a deontic logic based on dynamic epistemic logic deserves an elaboration. Dynamic epistemic logic extends the classical epistemic logic by introducing action models to show the transitions between two epistemic models. Each epistemic model is a fully description on the epistemic states of a group of agents at a certain moment. An action is a transition mechanism to update one epistemic model to a new epistemic model. This is the underlying strategy of dynamic epistemic logic to simulate the epistemic and factual changes.

Then we look back on Meyer (1988)'s dynamic deontic logic. From his perspective, an action ought to be done if and only if all consequences led to by the action are morally good. So if we graft his idea to the context of dynamic epistemic logic, we will probably be able to define that an action ought to be done if and only if all updated epistemic models derived via the action model are better than the initial epistemic model. Thus, the only problem left is how to compare two epistemic models and judge which one is morally better. In this way, priority structures play their role again as the criteria on betterness. Chapter 6 will mainly focus on these issues and propose a new logic of knowledge-based 'ought-to-do' obligations.

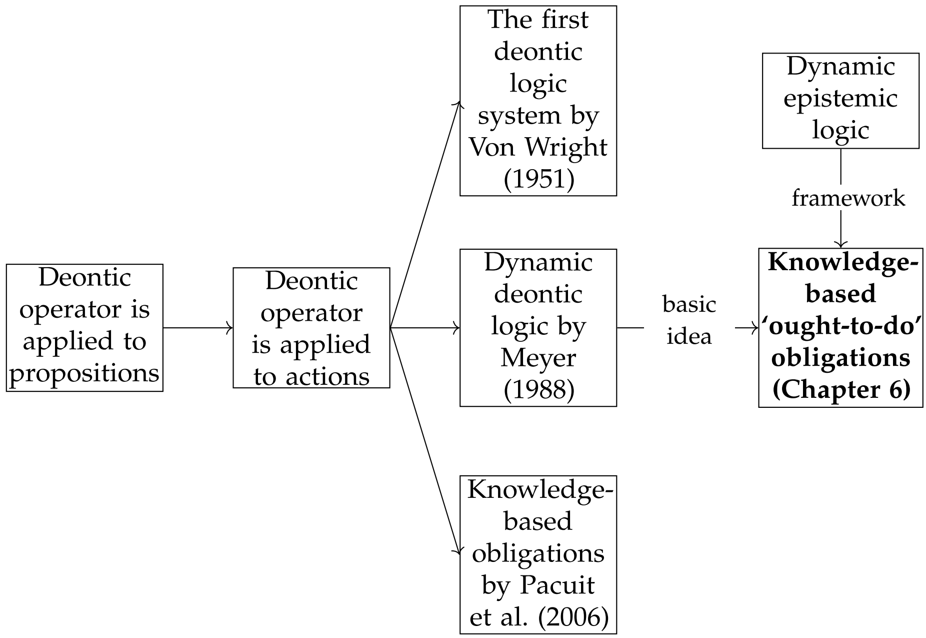


Figure 1.3: Background of Chapter 6

Figure 1.3 shows the background of our study on the knowledge-based 'ought-to-do' obligations which will be investigated in Chapter 6.

1.7 Outline of the chapters

The rest of this thesis is structured as follows.

- In **Chapter** 2, we present technical preliminaries for reading the rest of the thesis. The first and also the most important is Hansson's dyadic deontic logic which provides the technical basis for our framework. Then classical epistemic logic will also be shown. Next we introduce the notion of priority structure. It gives the relative ordering on several states of affairs. Finally, dynamic epistemic logic is shown to explain how action models change epistemic models. The approach that how to formalize information and factual change will play an important role in chapter 4 and Chapter 6.

- In **Chapter** 3, we present a new dyadic deontic operator to formalize the notion of *knowledge-based conditional obligation* in a static manner. Based on epistemic betterness structures, the agent's knowledge plays an important role in the definition of epistemic conditional obligation. By giving an explicit reading of this new operator and comparing it with objective conditional obligation, some characteristics of epistemic conditional obligations will be clarified. In the light of *epistemic detachment*, some real-life scenarios involving knowledge-based obligations can be described formally with our operator. A sound and strongly complete logic $\mathbb{KCDL}_1$ for single-agent knowledge-based conditional obligations is provided. The strategy for completeness proof comes from Parent (2014). A sound and strongly complete logic $\mathbb{KCDL}$ for multi-agent case is also provided. The completeness proof is constructed with the help of step-by-step method. This chapter is an extended version of a joint work with Davide Grossi, Barteld Kooi and Rineke Verbrugge, which has been published as:

  Xingchi Su. Knowledge-based conditional obligation. *Short Papers Advances in Modal Logic (AiML) 2020*, pages 112–116, 2020.

- In **Chapter** 4, we provide a formalism to capture the *dynamic* interaction between knowledge and obligations. We introduce the dynamic extension of the static knowledge-based obligations $\mathbb{DKCDL}$. We motivate the logic by analyzing several scenarios and by showing how it can capture in an original manner several fundamental deontic notions such as absolute, *prima facie* and all-things-considered obligations. Finally, in the dynamic epistemic logic tradition, we provide reduction axioms for the dynamic operator and Kangerian-Andersonian reduction for the dyadic deontic operator, on which the strong completeness of $\mathbb{DKCDL}$ is established. This chapter has been previously published as:

Davide Grossi, Barteld Kooi, Xingchi Su, and Rineke Verbrugge. How knowledge triggers obligation. In *International Workshop on Logic, Rationality and Interaction*, pages 201–215. Springer, 2021.

- **Chapter** 5 temporarily releases the context of knowledge. We provide a formalization of conditional obligations relativized to normative systems. The conditional obligation still describes what state of affairs that the agent ought to see to it that. However, differently from previous work, each conditional obligation appears in accordance with some certain normative system. Prescribing norms can be reflected on the updates on the normative system. In this way, we distinguish the descriptive and prescriptive uses of normative sentences and show the obligation changes due to norm changes. Based on the notion of successful updates, the Jørgensen's dilemma can be conceptualized in a novel and proper way.

- **Chapter** 6 develops knowledge-based obligation following the approach of 'ought-to-do' obligations. Obligations are formalized as compound actions in dynamic epistemic logic. A compound action can be a pointed action model or a non-deterministic choice among several action models. An obligation is able to change the current epistemic situation (the initial epistemic model) and bring it about updated epistemic situations where each consequence models are better than the initial one. A sound and strongly complete axiomatization $\mathbb{AKDL}$ is provided as a logic of knowledge-based 'ought-to-do' obligations. The approach of defining obligations as actions resolves several notorious paradoxes discussed in deontic logic, such as Ross's paradox, Weinberger's paradox, Forrester's paradox of gentle murder, etc.

Chapter 2

Preliminaries

In order to characterize knowledge-based conditional obligations and possible dynamics on them, we take Hansson's betterness structures as the basic framework. Epistemic logic will also be introduced for formalizing knowledge. Priority structures from Hansen (2006) are criteria on comparing states and therefore can determine betterness relations in betterness structures. Dynamic epistemic logic is to be introduced as a preparation for capturing information and factual change that would cause obligation change. Each logic or notions have been investigated adequately in their own fields. We only show the most standard aspects of these formal works as technical preliminaries for the rest of this thesis.

2.1 Hansson's dyadic deontic logic

Some philosophical background and the basic idea of Hansson's dyadic deontic logic have been introduced in Section 1.2. We will focus on the technical aspect in this section and give the semantics and syntax formally.

2.1.1 Language and semantics

The language for Hansson's dyadic deontic logic is $\mathcal{L}_{\text{DDL}}$, where DDL is an acronym for 'dyadic deontic logic'. Let $\mathbf{P}$ be a countable set of propositional letters.

Definition 1 (Language $\mathcal{L}_{\text{DDL}}$). *The language of Hansson's dyadic deontic logic is given by the following BNF:*

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid \Box\phi \mid \bigcirc(\phi|\phi)$$

where $p \in \boldsymbol{P}$.

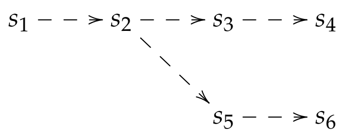


Figure 2.1: A betterness structure

The formula p is a propositional atom which represents the minimal or primary state of affairs. The formula $\neg\phi$ is the negation of ϕ. The formula $(\phi \wedge \psi)$ is the conjunction. The formula $\Box\phi$ is the universal formula which means that ϕ is always true or ϕ is true in all situations. The formula $\bigcirc(\phi|\psi)$ represents the conditional obligation which can be read as 'it ought to be ϕ given ψ'.

We mentioned in Section 1.2 that the language is interpreted over so-called *betterness structures*. Betterness structures are essentially assessment systems to order all the states, which can tell which states are *ideal* and which are sub-ideal or sub-sub-ideal, etc. The betterness relation $\leqslant$ can be defined based on different ethical theories, such as consequentialism/utilitarianism (see Mill (1859)), deontological theory (see Kant (2002, 2005)), social contractarianism (see Hobbes (1914), Rousseau (1795)), or contractualism (see Scanlon et al. (1998)).

Definition 2 (Betterness structures (Hansson (1969)))**.** *A betterness structure M is a triple of a set of states S, betterness relation $\leqslant: S \times S$ which is a partial order, and a valuation $V : \boldsymbol{P} \to \mathcal{P}(S)$ over S: $M = \langle S, \leqslant, V\rangle$.*

An example of a betterness structure is figured as Fig. 2.1. In the figure, $s_i \dashrightarrow s_j$ denotes $s_i \leqslant s_j$ for any $i, j \in \mathbb{N}$, which informally means that s_j is at least as good as s_i.

Several notations to be used in the rest of the thesis are also given here:

- $\|\phi\|_M = \{s \mid s \in S, M, s \models \phi\}$,
- $s < t$ iff $s \leqslant t$ and $t \nleqslant s$.

In Hansson's tradition, the *best* states in a particular subset of S with respect to $\leqslant$ are crucial if we want to define the semantics of the dyadic deontic formulas. The notion of *maximal elements* in one set with respect to some ordering can be used for capturing the best states in some subset.

Definition 3 (Maximal elements)**.** *Given a betterness structure $M = \langle S, \leqslant, V\rangle$ and $T \subseteq S$,*

$$s \in \max_{\leqslant} T \textit{ iff } s \in T \textit{ and } \forall t \in T (s \leqslant t \Rightarrow t \leqslant s)$$

Informally: within T, these are states such that no other state is strictly better in T. Therefore these states are "as good as it gets" within T.

Now we can give the semantics for Hansson's conditional obligations.

Definition 4 (Semantics of $\mathcal{L}_{\text{DDL}}$ (Hansson (1969))). *Let $M = \langle S, \leqslant, V \rangle$ be a betterness structure. The semantics of $\mathcal{L}_{\text{DDL}}$ is defined as follows:*

$$\begin{array}{lll} M, s \models p & \textit{iff} & s \in V(s). \\ M, s \models \neg\phi & \textit{iff} & M, s \not\models \phi \\ M, s \models (\phi \wedge \psi) & \textit{iff} & M, s \models \phi \textit{ and } M, s \models \psi \\ M, s \models \Box\phi & \textit{iff} & S = \|\phi\|_M. \\ M, s \models \bigcirc(\phi|\psi) & \textit{iff} & \max_{\leqslant}\|\psi\|_M \subseteq \|\phi\|_M \end{array}$$

Note that here both $\Box$ and $\bigcirc(\_|\_)$ are global modalities. The intuition of Hansson's conditional obligation is: all the best states satisfying ψ in M also satisfy ϕ. We will call it an *objective conditional obligation* in Chapter 3 to distinguish it from our epistemic conditional obligation since it does not depend on any agent's information.

Given the definition of the betterness relation, there is no guarantee that the set of maximal elements of some non-empty $\|\phi\|_M$ is non-empty as well. It can be the case that one can always find a strictly better state and never reach a maximal state. In order to exclude these cases, two alternative properties on betterness structures have been proposed by Parent (2014).

Definition 5. *Let $M = \langle S, \leqslant, V \rangle$ be a betterness structure. We define two properties of $\leqslant$ as follows:*

- *(Limitedness) if $\|\phi\|_M \neq \varnothing$, then $\max_{\leqslant} \|\phi\|_M \neq \varnothing$;*
- *(Smoothness) if $M, s \models \phi$, then either $s \in \max_{\leqslant} \|\phi\|_M$ or $\exists t : t > s$ and $t \in \max_{\leqslant} \|\phi\|_M$.*

Limitedness is the most straightforward property that guarantees non-emptiness. Smoothness guarantees the existence of maximal elements in a direct way. The idea of smoothness can be found in some early research on nonmonotonic reasoning where it is originally called *minimal modelability* by Bossu and Siegel (1985). Smoothness defined here is identical to the *smoothness condition* defined by Kraus et al. (1990). Their original definition of smoothness is defined only with respect to some particular subset of the model, i.e., $\|\phi\|_M$ is smooth, which means that for each $t \in \|\phi\|_M$, either t itself is maximal in $\|\phi\|_M$ or there exists some $s > t$ such that s is maximal in $\|\phi\|_M$.[1]

2.1.2 The axiom system

Since Parent (2014) systematically studied the axiomatization of Hansson's dyadic deontic logic, we mainly illustrate Parent's contributions in this section.

[1]If both $\|\phi\|_M$ and $\|\psi\|_M$ are smooth, then $\|\phi \vee \psi\|_M$ is smooth. However, if $\|\phi\|_M$ is smooth, it is not necessary that $\|\neg\phi\|_M$ is smooth. As a consequence, in Kraus et al. (1990)'s contexts, smoothness is not closed under the operations of Boolean algebras. Although smoothness is closed under finite disjunctions, it is not closed under infinite disjunctions.

Definition 6. *The axiom system **F**+(CM) consists of following axiom schemas and inference rules:*

(TAUT)	*All instances of tautologies*
$(\mathcal{S}5)$	$\mathcal{S}5$*-schemata for* $\Box$
(COK)	$\bigcirc(B \to C \mid A) \to (\bigcirc(B \mid A) \to \bigcirc(C \mid A))$
(Abs)	$\bigcirc(B \mid A) \to \Box \bigcirc (B \mid A)$
($\bigcirc$Nec)	$\Box A \to \bigcirc(A \mid B)$
(Ext)	$\Box(A \leftrightarrow B) \to (\bigcirc(C \mid A) \leftrightarrow \bigcirc(C \mid B))$
(Id)	$\bigcirc(A \mid A)$
(Sh)	$\bigcirc(C \mid A \wedge B) \to \bigcirc(B \to C \mid A)$
(D*)	$\neg\Box\neg A \to (\bigcirc(B \mid A) \to \neg \bigcirc (\neg B \mid A))$
(CM)	$(\bigcirc(B \mid A) \wedge \bigcirc(C \mid A)) \to \bigcirc(C \mid A \wedge B)$
(MP)	*If* $\vdash A$ *and* $\vdash A \to B$, *then* $\vdash B$
(N)	*If* $\vdash A$, *then* $\vdash \Box A$

The system **F**+(CM) given by Parent (2014) extends Åqvist (1987)'s system **F**. The axiom (COK) is the distribution axiom for the operator $\bigcirc$. The axiom (Abs) reflects the assumption that the ordering over possible states is not dependent on any specific states. The axiom ($\bigcirc$Nec) is the deontic counterpart of the well-known necessity rule. The axiom (Ext) permits the replacement of the antecedent of the deontic conditionals with an equivalent antecedent. The axiom (Id) means that the best cases where A is true, are cases where A is true. The axiom (Sh), named after Shoham (1988b), corresponds to a weaker version of the deduction theorem of a nonmonotonic preferential logic mentioned in Shoham (1988a) and the derived rule (S) in Kraus et al. (1990). The axiom (D*) intuitively means that if A is possible and we also have that under the condition that A, the best states are B-states, then it is impossible that under the condition that A, $\neg B$ is the best. The axiom (CM) corresponds to the (*Cautious Monotonicity*) rule in Kraus et al. (1990).

The system **F**+(CM) was shown to be complete with respect to reflexive and smooth betterness structures and it is also complete with respect to reflexive, total, transitive and smooth (limited) structures.

2.2 Epistemic Logic

Epistemic logic formally studies the reasoning about epistemic notions, such as knowledge and belief. In a narrow sense, epistemic logic is mainly concerned with knowledge and its representation. Although there are many different types of knowledge as indicated by Wang (2018) in recent years, standard epistemic logic still focuses on 'knowing that' which can be paraphrased by propositional knowledge [2]. In other words, epistemic logic treats knowledge as propositions

[2]There are many non-propositional knowledge. The expression 'knowing how' in natural language describes the *process knowledge* discussed in artificial intelligence or epistemology. For example,

and therefore the reasoning between knowledge can be represented by the logical inferences on propositions.

Let G be a finite set of agents and let $\mathbf{P}$ be a countable set of propositional variables.

Definition 7 (Language $\mathcal{L}_{\text{EL}}$)**.** *The language of epistemic logic is given by the following BNF:*

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid K_i\phi$$

where $p \in \mathbf{P}$ and $i \in G$.

The formula $K_i\phi$ is normally used for denoting 'agent i knows that ϕ'. Since $\mathcal{L}_{\text{EL}}$ is defined in an inductive way, the formula $K_i\phi$ can be a first-order knowledge $K_i p$ where p is a propositional atom, or a higher-order knowledge $K_iK_j\psi$ which means i knows that j knows that ψ.

The most widely-used semantics for epistemic logic is the modal approach where Kripke models play a crucial role.

Definition 8. *A Kripke model for epistemic logic is a tuple $\langle S, \sim_1, \cdots, \sim_n, V\rangle$ where*

- *S is a set of states;*
- *for each $i \in G$, $\sim_i \subseteq S \times S$ is an equivalence relation over S;*
- *$V \subseteq \mathbf{P} \to \mathcal{P}(\mathbf{S})$.*

We also call these Kripke models *epistemic models*. The domain S consists of states. Each state provides the basis for evaluating all propositions that we are concerned with via their truth values. A state s and the epistemic model M where s comes from make up a possible world (M, s). In the rest part of this thesis, we will also call a state s possible world if its model is clear. In the earliest formal studies on epistemic logic by Hintikka (1962), possible worlds are also called epistemic alternatives on which the information cannot be indefensible (p and $\neg p$ cannot hold in one alternative). A possible world can be regarded as an adequate description on a given situation with respect to a certain valuation V. Every $\sim_i$ represents the indistinguishability relation for the agent i. Understanding the epistemic relations as indistinguishability relation was first proposed by Lehmann (1984). The indistinguishability interpretation brings about that epistemic relations must be reflexive, symmetric and transitive, thus equivalence relations.

Figure 2.2 is an example of epistemic models. In Figure 2.2, four possible worlds differ from each other since the two concerned propositions p and q have

'knowing how to ride the bike' is a process knowledge about ability, which is obviously not propositional. The logic of knowing how can be referred to Wang (2015). The expression 'knowing why' is related to the explanation knowledge in philosophy of science. A logical study on 'knowing why' is given by Xu et al. (2021). The expression 'knowing what' corresponds to *descriptive knowledge*. For example, 'knowing what the password is' which is different from 'knowing that the password is 1234'. The earliest formalization refers to Plaza (2007).

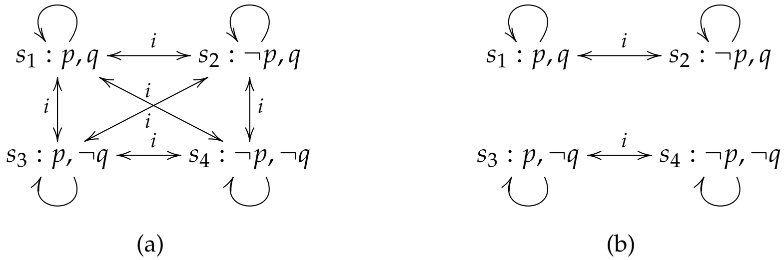


Figure 2.2: An example of epistemic models

different values over them. In Figure 2.2a, agent i cannot distinguish between any pair of the worlds, which represents that i neither knows the value of p nor q. But in Figure 2.2b, i can distinguish s_1 from s_3 and s_4, and also can distinguish s_2 from s_3 and s_4. It represents that i does know the value of q but i does not know the value of p.

The standard axiomatization of epistemic logic $\mathbb{EL}$ is established entirely based on the system $\mathbb{S}5$ in classical modal logic since epistemic relations are equivalence relation.

Definition 9. *The axiom system of* $\mathbb{EL}$ *consists of the following axiom schemas and inference rules:*

for each $i \in G$,

(TAUT)	*All instances of tautologies*
(K)	$K_i(\phi \to \psi) \to (K_i\phi \to K_i\psi)$
(T)	$K_i\phi \to \phi$
(4)	$K_i\phi \to K_iK_i\phi$
(5)	$\neg K_i\phi \to K_i\neg K_i\phi$
(MP)	*From* ϕ *and* $\phi \to \psi$, *infer* ψ
(N)	*From* ϕ, *infer* $K_i\phi$

The theories of $\mathbb{EL}$ are treated as characterization on principles of knowledge. The axiom (T) characterized reflexivity in modal logic. It is the only uncontroversial interpretation on knowledge: If agent i knows that ϕ, then ϕ is true. This validity perfectly expresses the *condition of truth*[3] of forming knowledge from the famous claim in epistemology 'knowledge is equivalent to justified true belief'. The axiom (K) says that if agent i knows that an implication is the case and i also knows that the antecedent is the case, then the consequence is also known by i. In other words, an agent's knowledge is closed under distributivity. The axiom (4) is also known as the principle of *positive introspection* which means that if

[3]There is a long-held justified-true-belief account of knowledge. It means that if all three conditions (justification, truth, belief) are met of a claim, we have knowledge of this claim (see Ichikawa and Steup (2018)).

agent i knows that ϕ is the case, then i also knows that he/she knows that ϕ. The *negative introspection*, the axiom (5), is not accepted by many philosophers since in most cases, people even cannot realize that he/she does not know something.

The soundness and strong completeness can be proved in the same way as modal logic S5 following the approach of canonical models (see Blackburn et al. (2002)). In this thesis, most completeness proofs, however, will not use the canonical model approach.

2.3 Epistemic plausibility models

As mentioned in Section 1.4, the core notion of this thesis, knowledge-based conditional obligations, is defined based on epistemic betterness structures, i.e., $M_{EB} = \langle W, \leqslant, \sim, V \rangle$ (single-agent case). Coincidentally, in doxastic logic, the epistemic plausibility models which are very similar to epistemic betterness structures, form an alternative semantic apparatus for modelling both knowledge and beliefs.

Definition 10 (Epistemic plausibility models Baltag and Smets (2006b)). *An epistemic plausibility model is a tuple $M_{EP} = \langle W, \leqslant_P, \sim, V \rangle$ where*

- *W is a set of states;*
- *$\sim \subseteq W \times W$ is the epistemic relation;*
- *$\leqslant_P \subseteq W \times W$ is the plausibility relation;*
- *$V \subseteq \mathbf{P} \to \mathcal{P}(\mathbf{S})$ is a valuation.*

For two states s and t, the relation $s \leqslant_P t$ holds if and only if s is at least as plausible as t. But there is a constraint on the plausibility relations: $\leqslant_P \subseteq \sim$, which means that $s \leqslant_P t$ implies $s \sim t$. In other words, only if the agent cannot epistemically distinguish two possible worlds, their plausibilities are comparable.

Analogous to the maximal elements in a betterness structure, there is a set of minimal states in an epistemic plausibility model, i.e., $\min_{\leqslant_P} S$ which represents the set of the most plausible states in $S \subseteq W$.

Definition 11 (Minimal elements). *Given an epistemic doxastic model $M = \langle W, \leqslant_P, \sim, V \rangle$ and $T \subseteq S$,*

$$s \in \min_{\leqslant_P} T \text{ iff } s \in T \text{ and } \forall t \in T (t \leqslant_P s \Rightarrow s \leqslant_P t)$$

Baltag and Smets (2006b) defined the notion of conditional beliefs based on epistemic plausibility models. The unconditional belief is a special conditional belief when the condition is a tautology. The formula $B^{\psi}\phi$ represents that the agent believes that ϕ after learning ψ. The semantics is given as follows:

$$M_{EP}, s \models B^{\psi}\phi \text{ iff } M_{EP}, t \models \phi \text{ for all } t \in \min_{\leqslant_P}([s]^{\sim} \cap \|\psi\|_{M_{EP}}).$$

The truth condition intuitively means that all the most plausible ψ-states that the agent epistemically cannot distinguish satisfy ϕ. We will see that this definition is very similar to our knowledge-based conditional obligation in Chapter 3.

2.4 Priority structure

The betterness relation in a betterness structure is given *a priori*. However, a priority structure is able to make a ranking for a set of states. Priority structures were originally introduced by Liu (2008) in the context of preference logic. But the idea of it in deontic logic can be dated back to Hansen (2006). We first introduce the notion of priority sequence and G-sequence.

Definition 12 (Priority sequence (Liu (2008)))**.** *A priority sequence is a finite ordered sequence of first-order formulas written as follows:*

$$C_1 \gg C_2 \cdots \gg C_n \ (n \in \mathbb{N})$$

According to Liu's explanations, they only think of finite domains, monadic predicates, simple formulas, usually quantifier free or even variable free. For example, the formula $C_n(x)$ represents that the object x has the property described by C_n. Based on a certain priority sequence, a preference ordering between objects can be derived as follows:

Definition 13 (Preference derived from priority sequence (Liu (2008)))**.** *Given a priority sequence of length n, and two objects x and y, $pref(x, y)$ is defined as follows:*

$$\begin{array}{lll} pref_1(x,y) & ::= & C_1(x) \wedge \neg C_1(y), \\ pref_{k+1}(x,y) & ::= & pref_k(x,y) \vee (Eq_k(x,y) \wedge C_{k+1}(x) \wedge \neg C_{k+1}(y)),\ k < n, \\ pref(x,y) & ::= & pref_n(x,y), \end{array}$$

where the auxiliary binary predicate $Eq_k(x,y)$ stands for $(C_1(x) \leftrightarrow C_1(y)) \wedge \cdots \wedge (C_k(x) \leftrightarrow C_k(y))$.

The above definition shows that the preference between two objects is decided by the highest formula for which they have different truth values in the priority sequence. Liu's priority sequence is actually very similar to the notion of ideality sequence to be introduced in Chapter 5. Now let us show a follow-up notion introduced by Liu which motivates the priority structures in deontic logic.

Definition 14 (G-sequence (Liu (2008)))**.** *A priority sequence $C_1 \gg C_2 \cdots \gg C_m$ gives rise to a G-sequence: $R_1 \geqslant R_2 \geqslant \cdots \geqslant R_{2^m}$ of length 2^m by the following way:*

$$(p \vee q) - - \succ p$$

Figure 2.3: $\mathcal{G}$

$$\begin{array}{lll} R_1 & : & C_1 \wedge C_2 \wedge \cdots \wedge C_m, \\ R_2 & : & C_1 \wedge C_2 \wedge \cdots \wedge C_{m-1} \wedge \neg C_m, \\ R_3 & : & C_1 \wedge C_2 \wedge \cdots \wedge \neg C_{m-1} \wedge C_m, \\ \vdots & & \\ R_{2^m-1} & : & \neg C_1 \wedge \neg C_2 \wedge \cdots \wedge \neg C_{m-1} \wedge C_m, \\ R_{2^m} & : & \neg C_1 \wedge \neg C_2 \wedge \cdots \wedge \neg C_{m-1} \wedge \neg C_m. \end{array}$$

It is obvious that for each $i, j \in \mathbb{N}$, if $i < j$, then R_i implies R_j, which is in line with the constraints of our core notion, priority structures. But priority structures are extensions of G-sequences since they are not only linear orders, but can also be any strict orders[4].

The domain of a priority structure is a finite set of propositional formulas. Let $\mathcal{L}_{\text{PL}}$ be the language for the propositional logic.

Definition 15 (Priority Structures (van Benthem et al. (2014)))**.** *A priority structure is a tuple $\mathcal{G} = \langle \Phi, \prec \rangle$ such that:*

- $\Phi \subset \mathcal{L}_{\text{PL}}$ *and* Φ *is finite;*
- $\prec$ *is a strict order on* Φ *such that for all formulas* $\phi, \psi \in \Phi$*, it holds that: if* $\phi \prec \psi$*, then* ψ *logically implies* ϕ *in the propositional logic.*

An example of a priority structure is shown in Fig 2.3, where a one-way dashed arrow from ϕ to ψ denotes $\phi \prec \psi$. In Figure 2.3, p is the most ideal proposition and $(p \vee q)$ is the sub-ideal proposition, which intuitively means that if p (the most ideal case) is not the case, then q is the best.

A priority structure supplies a criterion for assessing the relative ideality of states. Given a priority structure, a betterness relation can be derived from a domain of a betterness structure. In this way, priority structures serve a similar purpose to norms in van der Torre and Tan (1998). In this thesis, we follow the approach of van Benthem et al. (2014) to obtain betterness relations from priority structures.

Definition 16 (Betterness Structures Based on Priority Structures (van Benthem et al. (2014)))**.** *Given a priority structure $\mathcal{G} = \langle \Phi, \prec \rangle$ and a betterness structure $M = \langle S, \leqslant, V \rangle$, the relation $\leqslant$ is defined as follows, for any two states $s, s' \in S$:*

$$s \leqslant_{\mathcal{G}} s' \iff \forall \phi \in \Phi : s \in \|\phi\|_M \Rightarrow s' \in \|\phi\|_M$$

then we say M is a betterness structure based on $\mathcal{G}$.

[4]A relation $\prec$ is a strict order on a set S if it is irreflexive, asymmetric and transitive.

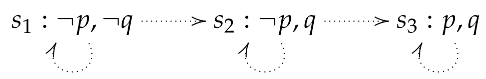


Figure 2.4: $M_{\mathcal{G}}$

Briefly speaking, the more the formulas in $\mathcal{G}$ that a possible world satisfies, the better the world is. And the betterness relations derived in this way are total preorders. An example of a betterness structure based on $\mathcal{G}$ is shown as $M_{\mathcal{G}}$ in Figure 2.4. According to $\mathcal{G}$, the state satisfying p is the best. So s_3 is the best. The state satisfying $(p \vee q)$ is better than those not satisfying it. So s_2 is better than s_1.

2.5 Dynamic epistemic logic

Dynamic epistemic logic generally covers all formal studies on changes of epistemic notions by modal logic that follow the approach of model change. It involves many subfields, such as public announcement logic, belief revision theory, action models for characterizing information and factual change, etc. In this section, we mainly introduce the standard system for dynamic epistemic logic $\mathbb{DEL}$ on which epistemic models can be changed by action models. More details can be referred to Kooi (2007) and van Ditmarsch et al. (2007).

2.5.1 Language $\mathcal{L}_{\text{DEL}}$ and semantics

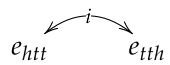
The logic $\mathbb{DEL}$ is different from other epistemic logics due to action models which are called 'models' but appear both in syntax and semantics. The action models are capable of changing information or facts in a gentle way. Let us first introduce the formal definition of action models.

Definition 17 (Action Models (Definition 2.4 in van Ditmarsch and Kooi (2008))). *An action model for a language $\mathcal{L}$ is a structure $U = \langle E, R_1, R_2, \cdots, R_n, pre, post \rangle$ where*

- *E is a finite non-empty set of events;*
- *for each $i \in G$, $R_i : E \times E$ is i's indistinguishability relation between events;*
- *$pre : E \to \mathcal{L}$ assigns to each event a precondition;*
- *$post : E \to (\mathbf{P} \to \mathcal{L})$ assigns to each event a postcondition for each atom. Each $post(e)$ function is required to change truth values of only finitely many propositions. The finite difference is called its domain: $dom(post(e))$.*

For each $e \in E$, (U, e) is called a pointed action model.

An action model consists of events, instead of states. An event can be regarded as an adequate description of all taken-into-account consequences caused by the action. For example, I flipped a coin on the table but I do not see which side was originally heads up (See the figure below, *htt* represents 'head to tail' and *tth* represents 'tail to head').



- The action model for the action of flipping the coin therefore contains two different events: one is that the coin was heads up and I flipped it to tails up and the other one is that the coin was tails up and I flipped it to heads up. These two events constitute the action of flipping.

- Binary relation R_i is the epistemic relation which represents that agent i cannot distinguish these two events. The indistinguishability captures the fact that I cannot see which side is heads up.

- The function *pre* is an abbreviation of *precondition*. For each possible world, it decides which events are executable on it. In our example, only if the coin was heads up, the event e_{htt} can happen on it. Otherwise, the event e_{tth} can happen.

- The function *post* represents *postcondition* for each event and each propositional atom. It is used for modelling factual changes due to the action. In the example, we have $post(e_{htt})(h) = \bot$ and $post(e_{htt})(t) = \top$, which means that if the coin was heads up, then the proposition 'the coin is heads up' (h) becomes false and the proposition 'the coin is tails up' (t) becomes true.

A pointed action model (U, e) represents that the event e is specified as the real event. Action models are not only functions operating on epistemic models, but also work as parts of syntax to show which actions are done.

Definition 18 (Language $\mathcal{L}_{\text{DEL}}$)**.** *The language $\mathcal{L}_{\text{DEL}}$ of dynamic epistemic logic is given by the following BNF:*

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid K_i\phi \mid [(U, e)]\phi,$$

where $p \in \mathbf{P}$, $i \in G$, and (U, e) is a pointed action model.

The formula $K_i\phi$ is the classical epistemic formula representing that agent i knows that ϕ is the case. The formula $[(U, e)]\phi$ means that after performing action (U, e), the formula ϕ becomes true.

The models for $\mathbb{DEL}$ are still epistemic models. Since action models can update epistemic models and the truth condition of $[(U,e)]\phi$ depends on the updated epistemic models, we need to introduce the notion of updated epistemic models.

Definition 19 (Updated epistemic model (van Ditmarsch et al. (2007))). *Given an epistemic model $M_E = \langle S, \sim_1, \cdots, \sim_n, V\rangle$ and an action model $U = \langle E, R_1, \cdots, R_n, pre, post\rangle$, the result of executing U in M_E is the model $M_E \otimes U = \langle S', \sim'_1, \cdots, \sim'_n, V'\rangle$:*

- $S' = \{(s,e) \mid s \in S, e \in E$ *and* $M_E, s \models pre(e)\}$;
- *for each* $i \in G$, $\sim'_i = \{((s,e),(t,f)) \mid (s,e),(t,f) \in S', (s,t) \in \sim_i, (e,f) \in R_i\}$;
- $V'(p) = \{(s,e) \mid M_E, s \models post(e)(p)\}$.

The new domain of states is a subset of the *cartesian product* of the original domain S and the set of events E. The subset only includes those pairs such that the state from S in the pair satisfies the precondition of the event from E in the same pair. And, given two pairs in the new domain, if the agent i can neither distinguish the states from the two pairs nor distinguish the two events, then i cannot distinguish these two pairs in the updated epistemic model. The new valuation is decided by the postcondition function of U.

Definition 19 shows the way to update one epistemic model by a pointed action model. Then the semantics of $\mathcal{L}_{\text{DEL}}$ comes naturally.

Definition 20. *The truth conditions of propositional atoms, Boolean formulas and epistemic formulas are identical to* $\mathbb{EL}$. *Let M be an arbitrary epistemic model.*

$$M, s \models [(U,e)]\phi \text{ iff } M, s \models pre(e) \text{ implies } M \otimes U, (s,e) \models \phi.$$

The antecedent of the truth condition $M, s \models pre(e)$ represents that the event e is executable on the current possible world s. And the consequence shows that after performing the action (U,e), the formula ϕ is true on the updated world (s,e).

2.5.2 Axiom system $\mathbb{DEL}$

The axiom system for dynamic epistemic logic with postconditions is called **UM** by van Ditmarsch and Kooi (2008). In order to make terminology unified and more readable, we name it as $\mathbb{DEL}$ in this thesis.

Definition 21. *The proof system* $\mathbb{DEL}$ *consists of the following axiom schemas and inference rules:*

for each $i \in G$,

(TAUT)	*All instances of tautologies*
(K)	$K_i(\phi \to \psi) \to (K_i\phi \to K_i\psi)$
(T)	$K_i\phi \to \phi$
(4)	$K_i\phi \to K_iK_i\phi$
(5)	$\neg K_i\phi \to K_i\neg K_i\phi$
(U-A)	$[(U,e)]p \leftrightarrow (pre(e) \to post(e)(p))$
(U-N)	$[(U,e)]\neg\phi \leftrightarrow (pre(e) \to \neg[(U,e)]\phi)$
(U-C)	$[(U,e)](\phi \wedge \psi) \leftrightarrow ([(U,e)]\phi \wedge [(U,e)]\psi)$
(U-K)	$[(U,e)]K_i\phi \leftrightarrow (pre(e) \to \bigwedge_{e'R_ie} K_i[(U,e')]\phi)$
(MP)	*From* ϕ *and* $\phi \to \psi$, *infer* ψ
(N)	*From* ϕ, *infer* $K_i\phi$
(RE)	*From* $\phi \leftrightarrow \psi$, *infer* $\chi \leftrightarrow \chi[\phi/\psi]$

Besides the classical $\mathcal{S}5$-axioms for epistemic operator K_i, the axiom (U-A), (U-N), (U-C) and (U-K) are reduction axioms for the dynamic operator $[(U,e)]$. By these reduction axioms and the inference rule (RE), we can reduce every $\mathcal{L}_{\text{DEL}}$-formula to an $\mathcal{L}_{\text{EL}}$-formula. In this way, the completeness for $\mathbb{DEL}$ can be reduced to the completeness for $\mathbb{EL}$. The basic proof strategy refers to Chapter 7.4 in van Ditmarsch et al. (2007) and Theorem 11 in Kooi (2007).

Chapter 3

Knowledge-based Conditional Obligation

不知者不罪。
No blame attaches to the unconscious doer of the wrong.
— 清·钱彩《说岳全传》

3.1 Introduction

The obligations of an agent can be affected by their knowledge. Extensive studies have already been published that formalize various possible interactions between obligations and knowledge as mentioned in Chapter 1.3. Motivated by several real-life scenarios, some of which are taken from literature, this chapter focuses on the 'knowledge-based obligation' and provides a logic for it. These scenarios are listed below.

Scenario 1, 3 and 4 are given by Pacuit et al. (2006) and Scenario 2 by Horty (2019).

Scenario 1. Uma is a doctor whose neighbour Sam is ill. And Sam is a patient at Uma's practice. But Uma does not know that Sam is ill. We intuitively think that Uma has no obligation to treat her neighbour.

Scenario 2. Tao places a coin on the table but Chiyo cannot see whether it is heads up or tails up. Chiyo must risk five euros for the opportunity to bet on heads or tails, with ten euros to win if Chiyo bets correctly (if the coin lands heads up and Chiyo bets on heads, or if the coin lands tails up and Chiyo bets on tails) and cannot get the five euros back if she bets incorrectly; or Chiyo can choose not to gamble, without any profit or loss.

Scenario 3. Uma is a doctor whose neighbour Sam is ill. Sam's daughter Ann comes to Uma and tells her this fact. Now Uma knows that Sam is ill and intuitively, we think Uma has an obligation to treat her neighbour.

Scenario 4. Uma is a doctor whose neighbour Sam is ill. Uma is working in her hospital and Sam is in the same hospital at the same time. So we think that Uma has an obligation to know whether Sam is ill although Sam's daughter does not come to tell her.

Scenario 5. Uma is a doctor whose neighbour Sam is ill. Sam's daughter Ann knows that Uma is a doctor and Ann also knows that Uma ought to treat Sam if she knows that Sam is ill.

Scenario 6. Zaha is a patient in hospital. She knows that she has a fatal disease. Driss, as a doctor in the hospital, is obliged to offer palliative care.

In Scenario 1, one would not say that Uma is obliged to treat Sam as she does not know that he is ill. Likewise, in Scenario 2, Chiyo knows that there is a right choice and she ought to bet on the face that the coin lands, but Chiyo does not know whether the coin lands heads up, which makes Chiyo refrain from bearing the obligation of betting correctly. But in Scenario 3, Uma ought to treat Sam since Ann has already told her. It is clear, based on the first three scenarios, that knowing some fact or not would directly decide whether the agent's obligation is triggered.

Scenario 4 emphasizes a special kind of obligation: epistemic obligation, which means that an agent is obliged to know something. Multiple agents are involved in Scenario 5 where we can also say that if Sam is ill, Ann ought to let Uma know that Sam is ill. In Scenario 6, under the condition that Driss knows that Zaha already knows that she has the disease, Driss ought to offer palliative care.

In all the scenarios, obligations are in the form of conditionals. In other words, whether an agent should fulfill their obligation depends on whether they know the condition (antecedent) of the conditional obligation. Out of this observation, we will formalize knowledge-based obligations in the form of conditional. Moreover, as Scenario 3 suggests, under the condition that Sam is ill, Uma is obliged to treat Sam but Ann is not. So conditional obligations are generally agent-dependent, which inspires us to introduce agents into the formal definition of conditional obligations. In what follows we will briefly review relevant literature.

Conditional obligations It has been illustrated in Chapter 1.2 that the paradoxes derived from standard deontic logic gave rise to the development of the notion of conditional obligations. They are defined to describe obligations that are 'triggered' by some conditions. Dyadic deontic operators represent conditional obligations by $\bigcirc(\phi|\psi)$, which intuitively means that 'it ought to be

ϕ given the condition that ψ' (see Von Wright (1956), Van Fraassen (1973), Horty (1993), Kooi and Tamminga (2008)). Hansson defined his dyadic obligation operator over preference-based models Hansson (1969), where the semantics of $\bigcirc(\phi|\psi)$ is: the best ψ-states are ϕ-states.

Hansson's conditional obligations are generally thought as nonmonotonic obligations in the sense that it *invalidates* the formula $\bigcirc(\phi|\psi) \to \bigcirc(\phi|\psi \wedge \chi)$. It is analogous to nonmonotonic reasoning as well as to variably strict conditionals. In nonmonotonic reasoning, the inference pattern *Left Strengthening*, from $\psi \rightsquigarrow \phi$ infer $\psi \wedge \chi \rightsquigarrow \phi$, is invalid shown by Kraus et al. (1990). In Lewis (1973)'s study on counterfactual, he suggests that counterfactual is a variably strict conditional, which makes a case possible that $\models \psi_1 \mapsto \phi$[1] but $\not\models (\psi_1 \wedge \psi_2) \mapsto \phi$. Nonmonotonicity of Hansson's conditional obligations suggests that 'ought to be ϕ' can be overridden by a stronger condition.

Conditional obligations are related to the issue of factual detachment which can be formalized as $(\bigcirc(\phi|\psi) \wedge \psi) \to \bigcirc(\phi|\top)$. It means that unconditional obligation follows from the truth of the antecedent of conditional obligation. For example, if there are conditional obligations $\bigcirc(\phi|\psi_1)$ and $\bigcirc(\neg\phi|\psi_2)$, then, under different facts, opposite obligations are invoked.

It deserves noting that the upcoming frames for knowledge-based conditional obligations in this chapter are very similar to epistemic plausibility models in context of doxastic logic and belief revision theory (see Baltag and Smets (2006a), van Benthem and Liu (2007), Baltag and Smets (2008) and Chapter 2.3). Given a well-founded preorder called plausibility relation, the agent's belief on ϕ is also interpreted with respect to the preorder.

Obligations involving multiple agents One agent may have some obligations involving other agents. For example, you could have an obligation to let others do something. Moreover, different agents have different obligations. It is possible that different agents ought to act differently, even oppositely sometimes, under some identical condition. For example, when a house is on fire, the firefighters ought to rush to the house but other people should keep away from the fire. These examples indicate that obligations, especially the conditional obligations, are strongly agent-dependent.

Multi-agent deontic logic has been studied in various approaches, such as stit logic (see Horty (1996), Lorini (2012)), coalition epistemic dynamic logic (see De Lima et al. (2010)), non-monotonic reasoning (see Horty (2001), Beirlaen and Straßer (2014)), deontic interpreted systems (See Lomuscio and Sergot (2001, 2003)), normative systems (see Ågotnes et al. (2007)) and game theory (see Tamminga and Duijf (2017)). Collective obligations form an important topic in the realm of multi-agent deontic logic. This thesis will not focus on collective obligations which is left as future work.

[1] $\mapsto$ is the strict implication.

Contributions of this chapter The aim of this chapter is to formalize the notion of knowledge-based obligations as conditional obligations in the multi-agent case. The semantic apparatus, called epistemic betterness structures, is defined by adding epistemic relations for each agents to Hansson's betterness structures. Hansson (1969) defines a dyadic deontic operator $\bigcirc(\_|\_)$ to formalize conditional obligations. We will follow Hansson's approach to define a dyadic deontic operator $\odot_i(\_|\_)$, called epistemic conditional obligation whose definition involves both i's epistemic relation and the betterness relation. We also investigate the new operator by comparing it to objective conditional obligation and finding its Boutilier form. That is, its translation into a standard modal language.

Since the motivation of this chapter is mostly rooted in the scenarios given above, each is modelled by some proper epistemic betterness structure. Over each structure, we can formalize the knowledge-based conditional obligations of different agents with the new dyadic deontic operator accordingly. Moreover, unconditional obligations (all-things-considered obligations) are obtained in the light of an epistemic version of the factual detachment, called epistemic detachment. Epistemic detachment is an inference pattern which intuitively indicates that the agent's uncertainty of the condition decides whether the knowledge-based conditional obligation is triggered. Therefore, over these structures where i bears the same knowledge-based conditional obligation, i might have different unconditional obligations.

The main technical result of this chapter is the axiom system $\mathbb{KCDL}$ for knowledge-based conditional obligations based on Parent (2014)'s system **F**+(CM) (see Chapter 2.1). We prove that the system is strongly complete with respect to an appropriate class of epistemic betterness structures, using the method known as the step-by-step construction (see Burgess (1984), Blackburn et al. (2002)).

Several assumptions on knowledge-based conditional obligations studied in this chapter should be explicitly pointed out:

1. We assume that an agent's knowledge-based conditional obligations not only depend on the deontic betterness relation, but also their epistemic information (and therefore may be different for different agents).

2. We assume that an agent has no uncertainty regarding their own knowledge-based conditional obligations.

3. We assume that an agent can have uncertainty on the state of affairs that may trigger conditional obligations.

Outline of this chapter Some formal background will be shown in Section 3.2. Section 3.3 defines the epistemic betterness structures as the semantics of our logic and investigates some properties of the new dyadic operator. The epistemic detachment can be validated over our frameworks and we can model the above six scenarios properly with epistemic detachment in Section 3.4. A sound and strongly complete logic of knowledge-based conditional obligations $\mathbb{KCDL}$ will be given in Section 3.5. We also compare our work with some related articles in Section 3.6.

3.2 Formal background

We have provided some background on Hansson's dyadic deontic logic in Chapter 2.5. This section will recall some key points and give some other new important notions to be used in the rest of this chapter. The formula $\bigcirc(\phi|\psi)$ represents Hansson's conditional obligation and $\odot_i(\phi|\psi)$ is to denote epistemic conditional obligation (note the dot). Unconditional obligations are interpreted as special cases of conditional ones: $\bigcirc(\phi|\top)$ and $\odot_i(\phi|\top)$, respectively. Given a countable set of propositional letters **P**, the language of Hansson's dyadic deontic $\mathcal{L}_{\mathrm{DDL}}$ is given by the following BNF:

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid \Box\phi \mid \bigcirc(\phi|\phi)$$

where $p \in \mathbf{P}$.

3.2.1 Betterness structures

The language is interpreted over so-called *betterness structures*. Betterness structures are essentially assessment systems to order all the states, which can tell which states are *ideal*.

Definition 22 (Betterness structures (Hansson (1969))). *A betterness structure M is a tuple $\langle S, \leqslant, V\rangle$, consisting of a set of states S, a betterness relation $\leqslant: S \times S$ which is a partial order, and a valuation $V : \boldsymbol{P} \to \mathcal{P}(S)$.*

We write $s < t$ iff $s \leqslant t$ and $t \nleqslant s$.

Definition 23 (Maximal elements). *Given a betterness structure $M = \langle S, \leqslant, V\rangle$ and $T \subseteq S$,*

$$s \in \max_{\leqslant} T \textit{ iff } s \in T \textit{ and } \forall t \in T (s \leqslant t \Rightarrow t \leqslant s)$$

Maximal elements in T are these states which no other state is strictly better than in T. Therefore these states are "as good as it gets" within T.

Now we show the semantics for Hansson's conditional obligations as a reminder.

Definition 24 (Hansson's conditional obligations (Hansson (1969))). *Let $M = \langle S, \leqslant, V\rangle$ be a betterness structure. The semantics of $\mathcal{L}_{\mathrm{DDL}}$ is defined as follows:*

$$\begin{array}{lll} M, s \models \Box\phi & \textit{iff} & S = \|\phi\|_M. \\ M, s \models \bigcirc(\phi|\psi) & \textit{iff} & \max_{\leqslant}\|\psi\|_M \subseteq \|\phi\|_M. \end{array}$$

where $\|\chi\|_M$ *denotes* $\{s \mid s \in S, M, s \models \chi\}$.

We will call Hansson's conditional obligation as an *objective conditional obligation* in Section 3.3 to distinguish it from our epistemic conditional obligation since it does not depend on the agent's information.

Given the definition of the betterness relation, there is no guarantee that the set of maximal elements of some non-empty $\|\phi\|_M$ is non-empty. It can be the case that one can always find a strictly better state and never reach a maximal state. In order to exclude this, besides limitedness and smoothness mentioned in Chapter 2.1, the third alternative property on betterness structures is also proposed here.

Definition 25. *Let* $M = \langle S, \leqslant, V \rangle$ *be a betterness structure. We give three properties of* $\leqslant$ *as follows (two of them have been introduced in Definition 5):*

- *(Limitedness) if* $\|\phi\|_M \neq \emptyset$*, then* $\max_{\leqslant} \|\phi\|_M \neq \emptyset$;
- *(Smoothness) if* $M, s \models \phi$*, then either* $s \in \max_{\leqslant} \|\phi\|_M$ *or* $\exists t : t > s$ *and* $t \in \max_{\leqslant} \|\phi\|_M$.
- *(Noetherianness)* $\leqslant$ *is a reflexive, transitive, converse weakly well-founded order.*

Noetherian betterness structures guarantee that there are maximal elements by ensuring that there are no infinite strictly ascending chains in the model. Smoothness guarantees the existence of maximal elements in a more direct way. All three properties guarantee that the subsets of betterness structures have maximal elements. They are however not equivalent. The following fact captures how these properties are related.

Fact 1. *The relations between limitedness, smoothness and Noetherianness can be shown as following figure:*

Noetherianness ⟶ smoothness ⟶ limitedness

Proof. It is easy to see that these implications hold from left to right. We only give the proofs for the two nontrivial cases that show the converse does not hold: (1) if M is smooth, M can be non-Noetherian. (2) If M is limited, M can be non-smooth.

(1) We define a betterness structure $M_1 = \langle W, \leqslant^*, V \rangle$ such that M_1 is smooth but not Noetherian. Let $W = S \cup T$ where $S = \{s_i \mid i \in \mathbb{N}\}$ and $T = \{t_i \mid i \in \mathbb{N}\}$. Let $\leqslant = \{\langle s_i, t_i \rangle \mid i \in \mathbb{N}\} \cup \{\langle s_i, s_{i+1} \rangle \mid i \in \mathbb{N}\}$. Let $\leqslant^*$ be the reflexive and transitive closure of $\leqslant$. Let $V(p) = W$ for each propositional letter p. See Figure 3.1 for a picture of this model.

Since all the states in W share the identical valuation, we can show by induction on ϕ that $\|\phi\| = W$ or $\|\phi\| = \emptyset$ for all ϕ. Let r be an arbitrary state such that $r \in \|\phi\|_M$. If $r \in S$, there is a state $t \in T$ such that $r \leqslant t$ and $t \in \max_{\leqslant} \|\phi\|_M$. If $r \in T$, $r \in \max_{\leqslant} \|\phi\|_M$. So M_1 is smooth. But it is obvious

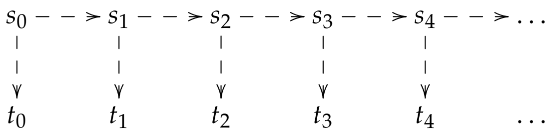


Figure 3.1: The model M_1 which is smooth but not Noetherian.

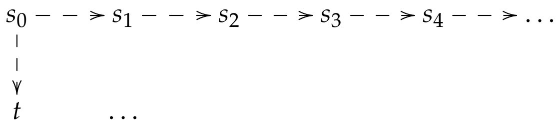


Figure 3.2: The model M_2 which is limited but not smooth.

that M_1 is not Noetherian since all the states in S form an infinite ascending chain.

(2) We define a betterness structure $M_2 = \langle W, \leqslant^*, V\rangle$ such that M_2 is limited but not smooth. Let $W = S \cup \{t\}$ where $S = \{s_i \mid i \in \mathbb{N}\}$. Let $\leqslant = \{\langle s_0, t\rangle\} \cup \{\langle s_i, s_{i+1}\rangle \mid i \in \mathbb{N}\}$. Let $\leqslant^*$ be the reflexive and transitive closure of $\leqslant$. $V(p) = W$ for each propositional letter p. See Figure 3.1 for a picture of this model.

Since all the states in W share the identical valuation, we can show by induction on ϕ that $\|\phi\| = W$ or $\|\phi\| = \emptyset$ for all ϕ. Therefore, for every formula ϕ such that $\|\phi\|_{M_2} \neq \emptyset$, $t \in \max_{\leqslant} \|\phi\|_{M_2}$. But no s_i $(i \geq 1)$ is a maximal ϕ-states. So M_2 is limited but not smooth.

□

3.2.2 Some properties of $\bigcirc(\phi|\psi)$

Nonmonotonicity Extensive literature has discussed the importance of nonmonotonicity in deontic logic (see Nute (2012)). For example, under the condition that you see a man is shot, you ought to help him. But if the man is a fugitive and is shot by a policemen trying to arrest him, you no longer have the obligation to help him. The example indicates that conditional obligations are defeasible in the sense that a conditional obligation that holds may fail when the condition is strengthened, i.e. the truth of $\bigcirc(\phi|\psi)$ does not imply that $\bigcirc(\phi|\psi \wedge \chi)$ holds.

In Hansson's framework, we have $\not\models (\Box(\phi \to \psi) \wedge \bigcirc(\theta|\psi)) \to \bigcirc(\theta|\phi)$ which shows that Hansson's conditional obligations are nonmonotonic. However, Kraus et al. (1990) proposed a weaker version of monotonicity – cautious monotonicity: $\models \bigcirc(\phi|\psi) \wedge \bigcirc(\theta|\psi) \to \bigcirc(\theta|\psi \wedge \phi)$. Cautious monotonicity expresses that adding a new fact that is already a good consequence of a condition should not invalidate another good consequence of the condition. This corresponds to the axiom (CM) in the system **F**+(CM) in Parent (2014).

Boutilier form It is worth noting that in several previous studies on conditional obligations or conditional logic (see Boutilier (1994), van Benthem et al. (2014), van der Torre (1997)), it is shown that over betterness structures, the definition of conditional obligations $\bigcirc(\phi|\psi)$ is equivalent to a formula which merely consists of classical modalities:

$$\mathcal{U}(\psi \to \langle\leqslant\rangle(\psi \wedge [\leqslant](\psi \to \phi))) \qquad (\star)$$

Here, $\mathcal{U}$ is the universal modal operator, while $\langle\leqslant\rangle$ and $[\leqslant]$ are the classical modalities based on the betterness relation $\leqslant$[2]. We call ($\star$) the *Boutilier form* of $\bigcirc(\phi|\psi)$. Intuitively, the Boutilier form expresses that for all the states satisfying ψ, there exists one better state satisfying ψ such that all of its better states either satisfy $\neg\psi$ or satisfy both ψ and ϕ. The Boutilier form not only indicates that $\bigcirc(\phi|\psi)$ can be reduced to classical modal operators over smooth structures, but describes that even on the non-smooth structures, we can always find a better ψ-state for every ψ-state such that each state better than it satisfies $\psi \to \phi$. In Section 3.3, we provide the Boutilier form of our new dyadic obligation operator as well.

3.3 Epistemic betterness structures

In this section we introduce the new operator for epistemic conditional obligations. After presenting the language and semantics in Sections 3.3.1 and 3.3.2, we compare the new operator to Hansson's conditional obligation in Section 3.3.3. We consider some validities and invalidities that help us read the new operator in Section 3.3.4. Finally we give a Boutilier form for the new operator in Section 3.3.5.

3.3.1 The Language for KCDL

Definition 26 (Language $\mathcal{L}_{\text{KCDL}}$)**.** *Let P be a countable set of propositional variables and G be a finite set of agents. The language $\mathcal{L}_{\text{KCDL}}$ is given by the following BNF:*

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid K_i\phi \mid \odot_i(\phi|\phi),$$

where $p \in \boldsymbol{P}$ and $i \in G$.

The formula $K_i\phi$ is read as "agent i knows that ϕ" and $\odot_i(\phi|\psi)$ is read as "if agent i knows that ψ, i ought to see to it that ϕ".

[2] For clarification, $M, s \models [\leqslant]\phi$ iff for all t such that $s \leqslant t$, $M, t \models \phi$ and $\langle\leqslant\rangle$ is the dual of $[\leqslant]$.

3.3.2 Epistemic betterness structures

In order to interpret this language, we need to define structures in which we can interpret both the deontic operator and the epistemic operator. We call these *epistemic betterness structures*. These combine the betterness structures defined in the previous section with models for epistemic logic. For the sake of simplicity we assume that the order of the states is not agent dependent, i.e. we assume all agents agree on all deontic issues. For each agent we add an equivalence relation to the model, which indicates which states are epistemically indistinguishable for that agent.

Definition 27 (Epistemic betterness structures)**.** *Given a set of propositional variables $\boldsymbol{P}$ and a set of agents $G = \{1, 2, \cdots, n\}$, $M = \langle S, \sim_1, \sim_2, \cdots, \sim_n, \leqslant, V\rangle$ is an epistemic betterness structure where:*

- *S is the set of states,*
- *For each $i \in G$, $\sim_i \subseteq S \times S$ is the epistemic equivalence relation for agent i,*
- *$\leqslant \subseteq S \times S$ is a betterness relation (partial order),*
- *$V \subseteq \boldsymbol{P} \rightarrow \mathcal{P}(S)$ is the valuation over S.*

Let $[s]^{\sim_i}$ be the set of states accessible from s by the epistemic relation $\sim_i$.

Epistemic betterness structures are very similar to epistemic plausibility models. When G is a singleton, epistemic betterness structures are almost same as *epistemic-plausibility models* that combine epistemic and doxastic logic (see Baltag and Smets (2006a), van Benthem and Liu (2007), Baltag and Smets (2008), Van Benthem (2010)). For these models it is required that the ordering of states is limited to epistemic equivalence classes. For epistemic betterness structures there is no such constraint.

The semantics of $\mathcal{L}_{\text{KCDL}}$ can now be defined.

Definition 28 (Semantics of $\mathcal{L}_{\text{KCDL}}$)**.** *The truth conditions are defined as follows:*

- $M, s \models p$ *iff* $s \in V(p)$;
- $M, s \models \neg\phi$ *iff* $M, s \not\models \phi$;
- $M, s \models (\phi \wedge \psi)$ *iff* $M, s \models \phi$ *and* $M, s \models \psi$;
- $M, s \models K_i\phi$ *iff* $[s]^{\sim_i} \subseteq \|\phi\|_M$;
- $M, s \models \odot_i(\phi|\psi)$ *iff* $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M) \subseteq \|\phi\|_M$.

Since $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)$ denotes these best ψ-states among the set $[s]^{\sim_i}$, $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)\|\psi\|_M \subseteq \|\phi\|_M$ means that each state in $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)$ satisfies ϕ. In other words, the semantics of $\odot_i(\phi|\psi)$ says that over all states that are indistinguishable to agent i from s, the best ψ-states also satisfy ϕ. In order to ensure that the set of maximal states of a non-empty $[s]^{\sim_i} \cap \|\psi\|_M$ not be empty, we require the epistemic betterness structures to be $\sim$-smooth. As mentioned in Section 3.2, three properties on betterness relations can be considered. In this chapter, we choose smoothness because when we prove soundness and completeness of $\mathbb{KCDL}$, limitedness is too weak since it cannot guarantee that every ϕ-state is worse than some best ϕ-state. Meanwhile, we cannot prove the smooth model established in the completeness proof is Noetherian. Moreover, smoothness is more widely used in the field of conditional logic and nonmonotonic reasoning (see Bossu and Siegel (1985), Kraus et al. (1990), Parent (2014)). Therefore, we assume that epistemic betterness structures are $\sim$-smooth whose definition is similar to smoothness.

Definition 29 ($\sim$-Smoothness)**.** *An epistemic betterness structure M is $\sim$-smooth if for every state s in $M = \langle S, \sim, \leqslant, V\rangle$, for each $i \in G$ and each $t \in [s]^{\sim_i}$, if $M, t \models \phi$, either statement (1) or (2) holds:*

1. $t \in \max_{\leqslant}([s]^{\sim_i} \cap \|\phi\|_M)$;
2. $\exists v \in [s]^{\sim_i} : v > t$ *and* $v \in \max_{\leqslant}([s]^{\sim_i} \cap \|\phi\|_M)$.

Note that $\sim$-smoothness is a generalization of smoothness. It means that for each $[s]^{\sim_i}$, M is smooth when $\leqslant$ is restricted on $[s]^{\sim_i}$. When G is a singleton and the only $\sim$ is total[3], $[s]^{\sim} = S$ and $\max_{\leqslant} \|\phi\|_M = \max_{\leqslant}([s]^{\sim} \cap \|\phi\|_M)$.

3.3.3 Objective vs. epistemic conditional obligations

We use the term 'objective' here to contrast it with 'subjective'. Given our assumptions that all agents agree on all deontic issues, which state is better than which is independent of the agents and in that sense we can view the betterness relation as an objective measure. Objective conditional obligations try to capture what is best under some particular condition independently of subjective aspects of the situation. For example, under the condition that someone is ill, the best state of affairs is that they are treated by a doctor. Hansson's definition of conditional obligations does not depend on the information of the agents involved either, and so is also not subjective in that sense. Recall Hansson's definition:

Definition 30 (Objective conditional obligations)**.** *Given an epistemic betterness structure $M = \langle S, \sim_1, \cdots, \sim_n, \leqslant, V\rangle$,*

$$M, s \models \bigcirc(\phi|\psi) \quad \textit{iff} \quad \max_{\leqslant} \|\psi\|_M \subseteq \|\phi\|_M.$$

[3]A binary relation $\sim$ is total over a set S iff for any $t_1, t_2 \in S$, $t_1 \sim t_2$ or $t_2 \sim t_1$.

It is easily seen that the formal semantics of objective conditional obligation does not involve the epistemic relation $\sim_i$ ($1 \leq i \leq n$), which implies that $\bigcirc(\_|\_)$ is a global operator: it does not depend on the state at which you evaluate it.

Let us consider the example above again. Does the doctor bear an obligation to treat a certain patient? We would hesitate to say so considering the possibility that the doctor knows nothing about the person's health state. The hesitation indicates the importance of the agent's information. In standard epistemic logic, the set of epistemically indistinguishable states for an agent decides what information the agent has. If we restrict our attention to those states that the agent considers possible, the information of the agent will affect their obligations. Therefore, we define epistemic conditional obligations for some agent i depending on the epistemic relation $\sim_i$, so $\odot_i(\_|\_)$ is local. We duplicate the formal definition of epistemic conditional obligations here for explicitness. In the remaining part of this paper, the term 'epistemic conditional obligation' and 'knowledge-based conditional obligation' are used interchangeably.

Definition 31 (Epistemic conditional obligations)**.** *Given an epistemic betterness structure $M = \langle S, \sim_1, \cdots, \sim_n, \leqslant, V\rangle$, for each $i \in G$,*

$$M, s \models \odot_i(\phi|\psi) \quad \textit{iff} \quad \max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M) \subseteq \|\phi\|_M.$$

The following facts capture the logical relations between the above two types of obligations.

Fact 2. *For each $i \in G$, $\not\models \bigcirc(\phi|\psi) \to \odot_i(\phi|\psi)$ and $\not\models \odot_i(\phi|\psi) \to \bigcirc(\phi|\psi)$*

It is easy to give two counter-models for these two cases.

Fact 3. *For every $i \in G$, if $[s]^{\sim_i} = S$, then $M, s \models \bigcirc(\phi|\psi) \leftrightarrow \odot_i(\phi|\psi)$.*

Proof. For every formula ψ, by $[s]^{\sim_i} = S$, $\max_{\leqslant} \|\psi\|_M = \max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)$. Therefore, $M, s \models \bigcirc(\phi|\psi) \leftrightarrow \odot_i(\phi|\psi)$. □

Fact 4. *For every $i \in G$ if $\leqslant \subseteq \sim_i$, then $\models \bigcirc(\phi|\psi) \to \odot_i(\phi|\psi)$, but $\not\models \odot_i(\phi|\psi) \to \bigcirc(\phi|\psi)$.*

Proof. Let (M, s) be an arbitrary pointed epistemic betterness structure such that $M, s \models \bigcirc(\phi|\psi)$. Since $\leqslant \subseteq \sim_i$, $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M) \subseteq \max_{\leqslant} \|\psi\|_M$.

Consider the epistemic betterness structure M' in Figure 3.3. There we have $M', s \models \odot_i(p|q)$, but $M', s \not\models \bigcirc(p|q)$.

□

3.3.4 How to read $\odot_i(\phi|\psi)$

In Section 3.1, we mentioned that the notion of knowledge-based conditional obligations studied in this chapter is based on three assumptions. Regarding assumption 1, the semantics of $\odot_i(\phi|\psi)$ was defined using the agents' information. As for assumption 2 and 3, the following formulas suggest that $\odot_i(\phi|\psi)$ is in line with those as well.

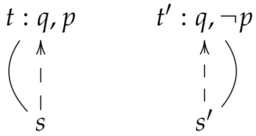
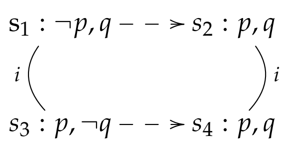


Figure 3.3: The epistemic betterness structure M', which shows that epistemic conditional obligations do not imply objective conditional obligations.

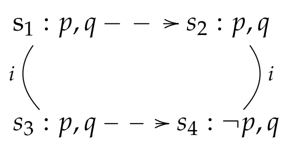
$\models \odot_i(\phi|\psi) \to K_i \odot_i(\phi|\psi)$ and $\models \neg \odot_i(\phi|\psi) \to K_i \neg \odot_i(\phi|\psi)$
The first validity indicates the agent's positive introspection of their epistemic conditional obligations which means that if there is an epistemic conditional obligation for the agent i, i knows it. The second one shows the agent's negative introspection of their non-obligations which means if the agent does not have an epistemic conditional obligation, they know that. Therefore, in this chapter, we only focus on those conditional obligations that are known by the agent.

$\models (\odot_i(\phi|\psi) \wedge K_i\psi) \to \odot_i(\phi|\top)$
This validity indicates that if the agent knows the antecedent of an epistemic conditional obligation, they have an unconditional obligation. The formula is also called 'epistemic detachment' which is discussed more elaborately in Section 3.4. Since $\not\models (\odot_i(\phi|\psi) \wedge \psi) \to \odot_i(\phi|\top)$, the epistemic conditional obligations do not necessarily imply unconditional obligations if the antecedent is merely a fact.

$\bigcirc(\phi|K_i\psi)$ vs. $\odot_i(\phi|\psi)$
The interpretation of $\bigcirc(\phi|K_i\psi)$ is quite similar to the interpretation of $\odot(\phi|\psi)$, which can be read as: under the condition that i knows that ψ, it ought to be that ϕ. But in our logic they are quite different, as can be seen by noting these two invalid formulas: $\not\models \bigcirc(\phi|K_i\psi) \to \odot(\phi|\psi)$ and $\not\models \odot_i(\phi|\psi) \to \bigcirc(\phi|K_i\psi)$. Let us see the following counter-model N_1:



Let s_1 be the factual world. We have $N_1, s_1 \models \bigcirc(p|K_iq) \wedge \neg \odot_i(p|q)$. See another counter-model N_2:



Let s_1 be the factual world. We have $N_2, s_1 \models \odot_i(p|q) \wedge \neg \bigcirc (p|K_i q)$.
These two invalid formulas indicate that the epistemic conditional obligations of agent i cannot be reduced to objective conditional obligations or, vice versa. The truth value of $\odot_i(\phi|\psi)$ is dependent on i's information.

Therefore, epistemic conditional obligation $\odot_i(\phi|\psi)$ can be read as: under the condition that *i knows that* ψ is the case, it ought to be the case *for i* that ϕ.

3.3.5 Boutilier form of $\odot_i(\phi|\psi)$

As mentioned in Section 3.2, we can also find the Boutilier form for $\odot_i(\phi|\psi)$. Although the definition of $\odot_i(\phi|\psi)$ is quite similar to $\bigcirc(\phi|\psi)$, we still need to introduce a new operator defined by classical modalities to give the equivalent Boutilier form of $\odot_i(\phi|\psi)$.

Definition 32 ($[\rightarrow_A \cap \rightarrow_B]\phi$). *Let $\rightarrow_A$ and $\rightarrow_B$ be different binary relations over a domain W. Let $\Box_A$ and $\Box_B$ be the classical modalities whose semantics are defined by $\rightarrow_A$ and $\rightarrow_B$, respectively, as follows:*

$$M, s \models \Box_A\phi \quad \textit{iff} \quad \textit{for all } t \textit{ such that } s \rightarrow_A t,\ M, t \models \phi;$$
$$M, s \models \Box_B\phi \quad \textit{iff} \quad \textit{for all } t \textit{ such that } s \rightarrow_B t,\ M, t \models \phi.$$

We define the semantics of $[\rightarrow_A \cap \rightarrow_B]$ as follows:

$$M, s \models [\rightarrow_A \cap \rightarrow_B]\phi \quad \textit{iff} \quad \textit{for all } t \textit{ such that both } s \rightarrow_A t \textit{ and } s \rightarrow_B t,\ M, t \models \phi.$$

Let $\langle\rightarrow_A \cap \rightarrow_B\rangle$ be the dual of $[\rightarrow_A \cap \rightarrow_B]$.

Proposition 1 (Boutilier form). *For all pointed epistemic betterness structure (M, s), the following two statements are equivalent:*

- $M, s \models \odot_i(\phi|\psi)$,
- $M, s \models K_i(\psi \rightarrow \langle\sim_i \cap \leqslant\rangle(\psi \rightarrow [\sim_i \cap \leqslant](\psi \rightarrow \phi)))$ $(\star_{K_i})$

where K_i is the classical epistemic operator for $\sim_i$.

Proof. We want to show that for every pointed epistemic betterness structure (M, s), $M, s \models \odot_i(\phi|\psi)$ iff $M, s \models K_i(\psi \rightarrow \langle\sim_i \cap \leqslant\rangle(\psi \rightarrow [\sim_i \cap \leqslant](\psi \rightarrow \phi)))$.

($\Rightarrow$) Suppose that $M, s \models \odot_i(\phi|\psi)$. By the semantics of $\odot_i(\phi|\psi)$, it follows that $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M) \subseteq \|\phi\|_M$. Let $t \in [s]^{\sim_i}$ such that $M, t \models \psi$. There must be an $r \in \max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)$ such that $t \leqslant r$ by $\sim$-smoothness (otherwise there would be one infinite ascending chain without any maximal element). As for r, $r \leqslant r$ and $M, r \models \psi \rightarrow \phi$. As for $r' \in [s]^{\sim_i}$ such that $r < r'$, $M, r' \models \neg\psi$ since $r \in \max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)$. Therefore, $M, r \models \psi \rightarrow [\sim_i \cap \leqslant](\psi \rightarrow \phi)$. Since t is an arbitrary state in $[s]^{\sim_i}$, we have $M, s \models K_i(\psi \rightarrow \langle\sim_i \cap \leqslant\rangle(\psi \rightarrow [\sim_i \cap \leqslant](\psi \rightarrow \phi)))$.

($\Leftarrow$) Assume that $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M) \not\subseteq \|\phi\|_M$. It implies that there must be a $t \in \max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)$ such that $M, t \not\models \phi$. So t is the only state such that $t \leqslant t$ and $M, t \models \psi$. Since $M, t \not\models \phi$, $M, t \not\models [\sim_i \cap \leqslant](\psi \to \phi)$. So $M, t \not\models \psi \to [\sim_i \cap \leqslant](\psi \to \phi)$. Therefore, $M, s \not\models K_i(\psi \to \langle \sim_i \cap \leqslant \rangle(\psi \to [\sim_i \cap \leqslant](\psi \to \phi)))$. $\Box$

When we compare this Boutilier form of $\odot_i(\phi|\psi)$ to the Boutilier form ($\star$) of $\bigcirc(\phi|\psi)$ given in Section 3.2.2, it is easy to find that we only replace the operators $\mathcal{U}$ and $\Box$ in ($\star$) with K_i and $[\sim_i \cap \leqslant]$, so as to restrict the set of states to these epistemically indistinguishable states.

3.4 Knowledge as the trigger: epistemic detachment

It seems intuitive to conclude that an agent has an unconditional obligation when the antecedent of a conditional obligation is satisfied. This inference pattern is called *factual detachment*. The earliest version of factual detachment in standard deontic logic is formalized as $(\phi \wedge (\phi \to \bigcirc\psi) \to \bigcirc\psi)$ and there is a variant, called deontic detachment, generally formalized as $(\bigcirc\phi \wedge \bigcirc(\phi \to \psi)) \to \bigcirc\psi$. These give rise to the counter-intuitive consequences of the contrary-to-duty paradox (see page 22 in van der Torre (1997), Greenspan (1975)). For example, we ought to keep a promise ($\bigcirc p$) and it ought to be the case that if we keep the promise, we do not apologize ($\bigcirc(p \to \neg q)$). But when we do not keep the promise, we should apologize ($\neg p \to \bigcirc q$). Now, it is the case that we do not keep the promise ($\neg p$). Then the set of formulas above will lead to two conflicting obligations by standard deontic logic: $\bigcirc\neg q \wedge \bigcirc q$.

As for objective conditional obligation, the validity $\bigcirc(\phi|\psi) \wedge \Box\psi \to \bigcirc(\phi|\top)$ provided by Prakken and Sergot (1997) describes the detachment of the antecedent due to its necessity. In the tradition of Hansson's framework, detachment was also studied in other contexts. By introducing the dynamic epistemic operator of public announcement $[!\psi]$, van Benthem et al. (2014) gave a form like $\bigcirc(\phi|\psi) \to [!\psi] \bigcirc (\phi|\top)$ to capture detachment. Here, the public announcement of the antecedent of the conditional obligation leads to an unconditional obligation.

In our framework, we provide an epistemic version of detachment based on epistemic conditional obligations as the following formula:

$$(\odot_i(\phi|\psi) \wedge K_i\psi) \to \odot_i(\phi|\top)$$

This is in line with our intuitions: if the agent has an obligation that ϕ under the condition that they know that ψ and they do know that ψ is the case, then the agent has an unconditional obligation that ϕ. Epistemic detachment is valid over epistemic betterness structures.

Fact 5. $\models \odot_i(\phi|\psi) \wedge K_i\psi \to \odot_i(\phi|\top)$.

Proof. Let (M, s) be an arbitrary pointed epistemic betterness structure such that $M, s \models \odot_i(\phi|\psi) \wedge K_i\psi$. By Definition 28 and the semantics of $K_i\psi$, $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M) \subseteq \|\phi\|_M$ and $[s]^{\sim_i} \subseteq \|\psi\|_M$. So $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M) = \max_{\leqslant}([s]^{\sim_i} \cap \|\top\|_M)$. Therefore, $\max_{\leqslant}([s]^{\sim_i} \cap \|\top\|_M) \subseteq \|\phi\|_M$ implies that $M, s \models \odot_i(\phi|\top)$. □

Conditional obligations can be activated by different 'triggers' in various approaches, such as facts, necessity of state of affairs, background assumptions, knowledge, etc. In order to model Scenario 1, 2, 3 and 4 from Section 1, we check which triggers transform conditional obligations into unconditional ones (all-things-considered obligations).

3.4.1 Analysis of the six scenarios

The conditional obligations in all scenarios of Section 1 are activated by the agent's knowledge. Using our formalism we can provide novel formalizations of the six scenarios. We use the following atomic propositions and letters to denote agents:

q : Sam is ill.	U : Uma
p : Uma will treat Sam in two minutes.	A : Ann
r : Sam is in the hospital.	
H : The coin is heads up.	C : Chiyo
T : The coin is tails up.	
BH : Chiyo bets heads.	
BT : Chiyo bets tails.	
NG : Chiyo does not gamble.	
c : Zaha is comforted.	D : Driss
d : Zaha gets a fatal disease.	Z : Zaha

We presuppose that 'Sam is not ill' ($\neg q$) is the best state of affairs and 'if Sam is ill, then Uma treats Sam' ($\neg q \vee p$) is better than those cases where 'Sam is ill but Uma does not treat him' ($q \wedge \neg p$).

Scenario 1 & 2: no knowledge, no trigger

Accordingly, we can define an epistemic betterness structure M_1 as Figure 3.4. Assume s_2 to be the factual world.

In Scenario 1, Uma does not know whether Sam is ill and does not know whether she will treat Sam in two minutes. So all four states are indistinguishable. We have $M_1, s_2 \models \odot_U(p|q)$, which means that Uma has an epistemic conditional obligation to treat Sam if she knows that Sam is ill. But $M_1, s_2 \not\models \odot_U(p|\top)$. That is to say that Uma's knowledge-based conditional obligation $\odot_i(p|q)$ is not triggered. The reason is Uma does not know that Sam is ill, i.e. $M_1, s_2 \not\models K_U q$.

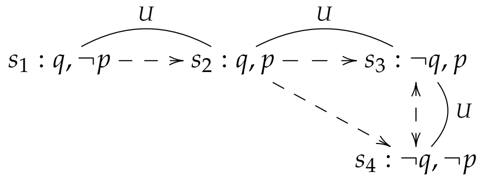


Figure 3.4: M_1

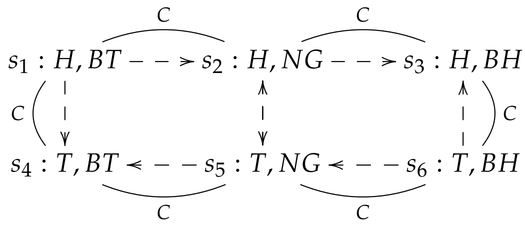


Figure 3.5: M_2

We can model Scenario 2 with M_2 as Figure 3.5. Assume s_1 to be the factual world. In Scenario 2, it is clear that Chiyo knows that under the condition that the coin lands heads up, she ought to bet heads and she also knows that under the condition that the coin lands tails up, she ought to bet tails. So Chiyo has two conditional obligations: $M_2, s_1 \models \odot_C(BT|T) \land \odot_C(BH|H)$. However, we have $M_2, s_1 \models \neg K_C T \land \neg K_C H$. It means that these two obligations are not triggered, which exactly corresponds to Chiyo's not bearing two unconditional obligations in this scenario (also in M_2), i.e. $M_2, s_1 \models \neg \odot_C(BT|\top) \land \neg \odot_C(BH|\top)$.

Scenario 3: knowledge triggers

In Scenario 3, Uma knows that Sam is ill (Figure 3.6). Here we have $M_3, s_2 \models K_U q \land \odot_U(p|q)$, which implies that $M_3, s_2 \models \odot_U(p|\top)$. We can say that Uma's knowledge triggers her knowledge-based conditional obligation, assigning her an unconditional obligation to treat Sam. In this case, the fact q cannot trigger the epistemic conditional obligation $\odot_U(p|q)$ although q is true on the factual world s_2. It corresponds to our intuition that the fact that 'Sam is ill' cannot bring (unconditional) obligations to Uma. Instead, the knowledge that q can.

Although the objective conditional obligation $\bigcirc(p|q)$ is satisfied over M_1 and M_3, and $\bigcirc(BT|T)$ and $\bigcirc(BH|H)$ are satisfied over M_2 as well, they can be triggered neither by the knowledge nor the fact. Therefore, epistemic conditional obligations have a better performance on modelling these scenarios.

Then we will formalize the epistemic obligation in Scenario 4.

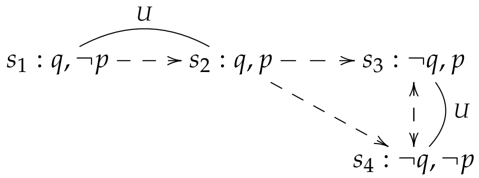


Figure 3.6: M_3

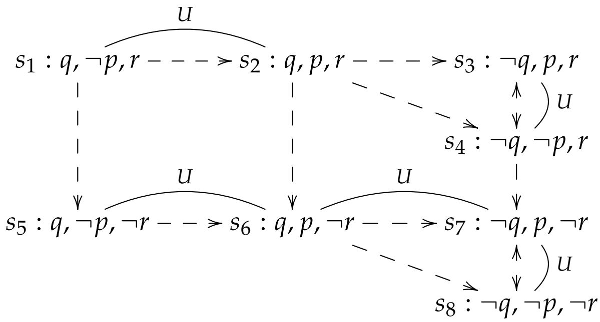


Figure 3.7: M_4

Scenario 4: epistemic obligation

In Scenario 4, 'Uma knows whether Sam is ill' is Uma's obligation under the condition that 'Sam is in the hospital'. We capture Scenario 4 with model M_4 shown as Figure 3.7, where Uma has epistemic obligations: $\odot_U(K_U q|r)$ or $\odot_U(K_U \neg q|r)$. If Sam is factually in the hospital and Uma knows that Sam is in the hospital, the factual world is s_1 or s_2 or s_3 or s_4 where we have $M_4, s_1(s_2) \models \odot_U(K_U q|r)$ and $M_4, s_3(s_4) \models \odot_U(K_U \neg q|r)$, which means that under the condition that Uma knows that Sam is in the hospital, Uma ought to know whether Sam is ill. On s_1 or s_2, $M_4, s_1(s_2) \models K_U r$, which triggers Uma's epistemic obligation that she ought to know that 'Sam is ill'. On s_3 or s_4, likewise, $M_4, s_3(s_4) \models K_U r$, which triggers Uma's epistemic obligation that she ought to know that 'Sam is not ill'.

Next, we formalize the obligations involving multiple agents in Scenario 5 and 6.

Scenario 5 & 6: obligations involving multiple agents

In Scenario 5, we focus on Ann's obligations. It is clear that Ann knows that Uma ought to treat Sam under the condition that Uma knows that Sam is ill. Moreover, Ann has an obligation to inform Uma that Sam is ill when she knows that Sam is ill. So we can model the scenario as M_5 in Figure 3.8.

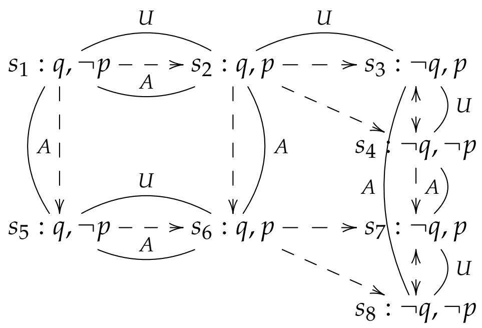


Figure 3.8: M_5

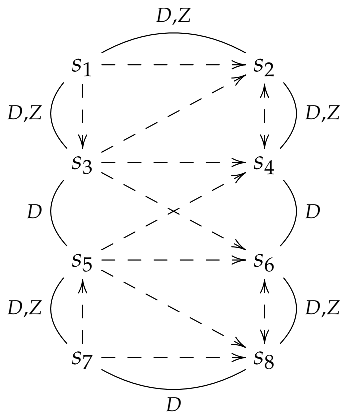


Figure 3.9: M_6

In M_5, assuming s_2 is the factual world, we have $M_5, s_2 \models K_A \odot_U(p|q) \wedge \odot_A(K_U q|q)$.

As for Scenario 6, Driss ought to comfort Zaha when Zaha knows that she has a fatal disease. But Driss ought to not comfort Zaha when Zaha does not know that she has a fatal disease. We can model Scenario 6 with model M_6 as shown in Figure 3.9.

In M_6, we have $M_6, s_3 \models \odot_D(\neg c \mid \neg K_D d \wedge d) \wedge \odot_D(c \mid \neg K_Z d)$.

3.5 A logic of knowledge-based conditional obligation: $\mathbb{KCDL}$

The axiom system of $\mathbb{KCDL}$ is mostly based on Parent's work on Hansson-style dyadic deontic logic (see a series of articles Parent (2008, 2014, 2015)) and other relevant work in conditional logic (see Kraus et al. (1990), Åqvist (1987)). Note that we are working with a multi-agent logic here.

3.5.1 Axiom system for $\mathbb{KCDL}$

	TAUT	(PL)
For each $i \in G$:	$\mathcal{S}5$-schemata for K_i for each $i \in G$	($\mathcal{S}5_n$)
	$\odot_i(B \to C\|A) \to (\odot_i(B\|A) \to \odot_i(C\|A))$	($\odot$K$_i$)
	$\odot_i(B\|A) \to K_i \odot_i(B\|A)$	($\odot_i$Abs)
	$K_i A \to \odot_i(A\|B)$	($\odot_i$Nec)
	$K_i(A \leftrightarrow B) \to (\odot_i(C\|A) \leftrightarrow \odot_i(C\|B))$	($\odot_i$Ext)
	$\odot_i(A\|A)$	($\odot_i$Id)
	$\odot_i(C\|A \wedge B) \to \odot_i(B \to C\|A)$	($\odot_i$Sh)
	$\neg K_i \neg A \to (\odot_i(B\|A) \to \neg \odot_i(\neg B\|A))$	($\odot_i$ D*)
	$(\odot_i(B\|A) \wedge \odot_i(C\|A)) \to \odot_i(C\|A \wedge B)$	($\odot_i$CM)
	If $\vdash A$ and $\vdash A \to B$, then $\vdash B$	(MP)
	If $\vdash A$, then $\vdash K_i A$	(K$_i$N)

$\mathbb{KCDL}$ is almost identical to the system **F**+(CM) in Parent (2014)'s paper which extends Åqvist (1987)'s system **F**. **F**+(CM) was shown to be complete with respect to reflexive and smooth betterness structures (or reflexive, total, transitive and smooth structures). For $\mathbb{KCDL}$ we drop the assumption that the structures are total and adapt the notion of smoothness to $\sim$-smoothness.

The axiom ($\odot$K$_i$) is the epistemic conditional obligation counterpart of the well-known distribution axiom K. Axiom ($\odot_i$Abs) reflects our assumption that we only consider conditional obligations that are known to the agent. Axiom ($\odot_i$Nec) indicates that once you know something is the case, it is the best under arbitrary conditions since its negation is impossible for you. The axiom ($\odot_i$Ext) permits the replacement of the antecedent of the epistemic deontic conditionals with the epistemically equivalent antecedent. Axiom ($\odot_i$Id) means that the best cases where A is true, are cases where A is true. Axiom ($\odot_i$Sh), named after Shoham (1988b), corresponds to a weaker version of the deduction theorem of a nonmonotonic preferential logic mentioned in Shoham (1988a) and the derived rule (S) in Kraus et al. (1990). ($\odot_i$ D*) intuitively means that if the agent considers A to be possible, then under the condition that A, the best states are B-states, it is impossible that under the condition that A, $\neg B$ is the best. Axiom ($\odot_i$CM) corresponds to the (*Cautious Monotonicity*) rule in Kraus et al. (1990).

3.5.2 Soundness of $\mathbb{KCDL}$

Theorem 1 (Soundness)**.** $\mathbb{KCDL}$ *is sound with respect to the class of epistemic betterness structures where* $\leqslant$ *is reflexive, transitive and* $\sim$*-smooth.*

We just show the validity of ($\odot_i$Sh) and ($\odot_i$CM) as examples. Readers are referred to Parent (2014) for more details.

Lemma 1. *The axiom* ($\odot_i$Sh) $\odot_i(C|A \wedge B) \rightarrow \odot(B \rightarrow C|A)$ is valid over epistemic betterness structures where $\leqslant$ is reflexive, transitive and $\sim$-smooth.

Proof. Let (M, s) be an arbitrary pointed epistemic betterness structure. Suppose that $M, s \models \odot_i(C|A \wedge B)$. By the semantics, this means that $\max_{\leqslant}([s]^{\sim_i} \cap \|A \wedge B\|_M) \subseteq \|C\|_M$.

Suppose $[s]^{\sim_i} \cap \|A \wedge B\|_M \neq \varnothing$. Let us assume that $M, s \not\models \odot_i(B \rightarrow C|A)$, which implies that there is a state t such that $t \in \max_{\leqslant}([s]^{\sim_i} \cap \|A\|_M)$, but $M, t \models B \wedge \neg C$. So $t \in \|A \wedge B\|_M$ but $t \notin \max_{\leqslant}([s]^{\sim_i} \cap \|A \wedge B\|_M)$. By $\sim$-smoothness of M, there must be a state $r > t$ such that $r \in \max_{\leqslant}([s]^{\sim_i} \cap \|A \wedge B\|_M)$. But $r \in \|A\|_M$ which is better than t. This contradicts that $t \in \max_{\leqslant}([s]^{\sim_i} \cap \|A\|_M)$. Therefore, $M, s \models \odot_i(B \rightarrow C|A)$.

Suppose $[s]^{\sim_i} \cap \|A \wedge B\|_M = \varnothing$. There are three cases.

1. When $[s]^{\sim_i} \cap \|A\|_M = \varnothing$, it is trivial that $M, s \models \odot_i(B \rightarrow C|A)$.

2. When $[s]^{\sim_i} \cap \|A\|_M \neq \varnothing$ and $\|B\|_M \cap [s]^{\sim_i} = \varnothing$, $M, s \models K_i \neg B$. Therefore, $M, s \models K_i(B \rightarrow C)$, which implies that $M, s \models \odot_i(B \rightarrow C|A)$ by axiom ($\odot_i$Nec).

3. When $[s]^{\sim_i} \cap \|A\|_M \neq \varnothing$ and $[s]^{\sim_i} \cap \|B\|_M \neq \varnothing$, it follows that $M, s \models \odot_i(\neg B|A)$. Then, by propositional logic and ($\odot_i$Nec), we obtain $M, s \models \odot_i(\neg B \rightarrow (B \rightarrow C)|A)$. We conclude that $M, s \models \odot_i(B \rightarrow C|A)$ by ($\odot_i$K).

Therefore, ($\odot_i$Sh) is sound. □

Lemma 2. *The axiom* ($\odot_i$CM) $(\odot_i(B|A) \wedge \odot_i(C|A)) \rightarrow \odot_i(C|A \wedge B)$ is valid over epistemic betterness structures where $\leqslant$ is reflexive, transitive and $\sim$-smooth.

Proof. Let (M, s) be an arbitrary pointed epistemic betterness structure. Suppose that $M, s \models \odot_i(B|A) \wedge \odot_i(C|A)$.

If $[s]^{\sim_i} \cap \|A\|_M \neq \varnothing$, it means that $\max_{\leqslant}([s]^{\sim_i} \cap \|A\|_M) \subseteq \|B\|_M$ and $\max_{\leqslant}([s]^{\sim_i} \cap \|A\|_M) \subseteq \|C\|_M$. Assume that $M, s \not\models \odot_i(C|A \wedge B)$, which implies that there is a state $t \in \max_{\leqslant}([s]^{\sim_i} \cap \|A \wedge B\|_M)$ such that $M, s \models \neg C$. So $t \notin \max_{\leqslant}([s]^{\sim_i} \cap \|A\|_M)$. By $[s]^{\sim_i} \cap \|A\|_M \neq \varnothing$ and $\sim$-smoothness, there must be a state r, such that $r > t$ and $r \in \max_{\leqslant}([s]^{\sim_i} \cap \|A\|_M)$. By $\max_{\leqslant}([s]^{\sim_i} \cap \|A\|_M) \subseteq \|B\|_M$, $r \in \|B\|_M$. So $r \in \|A \wedge B\|_M$ and $r > t$, which contradicts that $t \in \max_{\leqslant}([s]^{\sim_i} \cap \|A \wedge B\|_M)$. Therefore, $M, s \models \odot_i(C|A \wedge B)$.

If $[s]^{\sim_i} \cap \|A\|_M = \varnothing$, then $[s]^{\sim_i} \cap \|A \wedge B\|_M = \varnothing$ as well. Therefore, it is trivial that $M, s \models \odot_i(C|A \wedge B)$. □

3.5.3 Strong completeness of $\mathbb{KCDL}$

Completeness is usually proved by contraposition, which means that, for every consistent formula, we need to find a model that satisfies it. For strong completeness, for each consistent *set* of formulas, we need to find a model that satisfies all of them. The completeness proof of $\mathbb{KCDL}$ in the single-agent case follows Parent's approach in our short paper (see Su (2020) or Appendix of this thesis). However, in this chapter, we adopt the *step-by-step* strategy (more details on this method can be referred to Burgess (1984), Blackburn et al. (2002)) rather than Parent (2014) 's method of canonical models since the *step-by-step* strategy provides a lot of control over the model construction. The strategy does not involve all maximal $\mathbb{KCDL}$-consistent sets (MCS) to construct the canonical model. Rather, it starts with an initial set and builds up a bigger model stepwise with those sets needed to make it perfect.

Given an MCS Δ, let $K_i^{-1}\Delta = \{\phi \mid K_i\phi \in \Delta\}$ and $\Delta^{(i,\phi)} = \{\psi \mid \bigodot_i(\psi|\phi) \in \Delta\}$. Let CON be the set of all maximally $\mathbb{KCDL}$-consistent sets. The following lemmas and definitions are necessary for constructing the final model.

Lemma 3. *(Lindenbaum Lemma) If Γ_0 is a $\mathbb{KCDL}$-consistent set of formulas, then there is an MCS Γ such that $\Gamma_0 \subseteq \Gamma$.*

The Lindenbaum Lemma can be proved in the standard way.

Lemma 4. *For each $i \in G$:*

1. *Let Δ be an MCS. If $K_i\phi \notin \Delta$, then $K_i^{-1}\Delta \cup \{\neg\phi\}$ is consistent;*
2. *Let Δ and Δ_1 be two MCSs. If $K_i^{-1}\Delta \subseteq \Delta_1$ and $\phi \in \Delta_1$, then $\Delta^{(i,\phi)}$ is consistent;*
3. *Let Δ be an MCS. If $\bigodot_i(\phi|\phi \vee \psi) \notin \Delta$, then $\Delta^{(i,\phi\vee\psi)} \cup \{\neg\phi\}$ is consistent;*
4. *Let Δ be an MCS. If $\bigodot_i(\phi|\psi) \notin \Delta$, then $\Delta^{(i,\psi)} \cup \{\neg\phi\}$ is consistent;*
5. *For each $i \in G$, $\vdash_{\mathbb{KCDL}} (\bigodot_i(\theta|\psi \vee \theta) \wedge \bigodot_i(\phi|\psi)) \to \bigodot_i(\psi \to \phi|\theta)$;*
6. *For each $i \in G$, $\vdash_{\mathbb{KCDL}} \bigodot_i(\phi|\psi) \to \bigodot_i(\phi \vee \theta|\psi \vee \theta)$.*

The lemma above is already proved by Parent (2014).

Definition 33 (Network). *$N = \langle S, \sim_1, \cdots, \sim_n, \leqslant, v\rangle$ is a network where*

- *S is a countable set of states,*
- *for each $i \in G$, $\sim_i \subseteq S \times S$,*
- *$\leqslant \subseteq S \times S$,*
- *$v \subseteq S \to CON$.*

Let $[\![\psi]\!]_M$ be the set $\{s \in S \mid \psi \in v(s)\}$. Then we define the notion of $\sim_\in$-smoothness which is the counterpart of $\sim$-smoothness in networks.

Definition 34 ($\sim_\in$-smoothness). *Let $N = \langle S, \sim_1, \cdots, \sim_n, \leqslant, v\rangle$ be a network. N is $\sim_\in$-smooth if for every ϕ and every $s \in S$, for each $i \in G$ and each $t \in [s]^{\sim_i}$, if $\phi \in v(t)$, either statement 1 or 2 holds:*

1. $t \in \max_{\leqslant}([s]^{\sim_i} \cap [\![\phi]\!]_N)$;

2. $\exists u \in [s]^{\sim_i} : u > t$ *and* $u \in \max_{\leqslant}([s]^{\sim_i} \cap [\![\phi]\!]_N)$.

Definition 35 (Coherent network). *A network* $N = \langle S, \sim_1, \cdots, \sim_n, \leqslant, v\rangle$ *is coherent if it satisfies the following conditions:*

1. *there is a label function* $l : S \to \{\Sigma^{(i,\phi)} \mid i \in G,\ \phi \in \mathcal{L}_{\text{KCDL}},\ \exists\ t \in S(v(t) = \Sigma)\}$ *such that* $v(s)$ *extends* $l(s)$.

2. *for all* $s, t \in S$, *if* $s \sim_i t$, *then for each* $K_i\phi \in v(s)$, $\phi \in v(t)$.

3. *for all* $s, t \in S$, *if* $s \sim_i t$ *and* $l(t) = \Sigma^{(i,\phi)}$ *where there exists a* $u \in S$ *and* $v(u) = \Sigma$, *then* $s \sim_i u$.

4. *for each* $i \in G$, $\sim_i$ *is an equivalence relation,*

5. $\leqslant$ *is reflexive, transitive and* $\sim_{\in}$*-smooth.*

Coherent networks are analogous to epistemic betterness structures, but we need further requirements to make a coherent network effectively resemble an epistemic betterness structure.

Definition 36 (Perfect network). *A network* $N = \langle S, \sim_1, \cdots, \sim_n, \leqslant, v\rangle$ *is perfect if it is coherent and does not have any of the defects listed below:*

1. *Defect-1: S contains a state s such that* $K_i\phi \notin v(s)$. *But there is* no $t \in S$ *such that* $s \sim_i t$ *and* $\phi \notin v(t)$. *We can use* $(s, K_i\phi)$ *to denote the defect-1.*

2. *Defect-2: S contains a state s such that* $\bigodot_i(\phi|\psi) \notin v(s)$. *But there is* no $t \in S$ *such that* $t \in \max_{\leqslant}([s]^{\sim_i} \cap [\![\psi]\!]_M)$ *and* $\phi \notin v(t)$. *We can use* $(s, \bigodot_i(\phi|\psi))$ *to denote the defect-2.*

3. *Defect-3: S contains a state s such that* $\bigodot_i(\phi|\psi) \in v(s)$. *But there* exists *a* $t \in S$ *such that* $t \in \max_{\leqslant}([s]^{\sim_i} \cap [\![\psi]\!]_M)$ *and* $\phi \notin v(s)$. *We can use* $(s, \bigodot_i(\phi|\psi), t)$ *to denote the defect-3.*

It is straightforward to induce an epistemic betterness structure given a network by following v for the valuation of atomic propositions.

Definition 37 (Induced epistemic betterness structure). *Let* $N = \langle S, \sim_1, \cdots, \sim_n, \leqslant, v\rangle$ *be a coherent network.* $M_N = \langle S, \sim_1, \cdots, \sim_n, \leqslant, V_N\rangle$ *is the induced epistemic betterness structure regarding N where for each propositional atom p,* $V_N(p) = \{s \in S \mid p \in v(s)\}$.

The following lemma shows that a perfect network induces an epistemic betterness structure that is like a canonical model in the sense that set membership corresponds to truth.

Lemma 5 (Truth lemma). *Let $N = \langle S, \sim_1, \cdots, \sim_n, \leqslant, v\rangle$ be a perfect network. $M_N = \langle S, \sim_1, \cdots, \sim_n, \leqslant, V_N\rangle$ is the induced epistemic betterness structure regarding N. Then for all $\phi \in \mathcal{L}_{\text{KCDL}}$, and all states $s \in S$:*

$$M_N, s \models \phi \quad \textit{iff} \quad \phi \in v(s)$$

Proof. By induction on the structure of ϕ. When ϕ is a Boolean formula, the proof is standard.

When ϕ is in the form of $K_i\psi$:

- ($\Rightarrow$) Suppose that $M_N, s \models K_i\psi$. By the semantics, for each t such that $s \sim_i t$, $M_N, t \models \psi$. By the inductive hypothesis, $\psi \in v(t)$. Assume, to reach a contradiction, that $K_i\psi \notin v(s)$. Since for all t such that $s \sim_i t$, $\psi \in v(t)$, N has defect-1. Contradiction. Therefore, $K_i\psi \in v(s)$.
- ($\Leftarrow$) Suppose that $K_i\psi \in \Delta$. Since N is coherent, by Definition 35.2, for each $t \in S$ such that $s \sim_i t$, $\psi \in v(t)$. By the inductive hypothesis, $M_N, t \models \psi$. By semantics, $M_N, s \models K_i\psi$.

When ϕ is in the form of $\odot_i(\psi|\chi)$:

- ($\Rightarrow$) Suppose that $M_N, s \models \odot_i(\psi|\chi)$. By the semantics, for each $t \in \max_{\leqslant}([s]^{\sim_i} \cap \|\chi\|_{M_N})$, $M_N, t \models \psi$. By the inductive hypothesis, $\|\chi\|_{M_N} = [\![\chi]\!]_{M_N}$ and $\psi \in v(t)$. So $\max_{\leqslant}([s]^{\sim_i} \cap \|\chi\|_{M_N}) = \max_{\leqslant}([s]^{\sim_i} \cap [\![\chi]\!]_{M_N})$. Assume, to reach a contradiction, that $\odot_i(\psi|\chi) \notin v(s)$. It means that N has defect-3. Contradiction.
- ($\Leftarrow$) Suppose that $\odot_i(\psi|\chi) \in v(s)$. Assume, to reach a contradiction, that $M, s \not\models \odot_i(\psi|\chi)$. By the semantics, there is a $t \in S$ such that $t \in \max_{\leqslant}([s]^{\sim_i} \cap \|\chi\|_{M_N})$ and $M, t \models \neg\psi$. By the inductive hypothesis, $\|\chi\|_{M_N} = [\![\chi]\!]_{M_N}$ and $\psi \notin v(t)$. It means that N has defect-4. Contradiction.

□

Therefore, we know that by constructing a perfect network, we can find an epistemic betterness structure which can play the role of a canonical model in a completeness proof.

The sketch of constructing the perfect network is as follows. Let Γ be a $\mathbb{KCDL}$-consistent set which can be extended to a maximally consistent set Γ_0. Then we need to construct a perfect network N_Γ such that there exists one state s in N_Γ such that $v(s) = \Gamma_0$. A stock set of states, $S = \{s_{mn} \mid m, n \in \mathbb{N}\}$, is given beforehand to be used for constructing N_Γ step by step, providing the initial state and all the new states we need for repairing defects. S is obviously countable thereby rendering N_Γ countable as well.

Every defect-1 can be represented by one particular element from the Cartesian product $(S \times \{K_i \mid i \in G\} \times \text{ Form})$. Moreover, every two different defect-1 are represented by two different elements from $(S \times \{K_i \mid i \in G\} \times \text{ Form})$. Every defect-2 can be represented by one particular element from $(S \times \{\odot_i \mid i \in G\} \times \text{ Form} \times \text{Form})$ and every two different defect-2 are represented by two different elements. Every defect-3 can be represented by one particular element from $(S \times S \times \{\odot_i \mid i \in G\} \times \text{ Form} \times \text{Form})$ and every two different defect-3 are represented by two different elements. Therefore, all *potential* defects can be enumerated since the union set $(S \times \{K_i \mid i \in G\} \times \text{ Form}) \cup (\mathrm{S} \times \{\odot_\mathrm{i} \mid \mathrm{i} \in \mathrm{G}\} \times \text{ Form} \times \text{Form}) \cup (\mathrm{S} \times \mathrm{S} \times \{\odot_\mathrm{i} \mid \mathrm{i} \in \mathrm{G}\} \times \text{ Form} \times \text{Form})$ is countable. The enumeration provides an *order* that we use to determine which defect should be repaired at each step of constructing the final perfect network N_Γ.

The construction of N_Γ starts with an initial network N_0 which contains a single state $s_{00} \in S$ such that $v(s_{00}) = \Gamma_0$. Let $D(N_i) = \{d \mid d \text{ is a defect of } N_i\}$, which denotes the set of all defects of N_i. We will show that N_0 is coherent. If N_0 is perfect, just let $N_0 = N_\Gamma$. If N_0 is not perfect, N_0 must have some defect. Let $D_0 = D(N_0)$ where D_0 is the *D-set* of N_0. We repair the defect $d \in D_o$ which is the minimal in the enumeration of all potential defects. If d is a defect-1 or defect-2, we only add one new state s_{10} to N_0. If d is a defect-3, we add the set $\{s_{1n} \mid n \in \mathbb{N}\}$ to repair it. The concrete ways to repair it is to be shown in the following Lemma 6. In this way, we obtain N_1 which is an extension to N_0 without the defect d. Then let $D_1 = D(N_1) \cup (D_0 - \{d\})$ where D_1 if the D-set of N_1. If $D_1 = \varnothing$, let $N_1 = N_\Gamma$. If $D_1 \neq \varnothing$, then we repair the minimal defect d' in D_1. It is worth noting that D_1 contains all defects of N_1, which may include defects of N_0 except d. It is possible that after repairing d, some $d^\star \in D_0$ is also repaired. But we also need to repair $d^\star$ in a specific step for the convenience of proving the final model to be $\sim_\in$-smooth. Which defect should be repaired during each step and the process of repairing defects are shown in the pseudo-code Algorithm 1.

Algorithm 1 How to repair defects during each step

$i := 0$
$D_i := D(N_0)$
while $D_i \neq \varnothing$ **do**
  Repair the minimal defect d in D_i and obtain a network N_{i+1} which lacks d
  $D_{i+1} = D(N_{i+1}) \cup (D_i - d)$
  $i := i + 1$
end while

We continue repairing defects according to Algorithm 1. We either stop at some network N_n where $D_n = \varnothing$ or get a countably infinite sequence of networks $N_0, N_1, \cdots$. For the latter, let the union of all networks in this sequence be N_Γ. We will show that N_Γ is perfect in Theorem 2.

We call each network N_i $(i \geq 0)$, which is constructed as illustrated above, as *brick network*. The following *Repair Lemma* elaborates how to repair different types of defects from the D-set of a coherent brick network at some certain step. It plays a key role in the completeness proof.

Lemma 6 (Repair Lemma)**.** *Let $N_k = \langle S, \sim_1, \cdots, \sim_n, \leqslant, v\rangle$ be an arbitrary brick network and let D_k be the D-set of N_k. Assume $d \in D_k$ is the minimal according to the enumeration of defects. If N_k is coherent, then $N_{k+1} = \langle S', \sim'_1, \cdots, \sim'_n, \leqslant', v'\rangle$ is designed for repairing d such that N_k is a sub-network of N_{k+1} and N_{k+1} is coherent.*

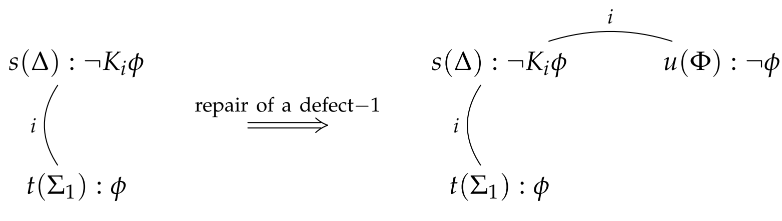
Proof. Let $N_k = \langle S, \sim_1, \cdots, \sim_n, \leqslant, v\rangle$ be an arbitrary coherent brick network. We assume d is the minimal in D_k. We prove the lemma by showing that d can be repaired based on its type.

If d is a defect-1 $(s, K_i\phi)$: Let $v(s) = \Delta$. By Lemma 4.1, $K_i^{-1}\Delta \cup \{\neg\phi\}$ is consistent. By the Lindenbaum Lemma, it can be extended to a maximally consistent set Δ_1. By Lemma 4.2, $K_i^{-1}\Delta \subseteq \Delta_1$ and $\neg\phi \in \Delta_1$ implies that $\Delta^{(i,\neg\phi)}$ is consistent. So it can be extended to a MCS Φ. Now we take a new state u, i.e. $u \notin S$. We define $N_{k+1} = \langle S', \sim'_1, \cdots, \sim'_n, \leqslant', v'\rangle$ as follows:

1. $S' = S \cup \{u\}$,
2. $\sim'_i = \sim_i \cup([s]^{\sim_i} \cup \{u\})^2$,

   For each j such that $j \neq i$, $\sim'_j = \sim_j \cup\{\langle u, u\rangle\}$,
3. $\leqslant' = \leqslant \cup\{\langle u, u\rangle\}$,
4. $v'(s') = \begin{cases} \Phi & \text{if } s' = u \\ v(s') & \text{if } s' \neq u \end{cases}$

Repairing a defect-1 can be illustrated with the following figure where the reflexive and transitive closure is omitted (the same for all the following figures):



In this way, d is no longer in D_{k+1}.

- N_{k+1} **is coherent**:
  1. Define the new label function l' as follows:

     $l'(s') = \begin{cases} \Delta^{(i,\neg\phi)} & \text{if } s' = u \\ l(s') & \text{if } s' \neq u \end{cases}$

2. We know that Φ extends $\Delta^{(i,\neg\phi)}$, $v'(s) = \Delta$, and $v'(u) = \Phi$. For each $K_i\pi \in v'(s)$, by ($\odot_i$Nec), $\odot_i(\pi|\neg\phi) \in v'(s)$. So $\pi \in \Phi$, i.e. $\pi \in v'(u)$. For each $K_i\pi \in v'(u)$, we assume, to reach a contradiction, that $\pi \notin v'(s)$. So $\neg\pi \in v'(s)$. By an $\mathcal{S}5$-schema $\vdash_{\mathbb{KCDL}} \neg\pi \to K_i\neg K_i\pi$, we have $K_i\neg K_i\pi \in v'(s)$. So $\neg K_i\pi \in v'(u)$. Contradiction. Therefore, $\pi \in v'(s)$.

   For other $t \in [s]^{\sim_i}$, the proof is similar.

3. For each $t \in [s]^{\sim_i}$, $t \sim'_i u$ by the definition of $\sim'_i$. We know that $t \sim'_i u$, $l'(u) = \Delta^{(i,\neg\phi)}$, and $v(s) = \Delta$. By $t \sim_i s$, $t \sim'_i s$.

4. It is clear that for each $i \in G$, $\sim'_i$ is still an equivalence relation.

5. It is straightforward to see that $\leqslant'$ is reflexive and transitive. As for $\sim_\in$-smoothness, we can also easily prove it. Let s' be an arbitrary state in S', let π be an arbitrary formula in $v(s')$ and let j be an arbitrary agent in G. If $s' \in S$, we can find a state in $\max_{\leqslant'}([s']^{\sim_j} \cap [\![\phi]\!]_{N_k})$ such that it is strictly better than s' since u is an isolated state with respect to $\leqslant'$ and N_k is $\sim_\in$-smooth. If $s' = u$, $u \in \max_{\leqslant'}([u]^{\sim'_j} \cap [\![\phi]\!]_{N_{k+1}})$ since u is isolated with respect to $\leqslant'$.

If d is a defect-2 $(s, \odot_i(\phi|\psi))$**:** Let $v(s) = \Delta$. We can simply throw a new state u into S such that $u \in \max_{\leqslant'}([s]^{\sim'_i} \cap [\![\psi]\!]_{N_{k+1}})$ and $\neg\phi \in v'(u)$. By Lemma 4.4, $\Delta^{(i,\psi)} \cup \{\neg\phi\}$ is consistent. So it can be extended to a MCS Φ. Now we take a new state u, i.e. $u \notin S$. We define $N_{k+1} = \langle S', \sim'_1, \cdots, \sim'_n, \leqslant', v'\rangle$ as follows:

1. $S' = S \cup \{u\}$,

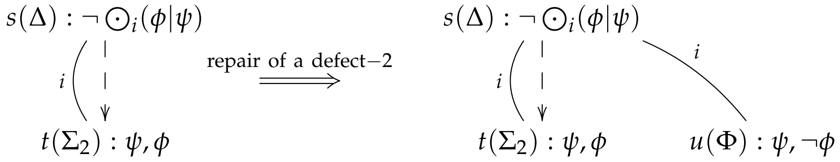
2. $\sim'_i = \sim_i \cup([s]^{\sim_i} \cup \{u\})^2$,

   For each j such that $j \neq i$, $\sim'_j = \sim_j \cup\{\langle u, u\rangle\}$,

3. $\leqslant' = \leqslant \cup\{\langle u, u\rangle\}$,

4. $v'(s') = \begin{cases} \Phi & if\ s' = u \\ v(s') & if\ s' \neq u \end{cases}$

Repairing defect-2 can be illustrated with the following figure:



In this way, d is no longer in D_{k+1}.

- N_{k+1} **is coherent**:

  1. Define the new label function l' as follows:

     $$l'(s') = \begin{cases} \Delta^{(i,\psi)} & \text{if } s' = u \\ l(s') & \text{if } s' \neq u \end{cases}$$

  2. We know that Φ extends $\Delta^{(i,\psi)}$, $v'(s) = \Delta$, and $v'(u) = \Phi$. Then the proof is similar to that in repairing defect-1.
  3. For each $t \in [s]^{\sim_i}$, $t \sim_i' u$ by the definition of $\sim_i'$. We know that $t \sim_i' u$, $l'(u) = \Delta^{(i,\psi)}$, and $v(s) = \Delta$. By $t \sim_i s$, $t \sim_i' s$.
  4. It is clear that for each $i \in G$, $\sim_i'$ is still an equivalence relation.
  5. It is straightforward to see that $\leqslant'$ is reflexive and transitive. $\sim_\in$-smoothness is also preserved since Φ is an isolated state w.r.t. $\leqslant'$.

If d is a defect-3 $(s, \bigodot_i(\phi|\psi), t)$: Let $v(s) = \Delta$ and $v(t) = \Sigma$. In this case, we must add a countable set of new states $X = \{u_1, u_2, \cdots\}$ into S such that every $u_m \in X$ is strictly better than t and $\psi \wedge \phi \in v(u_m)$. There are two cases:

If $l(t) = \Lambda^{(i,\theta)}$ for some $t' \in M$ such that $v(t') = \Lambda$ and for some formula θ (note that it must be the agent i), we can prove that $\Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\}$ is consistent. Assume it is inconsistent. This implies that there are $\theta_1, \theta_2, \cdots, \theta_n \in \Lambda^{(i,\psi\vee\theta)}$ such that $\vdash (\theta_1 \wedge \cdots \wedge \theta_n) \to \theta$. By axiom ($\bigodot_i$K), $\bigodot_i(\theta|\psi \vee \theta) \in \Lambda$ (that is $v(t')$). By Definition 35.3, we have $s \sim_i t'$. Then, by Definition 35.2, we obtain $\bigodot_i(\phi|\psi) \in v(t')$ since $K_i \bigodot_i(\phi|\psi) \in v(s)$. By Lemma 4.5, we have $\bigodot_i(\psi \to \phi|\theta) \in v(t')$. So $(\psi \to \phi) \in v(t)$. According to Algorithm 1, we know that d must be a defect of some N_e where $e \leq m$. So we have $t \in \max_{\leqslant}([s]^{\sim_i} \cap [\![\psi]\!]_{N_e})$. By $(\psi \to \phi) \in v(t)$, we have $\phi \in v(t)$. Contradiction. So $\Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\}$ is consistent. Since $\bigodot_i(\phi|\psi) \in v(t')$, by Lemma 4.5, $\bigodot_i(\phi \vee \theta|\psi \vee \theta) \in v(t')$. In other words, $\bigodot_i(\phi \vee \theta|\psi \vee \theta) \in \Lambda$. Therefore, $(\phi \vee \theta) \in \Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\}$. So we obtain $\phi \in \Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\}$.

We know that Σ contains countably many formulas, we can enumerate them as $\Sigma = \{\chi_m \mid m \in \mathbb{N}\}$. For each $\chi_m \in \Sigma$, if $\Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\} \cup \{\chi_m\}$ is consistent, let Φ_{χ_m} be an MCS extending $\Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\} \cup \{\chi_m\}$; if $\Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\} \cup \{\chi_m\}$ is inconsistent, let Φ_{χ_m} be an MCS extending $\Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\}$.

Let $X = \{u_m \mid m \in \mathbb{N}\}$ where for each $u_m \in X$, $u_m \notin S$. We define a new function v' such that for each $u_m \in X$, $v'(u_m) = \Phi_{\chi_m}$. X is used for repairing d. Note that for each $\chi \in \Sigma$ such that $\Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\} \cup \{\chi\}$ is consistent, X only contains unique state, e.g. u_m, such that $v'(u_m)$ extends $\Lambda^{(i,\psi\vee\theta)} \cup \{\neg\theta\} \cup \{\chi\}$. Since Σ only contains countably many formulas, we have that X is at most countable.

If $s = t$ and $l(t) = \Lambda^{(j,\theta)}$ for some $j \neq i$, some MCS Λ and some formula θ, let $v(s) = \Delta = \{\chi_m \mid m \in \mathbb{N}\}$. For each $\chi_m \in \Delta$, if $\Delta^{(i,\psi)} \cup \{\chi_m\}$ is consistent, let Φ_{χ_m} be an MCS extending $\Delta^{(i,\psi)} \cup \{\chi_m\}$; if $\Delta^{(i,\psi)} \cup \{\chi_m\}$ is inconsistent, let Φ_{χ_m} be an MCS extending $\Delta^{(i,\psi)}$. Let $X = \{u_m \mid m \in \mathbb{N}\}$ where for each $u_m \in X$, $u_m \notin S$. Then we define a new function v' such that for each $u_m \in X$, $v'(u_m) = \Phi_{\chi_m}$. X is used for repairing d.

If $s \neq t$ and $l(t) = \Lambda^{(j,\theta)}$ for some $j \neq i$, the case is impossible.

We can define $N_{k+1} = \langle S', \sim'_1, \cdots, \sim'_n, \leqslant', v' \rangle$ as follows:

1. $S' = S \cup X$.

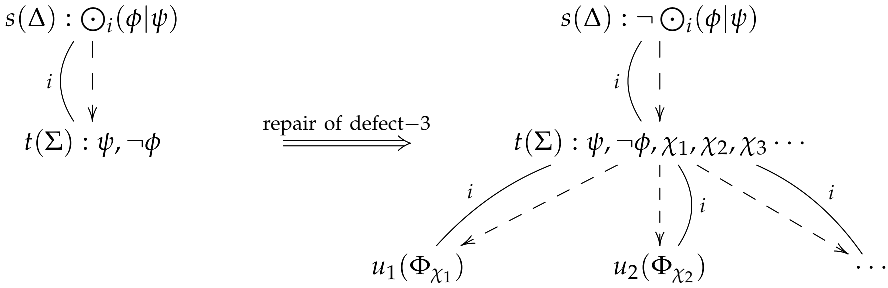
2. $\sim'_i = \sim_i \cup ([s]^{\sim_i} \cup X)^2$.

   For each j such that $j \neq i$, $\sim'_j = \sim_j \cup \{\langle u, u \rangle \mid u \in X\}$.

3. $\leqslant' = \leqslant \cup \{\langle t', u \rangle \mid t' \leqslant t, u \in X\} \cup \{\langle u, u \rangle \mid u \in X\}$.

4. $v'(s') = \begin{cases} \Phi_{\chi_m} & \text{if } s' = u_m \text{ where } u_m \in X \\ v(s') & \text{if } s' \notin X \end{cases}$

Repairing defect-3 can be illustrated with the following figure:



In this way, d is no longer in D_{k+1}.

- N_{k+1} **is coherent**:

  1. Define the new label function l' as follows:

     (1) If $l(t) = \Lambda^{(i,\theta)}$:

     $l'(s') = \begin{cases} \Lambda^{(i,\psi\vee\theta)} & \text{if } s' \in X \\ l(s') & \text{if } s' \notin X \end{cases}$

     (2) If $l(t) \neq \Lambda^{(i,\theta)}$:

     $l'(s') = \begin{cases} \Delta^{(i,\psi)} & \text{if } s' \in X \\ l(s') & \text{if } s' \notin X \end{cases}$

  2. Let u_m be an arbitrary state in X.

- If $l(u_m) = \Lambda^{(i,\psi\vee\theta)}$:

  Let $v(t') = \Lambda$ where $t' \in S$. For each $K_i\pi \in v(t')$, by ($\bigodot_i$Nec), $\bigodot_i(\pi|\psi \vee \theta) \in v(t')$. So $\pi \in v(u_m)$.

  We know that $t \sim_i t'$ by Definition 35.3. Then, by Definition 35.2, for each $K_i\pi \in v(t)$, by an $\mathcal{S}5$-schema $\vdash_{\mathbb{KCDL}} K_i\pi \to K_iK_i\pi$, we have $K_i\pi \in v(t')$. So we obtain that $\pi \in v(u_m)$.

  Similarly, for each $t'' \in S$, if $t'' \sim_i t$, we can prove that for each $K_i\pi \in v(t'')$, $\pi \in v(u_m)$.

  For each $K_i\pi \in v(u_m)$, assume, to reach a contradiction, that $\pi \notin v(t'')$. So $\neg\pi \in v(t'')$. By an $\mathcal{S}5$-schema $\vdash_{\mathbb{KCDL}} \neg\pi \to K_i\neg K_i\pi$, we have $K_i\neg K_i\pi \in v(t'')$. By the above results, $\neg K_i\pi \in v(u_m)$. Contradiction. Therefore, $\pi \in v(t'')$.

  Let u_n be a different state in X. For each $K_i\pi \in v(u_m)$, by an $\mathcal{S}5$-schema $\vdash_{\mathbb{KCDL}} K_i\pi \to K_iK_i\pi$, $K_iK_i\pi \in v(u_m)$. By the above results, $K_i\pi \in v(t)$. By the above results again, $\pi \in v(u_n)$.

- If $l(u_m) \neq \Lambda^{(i,\psi\vee\theta)}$ for the agent i:

  That means $l'(u_m) = \Delta^{(i,\psi)}$. Then we can prove it similarly to the case when we repair defect-1.

3. Let u_m be an arbitrary state in X. We know that $u_m \sim_i' t$.
   - If $l(u_m) = \Lambda^{(i,\psi\vee\theta)}$ for some $t' \in S$ such that $v'(t') = \Lambda$:

     Since N_k is coherent, by Definition 35.3, $t \sim_i t'$. By the definition of $\sim_i'$, $t \sim_i' t'$, thereby $u_m \sim_i' t'$.
   - If $l(u_m) \neq \Lambda^{(i,\psi\vee\theta)}$ for the agent i:

     It means that $s = t$ and $l(\Delta) = \Theta^{(j,\theta)}$ for some $j \neq i$. We do not need to consider this case since u_m only connected with some states in S by $\sim_i$.

4. It is clear that for each $i \in G$, $\sim_i'$ is still an equivalence relation.
5. It is also straightforward to see that $\leqslant'$ is reflexive and transitive. We need to show that $\leqslant'$ is $\sim_\in$-smooth. Let w be an arbitrary state in N_{k+1}. Let π be an arbitrary formula in $v'(w)$. And let j be an arbitrary agent in G. We only need to consider the case where $j = i$ since X is only connected to S by $\sim_i'$.
   - If $w \in [s]^{\sim_i}$, we know that there must exist a state w' in $\max_{\leqslant}([s]^{\sim_i} \cap [\![\pi]\!]_{N_k})$ such that $w' > w$ since N_k is $\sim_\in$-smooth.
     * If $w' = t'$ where $t' \leqslant t$:

       When there is a $u_m \in X$ such that $\pi \in v'(u_m)$, $u_m \in \max_{\leqslant'}([s]^{\sim_i'} \cap [\![\pi]\!]_{N_{k+1}})$ by the definition of $\leqslant'$ and $u_m > w$ by transitivity.

When there does not exist a $u_m \in X$ such that $\pi \in v(u_m)$, $w' \in \max_{\leqslant'}([s]^{\sim'_i} \cap [\![\pi]\!]_{N_{k+1}})$.

* If $w' \nleqslant t$ and $t \nleqslant w'$, then by the definition of $\leqslant'$, w' must be in $\max_{\leqslant'}([s]^{\sim'_i} \cap [\![\pi]\!]_{N_{k+1}})$.

– If $w \in S - [s]^{\sim_i}$, by the definition of $\sim'_i$, w is not connected to X by $\sim_i$.

– If $w \in X$, by the definition of $\leqslant'$, w must be in $\max_{\leqslant'}([s]^{\sim_i} \cap [\![\pi]\!]_{N_{k+1}})$.

Therefore, $\leqslant'$ is $\sim_\in$-smooth.

□

Repair Lemma elaborates which defect should be repaired during each step and how it can be fixed. We are fully prepared for the completeness proof now. In the proof for Theorem 2, we formally describe the process of constructing the perfect coherent network N_Γ and thereafter the completeness is proved.

Theorem 2 (Completeness of $\mathbb{KCDL}$)**.** $\mathbb{KCDL}$ *is strongly complete with respect to the class of* $\sim$*-smooth epistemic betterness structures.*

Proof. Let $S = \{s_{mn} \mid m, n \in \mathbb{N}\}$. $(S \times \{K_i \mid i \in G\} \times \text{ Form}) \cup (\mathrm{S} \times \{\bigodot_\mathrm{i} \mid \mathrm{i} \in \mathrm{G}\} \times \text{ Form} \times \text{Form}) \cup (\mathrm{S} \times \mathrm{S} \times \{\bigodot_\mathrm{i} \mid \mathrm{i} \in \mathrm{G}\} \times \text{ Form} \times \text{Form})$ be the set of all potential defects. We can enumerate all its elements. Given is a $\mathbb{KCDL}$-consistent set of formulas Γ. Then the proof consists of a number of steps:

Constructing N_Γ: Since Γ is consistent, it can be extended to a maximal consistent set Γ_0 by the Lindenbaum Lemma. We define a network $N_0 = \langle S_0, \sim^0_1, \sim^0_2, \cdots, \sim^0_{|G|}, \leqslant_0, v_0\rangle$ where

- $S_0 = \{s_{00}\}$, where $s_{00} \in S$.
- $\sim^0_i = \{\langle s_{00}, s_{00}\rangle\}$ for each $i \in G$.
- $\leqslant_0 = \{\langle s_{00}, s_{00}\rangle\}$.
- $v_0(s_{00}) = \Gamma_0$.

- N_0 **is coherent**:

1. Define the label function l_0 as: $l_0(s_{00}) = \Gamma_0^{(i,\phi)}$ for an arbitrary $i \in G$ and an arbitrary $\phi \in \mathcal{L}_{\mathrm{KCDL}}$.
2. For each $K_i\pi \in v(s_{00})$, by $\mathbb{KCDL} \vdash K_i\pi \to \pi$, $\pi \in v(s_{00})$.
3. Since $l_0(s_{00}) = \Gamma_0^{(i,\phi)}$, by the definition of $\sim^0_i$, $s_{00} \sim^0_i s_{00}$.

4. It is obvious that for each $i \in G$, $\sim_i'$ is an equivalence relation.
5. It is straightforward to see that $\leqslant'$ is reflexive, transitive and $\sim_{\in}$-smooth.

- **Repairing the minimal defect d in D_m:**

  Let $m \geq 0$. Suppose that $N_m = \langle S_m, \sim_1^m, \sim_2^m, \cdots, \sim_{|G|}^m, \leqslant_m, v_m \rangle$ is a coherent network consisting of states from S.

  • If N_m is perfect, namely N_m does not have any defects, let $N_m = N_\Gamma$.

  • If N_m has some defects and $m = 0$, let $D_m = \{d \mid d \text{ is a defect of } N_m\}$. If N_m has some defects and $m > 0$, $D_m = \{d \mid d \text{ is a defect of } N_m\} \cup (D_{m-1} - d')\}$ where d' is the minimal defect in D_{m-1} according to our enumeration.

  Then we repair the minimal defect d in D_m via the approaches shown in Repair Lemma. There are three types of defects:

  – d is a defect-1 or a defect-2:

    According to Repair Lemma, we only need one new state to repair d. Then we get a new networking lacking d: $N_{m+1} = \langle S_{m+1}, \sim_1^{m+1}, \sim_2^{m+1}, \cdots, \sim_{|G|}^{m+1}, \leqslant_{m+1}, v_{m+1} \rangle$, where $S_{m+1} = S \cup \{s_{(m+1)0}\}$ ($s_{(m+1)0} \in S$). In other words, we let the new state u added for repairing d be the state $s_{(m+1)0}$ in S. By Repair Lemma, we know that d will no longer be a defect of any network extending N_m and N_{m+1} is coherent.

  – d is a defect-3:

    According to Repair Lemma, we need a countable set of new states to repair d. Then we get a new network lacking d: $N_{m+1} = \langle S_{m+1}, \sim_1^{m+1}, \sim_2^{m+1}, \cdots, \sim_{|G|}^{m+1}, \leqslant_{m+1}, v_{m+1} \rangle$, where $S_{m+1} = S \cup \{s_{(m+1)n} \mid n \in \mathbb{N}\}$. In other words, we let the new set of states X added for repairing d be the set $\{s_{(m+1)n} \in S \mid n \in \mathbb{N}\} \subset S$. Since $\{s_{(m+1)n} \in S \mid n \in \mathbb{N}\}$ is a countable set, it is enough to repair the defect-4. By Repair Lemma, we know that d will no longer be a defect of any network extending N_m and N_{m+1} is coherent.

- **If there is no $m \in \mathbb{N}$ such that $N_{(m+1)}$ is perfect**

  In this case, we can define $N_\Gamma = \langle S_\Gamma, \sim_1, \sim_2 \cdots, \sim_{|G|}, \leqslant, v \rangle$, where

  – $S_\Gamma = \bigcup_{m \in \mathbb{N}} S_m$,
  – for each $i \in G$, $\sim_i = \bigcup_{m \in \mathbb{N}} \sim_i^m$,
  – $\leqslant = \bigcup_{m \in \mathbb{N}} \leqslant_m$,
  – for an arbitrary $s_{xy} \in S_\Gamma$, $v(s_{xy}) = v_x(s_{xy})$.

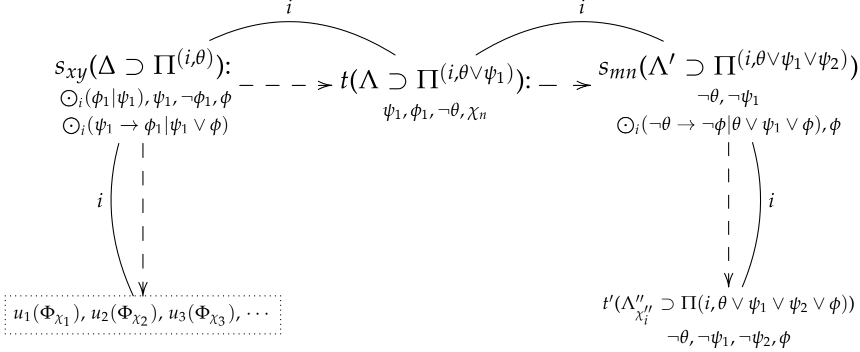
According to our construction of each brick network N_m ($m \in \mathbb{N}$) and Repair Lemma, s_{xy} must be added to repair some defect at the xth step. So s_{xy} must be included in network N_x.

- **N_Γ is coherent**

  • If $N_\Gamma = N_m$ for some $m \in \mathbb{N}$: We know that N_0 is coherent. According to Repair Lemma, every step that repairs some defect will form a new coherent network. So N_m must be coherent.

  • If $N_\Gamma \neq N_m$ for all $m \in \mathbb{N}$: It means that N_Γ is constructed by the countably infinite union defined above. Let $N_\Gamma = \langle S_\Gamma, \sim_1, \sim_2 \cdots, \sim_{|G|}, \leqslant, v\rangle$. We are to show that N_Γ is coherent.

  1. Define the label function l_Γ as follows:
     For each $s_{xy} \in S_\Gamma$, $l_\Gamma(s_{xy}) = l_x(s_{xy})$.
  2. Let $s_{x_1y_1}$ and $s_{x_2y_2}$ be two arbitrary states in S. Suppose that $s_{x_1y_1} \sim_i s_{x_2y_2}$ and $x_1 \leq x_2$. So N_{x_2} contains both $s_{x_1y_1}$ and $s_{x_2y_2}$. We know that N_{x_2} is coherent. Therefore, for each $K_i\pi \in v(s_{x_1y_1})$, $\pi \in v(s_{x_2y_2})$.
  3. Let $s_{x_1y_1}$ and $s_{x_2y_2}$ be two arbitrary states in S. Suppose that $s_{x_1y_1} \sim_i s_{x_2y_2}$, $x_1 \leq x_2$, and $l_\Gamma(s_{x_2y_s}) = \Sigma^{(i,\phi)}$ where there exists a $s \in S_\Gamma$ such that $v(s) = \Sigma$. So N_{x_2} contains $s_{x_1y_1}$, $s_{x_2y_2}$, and s. Since N_{x_2} is coherent, $s_{x_1y_1} \sim_i^{x_2} s$. Therefore, $s_{x_1y_1} \sim_i s$.
  4. It is easy to see that for each $i \in G$, $\sim_i$ is an equivalence relation.
  5. It is straightforward to prove that $\leqslant$ is reflexive and transitive. We need to check if it is also $\sim_\in$-smooth.
     Let s_{xy} be an arbitrary state in N_Γ such that $v(s_{xy}) = \Delta = \{\chi_k \mid k \in \mathbb{N}\}$. So N_x is the first network where s_{xy} is included. Suppose that $\phi \in v(s_{xy})$ and $s_{xy} \notin \max_{\leqslant}([s_{xy}]^{\sim_i} \cap [\![\phi]\!]_{N_\Gamma})$. Some key points in the proof are shown in the following figure:



• When $l(s_{xy}) = \Pi^{(i,\theta)}$ for some $s' \in N_x$ such that $v(s') = \Pi$ and some formula θ:
Since $s_{xy} \notin \max_{\leqslant}([s_{xy}]^{\sim_i} \cap [\![\phi]\!]_{N_\Gamma})$, there is a state t such that $v(t) = \Lambda$ and $s < t$. By Repair Lemma, it must be the case that t is added for repairing a defect-4 d on some state s and s_{xy}. Suppose that d is: there is a formula $\odot_i(\phi_1|\psi_1) \in v(s)$ such that $s_{xy} \in \max_{\leqslant}([s]^{\sim_i} \cap [\![\psi_1]\!]_{N_x})$ and $\neg\phi_1 \in v(s_{xy})$. So $v(t)$, namely Λ, extends $\Pi^{(i,\theta\vee\psi_1)} \cup \{\neg\theta\} \cup \{\chi_n\}$ for some $\chi_n \in \Delta$ or $\Pi^{(i,\theta\vee\psi_1)} \cup \{\neg\theta\}$.
Since $s \sim_i^x s_{xy}$, by N_x is coherent, $\odot_i(\phi_1|\psi_1) \in v(s_{xy})$. By axiom ($\odot_i$Id), $\odot_i(\psi_1 \vee \phi|\psi_1 \vee \phi \vee \psi_1) \in v(s_{xy})$. By $\odot_i(\phi_1|\psi_1) \in v(s_{xy})$ and Lemma 4.5, we obtain that $\odot_i(\psi_1 \to \phi_1|\psi_1 \vee \phi) \in v(s_{xy})$. It means that there must be another defect-4 d' on s_{xy}: $\odot_i(\psi_1 \to \phi_1|\psi_1 \vee \phi) \in v(s_{xy})$, but $s_{xy} \in \max_{\leqslant}([s_{xy}]^{\sim_i} \cap [\![\psi_1 \vee \phi]\!]_{N_x})$ and $\psi_1 \to \phi_1 \notin v(s_{xy})$. To repair d', there must be a countable set of states $X = \{u_k \mid k \in \mathbb{N}\}$ such that for each $u_k \in X$, $v(u_k) = \Phi_{\chi_k}$, and Φ_{χ_k} extends $\Pi^{(i,\theta\vee\psi_1\vee\phi)} \cup \{\neg\theta\} \cup \{\chi_k\}$ or $\Pi^{(i,\theta\vee\psi_1\vee\phi)} \cup \{\neg\theta\}$. And for each $u_k \in X$, $u_k > s_{xy}$.
We know that for any $u' \in [s]^{\sim_i}$, if $u' > u_k$, then $\phi \notin v(u')$. If there is a $u_k \in X$ such that $\phi \in v(u_k)$, then we find the state $u_k \in \max_{\leqslant}([s]^{\sim_i} \cap [\![\phi]\!]_{N_x})$. If there is no $u_k \in X$ such that $\phi \in v(u_k)$, it means that $\Pi^{(i,\theta\vee\psi_1\vee\phi)} \cup \{\neg\theta\} \cup \{\phi\}$ is inconsistent. So it implies that there are $\theta_1, \theta_1, \cdots, \theta_n \in \Pi^{(i,\theta\vee\psi_1\vee\phi)}$ such that $\vdash (\theta_1 \wedge \theta_2 \wedge \cdots \wedge \theta_n) \to (\neg\theta \to \neg\phi)$. So $\odot_i(\neg\theta \to \neg\phi|\psi_1 \vee \phi \vee \theta) \in \Pi$. By N_x is coherent, we know that $s' \sim_i^x s_{xy}$. So we have $\odot_i(\neg\theta \to \neg\phi|\psi_1 \vee \phi \vee \theta) \in \Delta$ as well.

For any $s_{mn} \in [s]^{\sim_i}$, if $s_{mn} > t$, we know that $v(s_{mn})$ extends $\Pi^{(i,\theta\vee\psi_1\vee\psi_2)} \cup \{\neg\psi_1 \wedge \neg\theta\} \cup \{\chi'\}$ for some $\chi' \in v(t)$ and some formula ψ_2, or $v(s_{mn})$ extends $\Pi^{(i,\theta\vee\psi_1\vee\psi_2)} \cup \{\neg\psi_1 \wedge \neg\theta\}$. Since N_m is coherent, $\odot_i(\neg\theta \to \neg\phi|\psi_1 \vee \phi \vee \theta) \in v(s_{mn})$. Let $v(s_{mn}) = \Lambda' = \{\chi''_k \mid k \in \mathbb{N}\}$.

If $\phi \in v(s_{mn})$, s_{mn} has a defect-4 $d^\star$ in N_m: $\odot_i(\neg\theta \to \neg\phi|\psi_1 \vee \phi \vee \theta) \in v(s_{mn})$ and $s_{mn} \in \max_{\leqslant}([s]^{\sim_i^m} \cap [\![\psi_1 \vee \phi \vee \theta]\!]_{N_m})$, but $\neg(\neg\theta \to \neg\phi) \in v(s_{mn})$. To repair $d^\star$, there must exist a set of states $Y = \{w_k \mid k \in \mathbb{N}\}$ such that for each $w_k \in Y$, $v(w_k) = \Lambda''_{\chi''_k}$ where $\Lambda''_{\chi''_k}$ extends $\Pi^{(i,\theta\vee\psi_1\vee\psi_2\vee\phi)} \cup \{\neg\theta \wedge \neg\psi_1 \wedge \neg\psi_2\} \cup \{\chi''_i\}$ or $\Pi^{(i,\theta\vee\psi_1\vee\psi_2\vee\phi)} \cup \{\neg\theta \wedge \neg\psi_1 \wedge \neg\psi_2\}$. Since $\neg\theta \wedge \neg\psi_1 \wedge \neg\psi_2 \in v(w_k)$, $\phi \in v(w_k)$. But we also know that for any $w' \in [s]'$, if $w' > w_k$, then $\neg\phi \in v(w')$. So $w_k \in \max_{\leqslant}([s_{mn}]^{\sim_i} \cap [\![\phi]\!]_{N_\Gamma})$ and $w' > s_{xy}$ by transitivity.

• When $l(s_{xy}) = \Pi^{(i,\theta)}$ for some $j \neq i$: the proof can go through similarly.
Therefore, we proved that N_Γ is $\sim_\in$-smooth.

N_Γ **is perfect:** We have checked that N_Γ is coherent. The only remaining work is to show that N_Γ does not have any defect.

Assume, to reach a contradiction, that N_Γ has a defect d. This defect must have appeared at some step of our construction, let us say at step n. This means that $d \in D_n$. Note that d occurs somewhere in our enumeration of all possible defects, say it appears as its m-th element. Then we know that there are at most m defects preceding it, and so at step $n+m$ we know that defect d has been repaired. Since step $n+m$ is included in N_Γ, it is not possible that N_Γ has defect d. Contradiction. Therefore, N_Γ does not have any defect. Therefore, N_Γ is perfect.

M_{N_Γ} **is the final model:** Since network N_Γ is perfect, it follows by Truth Lemma that the induced epistemic betterness structure regarding N_Γ, namely M_{N_Γ}, is $\sim$-smooth. Therefore, by Truth Lemma, we find an epistemic betterness structure to satisfy the consistent set Γ:

$$\text{for each } \phi \in \Gamma,\ M_{N_\Gamma}, s_{00} \models \phi.$$

Therefore, $\mathbb{KCDL}$ is complete with respect to the class of epistemic betterness structures of which the epistemic relations are equivalence relations and the betterness relation is reflexive, transitive and $\sim$-smooth.

□

For reviewing the whole completeness proof, M_{N_Γ} is induced by a perfect network N_Γ. N_Γ consists of elements from a stock set $S = \{s_{mn} \mid m, n \in \mathbb{N}\}$, in which we can enumerate all potential defects. N_Γ is established from an initial coherent network N_0 consisting of single state $s_{00} \in S$. Then, so as to repair the minimal defect in the current network, one state or countably many states are added step by step. For any particular $n \in \mathbb{N}$, if N_n is perfect, we find a network which is able to induce an epistemic betterness structure M_{N_n}, by Truth Lemma, to satisfy Γ. If we can never find a $n \in \mathbb{N}$ such that N_n is perfect, the union of all N_n ($n \in \mathbb{N}$), namely $\bigcup_{n\in\mathbb{N}} N_n$, is perfect. Finally, it follows by Truth Lemma that the induced epistemic betterness structure (M_{N_n} or M_{N_Γ}) is able to satisfy the consistent set Γ.

3.6 Discussion and conclusion

This paper focuses on the interaction between knowledge and obligations. With a special focus on the notion of knowledge-based obligation, which we formalized using Hansson's preference-based framework. We compare our work with several related articles on deontic logic and then discuss future work inspired by several challenges that remain.

3.6.1 Related work

The logic of objective conditional obligations Parent (2014) systematically studied the logic of objective conditional obligations using Hansson's approach. Since there is no epistemic operator in Parent's system **F**+(CM), it does not distinguish the objective and knowledge-based obligations that play a role in Scenario 1 and 3. In contrast, we are motivated by these scenarios. Just as shown in Section 3.4, epistemic detachment explicates how the agents in the scenarios get an obligation when they know some information.

At the level of technical results, our paper solves a problem that Parent left open. Parent proved that **F**+(CM) is complete with respect to the class of reflexive and smooth betterness structures, under both maximality and optimality rules. Parent (2014)'s main contributions are given in the following table[4]:

maximality rule	totalness	totalness + transitivity	transitivity
reflexivity + smoothness	**F**+(CM)	**G**	?

In terms of $\mathbb{KCDL}$ studied in the current paper, when restricting to the single-agent case and assuming $\sim$ (there is only one $\sim$ in single-agent case) is universal, epistemic betterness structures boil down to betterness structures in Hansson's framework and $\odot_i(\_|\_)$ is logically equivalent to $\bigcirc(\_|\_)$. Therefore, the argument for Theorem 2 can be applied also to prove the open question in Parent (2014) (marked by ? in the above table).

Corollary 1. ***F***+(CM) *is strongly complete with respect to the class of betterness structures that is reflexive, transitive and smooth under the maximality rule.*

Such a completeness result is also discussed by Gabbay et al. (forthcoming).

Reviewing the formal definition of epistemic conditional obligations and objective conditional obligations, we can say the epistemic conditional obligation is a relativized objective conditional obligation with respect to the agents' epistemic information. In other words, the objective conditional obligation of some agent i is the epistemic conditional obligation when i does not have any epistemic information. Therefore, multiple agents can be introduced into $\mathbb{KCDL}$ to formalize the notion of epistemic conditional obligation which is formally agent-dependent.

Knowledge-based Obligation The motivation of our paper is very much inspired by Pacuit et al. (2006)'s paper on *knowledge-based obligations* and Horty (2019)'s work on *epistemic oughts*. We firstly compare our approach to Pacuit et al. (2006).

What makes their approach different is that they define $G(a)$ as 'action a is a morally good action' over history-based models. With the concept of 'good

[4]**G** is obtained by supplementing **F** with $(\neg \bigcirc (\neg B|A) \land \bigcirc(C|A)) \to \bigcirc(C|A \land B)$.

action', 'agent i is obliged to perform action a' is defined as i can perform a and i knows that a is good. As they mentioned in their article, their knowledge-based obligations are *absolute obligations* in the sense that the agent bears these obligations until they practically fulfilled them. Rather, in our paper, we formalized knowledge-based conditional obligations which cannot come into force unless the agent knows that the conditions are the case.

On the other hand, the notion of good actions is analogous to Hansson's conditional obligation in the sense of describing objective obligations, no epistemic information involved. In their frameworks, only if you know that the action to fulfill your obligation is morally good, you bear the knowledge-based obligation. As for formalizing Scenario 1 and 3, given a specific model for Scenarios, Pacuit et al. use two formulas to express whether Uma bears a knowledge-based obligation: (1) $(K_{Uma}\text{Sick} \wedge \langle \text{Ann tells Uma} \rangle \top) \rightarrow K_{Uma}G(\text{Uma treats Sam})$; (2) $\neg K_{Uma}\text{Sick} \rightarrow \neg K_{Uma}G(\text{Uma treats Sam})$. Sentence (2) represents the situation in Scenario 1 that when Uma does not know that Sam is ill, Uma does not have a knowledge-based obligation to treat Sam. Sentence (1) expresses the situation in Scenario 3 that when Uma knows that Sam is ill and it is an available action that Ann tells the fact to Uma, Uma has a knowledge-based obligation to treat Sam. It is very analogous to the principle of epistemic detachment discussed in Section 3.4. In $\mathbb{KCDL}$, only if you bear some knowledge-based conditional obligation and know that the antecedent is the case, you bear an unconditional obligation. For brevity, knowing something triggers the agent's obligation.

However, we investigate the notion of knowledge-based obligation from a perspective of conditionals. In our framework, epistemic detachment is valid over *all* epistemic betterness structures rather than satisfied over some specific model. We think validity of epistemic detachment characterizes the obligations which are essentially *knowledge-based*.

Let us continue by comparing our work to Horty's work on knowledge and obligations. As Scenario 2 in Section 3.1 indicates, Horty (2019) specifically investigates the notion of *epistemic oughts* whose violation invites criticism of the agent only if the agent knows exactly what action is good. Horty embeds an epistemic element into the deontic operator in stit logic. He defines agent i has an epistemic ought to see to it that ϕ with a new stit operator: $\odot[i\ kstit]\phi$, whose definition intuitively means that for each optimal action type over all the currently indistinguishable moments, i sees to it that ϕ. More specifically, $\odot[i\ kstit]\phi$ not only means that 'it ought to be that i sees to it that ϕ', but also i knows which action (type) they should perform over all the currently indistinguishable moments.

We find that our basic idea of defining knowledge-based obligation is very close to Horty's. Horty defines epistemic oughts by focusing on optimal action types over those epistemically indistinguishable moments. Analogously, we define knowledge-based obligations by checking whether these best (optimal in Horty's sense) states uniformly satisfy some formula over those epistemically indistinguishable states. Moreover, according to Horty's frameworks, the following formula is valid in his logic: $\models \odot[i\ kstit]\phi \to K_i \odot[i\ kstit]\phi$. It is intuitively similar to the axiom ($\odot_i$Abs) in $\mathbb{KCDL}$. So it follows that both logics formalize the type of obligations that the agent already knows.

Horty also introduces conditional epistemic oughts as: $\odot([i\ kstit]\phi/\psi)$. He also mentions that his logic can invalidate the standard factual detachment: $\not\models (\odot([i\ kstit]\phi/\psi) \wedge \psi) \to \odot([i\ kstit]\phi/\top)$. But he does not point out when the conditions can be detached. So $\mathbb{KCDL}$ can formalize Scenario 2 more precisely by introducing the conditional obligations and formalizing the gamble in Scenario 2 with epistemic detachment.

In addition, there is also a common difference between our knowledge-based conditional obligations and Pacuit et al. (2006) and Horty (2019): the object of 'knowledge'. The term 'knowledge-based' does not explicitly indicate what knowledge should be involved in knowledge-based obligations. In Pacuit et al. (2006) and Horty (2019)'s research, they specified that knowing what good/optimal actions are is the main characteristic of knowledge-based obligations. Broersen (2008) focused on knowing agents' abilities to achieve goals. In contrast, this chapter does not specify the objects of the knowledge. We mainly focus on the relation between agents' knowledge and obligations. The epistemic detachment theorem indicates this relation: $\models_{\mathbb{KCDL}} (\odot_i(\phi|\psi) \wedge K_i\psi) \to \odot_i(\phi|\top)$. The formula ψ, as the object of the knowledge, can be a proposition describing an action or a state of affairs.

Multi-agent conditional obligations Most literature on conditional obligations involving multiple agents focuses on moral conflicts or collective obligations. The current article does not investigate collective obligations, though, it is still interesting to mention related work.

In the approach of stit logic, obligations of an agent are usually defined with respect to some optimal choices of the agent. Restricting to these ψ-histories of every choice for an agent i, Kooi and Tamminga (2008) define the collective conditional obligation $\odot_G^F(\phi|\psi)$ which means group G ought to see to it that ϕ under the situation that ψ in the interest of group F. Since different groups induce different choices, the conditional obligations of them are consequently distinct. It is analogous to Hansson's conditional obligation in the sense that only the necessity of antecedent can detach their conditional obligation. So $\odot_G^F(\phi|\psi)$ is also objective.

Furthermore, based on the framework of Kooi and Tamminga (2008), Tamminga (2013) extended the language with action propositions which renders conditional obligations defined with actions, instead of state of affairs. In the new framework, formula $\bigcirc_G^F(\alpha_G|\alpha_H)$ expresses that if group H performs action α_H, then, in the F's interest, group G ought to perform α_G. $\bigcirc_G^F(\alpha_G|\alpha_H)$ is designed for expressing what one group ought to do when other group has already performed some particular actions. Tamminga tries to formalize a specific type of conditional obligations: what one ought to do when others perform some

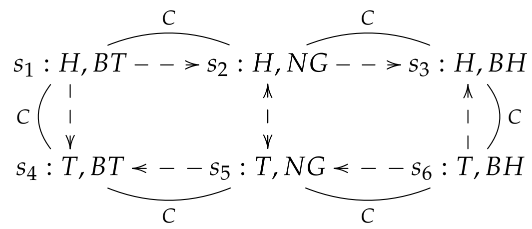


Figure 3.10: M_2

action. It is different from our aim shown in the current paper. We intend to show what it ought to be under some specific state of affairs according to one's epistemic information.

3.6.2 Horty's criticism

Recall Scenario 2. Horty (2019) illustrated that it is hard to say that Chiyo ought to gamble (bet on heads or tails). This is because refraining from gambling does not bring any loss but betting incorrectly makes Chiyo lose 5 euros. However, based on Figure 3.5, we have $M_2, s_1 \models \odot(BH \vee BT|\top)$, which seems to be counter to this intuition.[5]

In context of Horty's illustration, *choosing to gamble* is regarded as an action and the consequences of gambling are not always better than not gambling. Recall the model M_2 for Scenario 2. In our framework, we only describe the static states of all possible consequences. In M_2, the best consequences (regardless of any conditions) are those where Chiyo bets correctly, instead of all consequences of choosing to gamble. However, we are inspired by Horty that $\mathbb{KCDL}$ can be dynamified to describe actions themselves, not only the consequences of actions.

We plan to extend our framework by letting our epistemic betterness structures be based on priority structures, put forward by Liu (2008), van Benthem et al. (2014), and combining the system with dynamic logic. For example, before Chiyo decides to gamble or not, there are only two indistinguishable possible worlds for her, which are the coin heads up and the coin tails up. Formalizing 'betting heads', 'betting tails' and 'no gambling' with three different action models, there will be three corresponding consequence models after executing three action models on the initial model. Comparing with three updated models, we cannot tell which model is the best based on the priority structures defined beforehand. Therefore, we cannot say choosing to gamble is the best action. Chapter 6 will extensively study this approach.

[5]Horty also gives a solution to this problem within the framework of stit logic.

3.6.3 Conclusions

The chief research object of this chapter is the notion of knowledge-based obligation. Motivated by several real-life scenarios, we formalize the notion with the dyadic operator $\odot_i(\_|\_)$ as a conditional. By introducing epistemic relations into betterness structures, the new epistemic betterness structures can distinguish objective conditional obligations and epistemic conditional obligations. Following Hansson's approach to dyadic deontic logic, the semantics of $\odot_i(\_|\_)$ makes it agent-dependent, knowledge-triggered and already-known. With the help of epistemic detachment, we can explain how the obligations of different agents in six real-life scenarios are triggered. In the technical part, we provide a sound and strongly complete axiomatization for the logic with multiple agents $\mathbb{KCDL}$ with respect to the class of $\sim$-smooth epistemic betterness structures where betterness relations are reflexive, transitive.

Chapter 4

How Knowledge Triggers Obligation: A Dynamic Logic of Epistemic Conditional Obligation

4.1 Introduction

In Chapter 3, we investigated epistemic conditional obligations which are a type of conditional obligations where the consequent is triggered by the knowledge of the antecedent. A static logic of epistemic conditional obligations ($\mathbb{KCDL}$) has been presented in the previous chapter, which defines operators $\bigodot_i(\phi|\psi)$ as: the best ψ-states that are *epistemically indistinguishable for agent i* also satisfy ϕ.

In this chapter, priority structures are introduced as representations of norms that remain *static* throughout. These norms somehow determine obligations. Accordingly, we build on the logic of epistemic conditional obligations to study the *dynamic* process whereby the acquisition of new information, or the change of factual circumstances, triggers changes of obligations. To do so we introduce a dynamic operator formalizing obligation change, and show how the new logic can systematize some fundamental deontic notions. The proposed logic is motivated by the following scenarios, among which Scenario 7 is taken from Pacuit et al. (2006) and Scenario 8 is a variant of an example from Horty (2019).

Scenario 7. Uma is a doctor whose neighbour Sam is ill. And Sam is a patient at Uma's practice. But Uma does not know that Sam is ill. We intuitively think that Uma has no obligation to treat her neighbour. Then Sam's daughter Ann shouts loudly on the street that "My dad is ill, any help please?" Now Uma knows that Sam is ill and has an obligation to treat Sam.

Scenario 8. One coin is tossed and covered by a cup. Fumio and Chiyo have an obligation to bet correctly (if the coin lands heads up and they bet on heads, or if the coin lands tails up and they bet on tails). Chiyo then lifts the cup, looks at the coin and ensures that the coin is heads up by some sleight of hand. Fumio observes Chiyo looking at the coin, and he considers it possible that Chiyo has flipped the coin. So before Chiyo looks at the coin, they do not have obligations to bet on heads (or on tails). After Chiyo flips the coin, Chiyo has an obligation to bet on heads but Fumio still does not have an obligation to bet on heads.

Scenario 9. Driss promised to his friend that he will go to the party on time. But when he is on the way to the party, he sees a car accident happening. Now, Driss ought to call an ambulance and help the people in the car, although it could make himself be late for the party. Driss has a new obligation to call an ambulance which overrides the obligation to keep the promise.

Outline of this chapter We will review some necessary technical background in Section 4.2. Section 4.3 gives the language and semantics for the dynamic epistemic conditional obligation, which is then used for modelling the scenarios mentioned above. Section 4.4 uses our logic to provide novel formalizations of important deontic notions such as absolute, *prima facie* and all-things-considered obligations, as well as of a new type of obligation, which we call safe obligations. In developing these notions we highlight how the existing body of theory on conditional belief dynamics in Dynamic Epistemic Logic (see Baltag and Smets (2006b)) bears significance for the understanding of deontic conditionals and their dynamics. Finally, Section 4.5 provides a sound and complete axiom system for logic $\mathbb{DKCDL}$, based on standard reduction axioms and the Kangerian-Andersonian reduction of deontic operators (see Anderson (1958), Kanger (1970)). This chapter is written based on our paper published in the Eighth International Conference on Logic, Rationality and Interaction in 2021 Grossi et al. (2021). The proof details of strong completeness are entirely shown in this chapter. Some relative discussions and future works are complemented in the last section.

4.2 Preliminaries

Let **P** be a countable set of propositional atoms and let $G = \{1, \cdots, n\}$ be a finite set of agents. The semantics of the static logic of knowledge-based conditional obligations is shown as a reminder.

Definition 38 (Semantics of $\mathcal{L}_{\mathrm{KCDL}}$)**.** *The truth conditions of formulas can be defined over M as follows (only the non-trivial cases are shown):*

- $M, s \models K_i\phi$ *iff* $[s]^{\sim_i} \subseteq \|\phi\|_M$*;*
- $M, s \models \odot_i(\phi|\psi)$ *iff* $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M) \subseteq \|\phi\|_M$

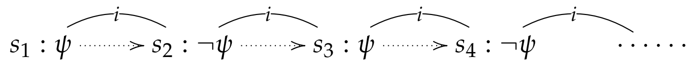
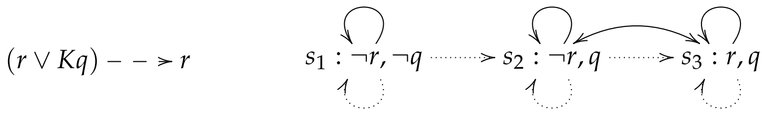


Figure 4.1: M ($\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)$ is empty)



(a) $\mathcal{G}_k$ (b) $M_{\mathcal{G}k}$

Figure 4.2: Two Examples

Observe that for non-empty $[s]^{\sim_i} \cap \|\psi\|_M$, the set $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)$ can be empty, which makes $\odot_i(\phi|\psi)$ trivially true. Figure 4.1 exemplifies the case:

In Figure 4.1, a directed dotted arrow from s_n to s_m denotes $s_n \leqslant s_m$. The solid line between s_n and s_m represents that $s_n \sim_i s_m$. The reflexive and transitive closure of $\sim_i$ and $\leqslant$ are omitted. The epistemic betterness structure M consists of an infinite strictly ascending chain where ψ is satisfied alternatively on the sequence of states. Therefore, $\max_{\leqslant}([s]^{\sim_i} \cap \|\psi\|_M)$ is empty and for an arbitrary formula ϕ, the formula $\odot_i(\phi|\psi)$ is trivially true.

The class of all epistemic models is denoted by $S5$. An epistemic betterness structure is also an epistemic model extended with a betterness relation $\leqslant$ over the set of states. In the study on $\mathbb{KCDL}$, betterness relations between states are given *a priori*. In this chapter, *priority structures* will be introduced for ordering states. In Chapter 2.4, we have given the original definition of priority structures whose domains are finite sets of $\mathcal{L}_{\text{PL}}$-formulas. In this chapter, we define it within $\mathcal{L}_{\text{EL}}$-formulas. Priority structures enable us to define the betterness relations between states according to the $\mathcal{L}_{\text{EL}}$-formulas that are satisfied on the states.

Definition 39 ($\mathcal{L}_{\text{EL}}$-Priority Structures)**.** *Given the language of the classical epistemic logic $\mathcal{L}_{\text{EL}}$, an $\mathcal{L}_{\text{EL}}$-priority structure is a tuple $\mathcal{G} = \langle \Phi, \prec \rangle$ such that:*

- $\Phi \subset \mathcal{L}_{\text{EL}}$ *and* Φ *is finite;*
- $\prec$ *is a strict order on* Φ *such that for all formulas* $\phi, \psi \in \Phi$*, it holds that: if* $\phi \prec \psi$*, then* ψ *logically implies* ϕ*.*

An example of an $\mathcal{L}_{\text{EL}}$-priority structure is shown in Fig.4.2a, where a one-way dashed arrow from ϕ to ψ denotes $\phi \prec \psi$.

A priority structure supplies a criterion for assessing the relative ideality of states. Given an $\mathcal{L}_{\text{EL}}$-priority structure, a betterness relation can be derived from a domain of an epistemic model. In this way, priority structures serve a similar purpose to norms in van der Torre and Tan (1998). In this chapter, we follow the approach of van Benthem et al. (2014) to obtain betterness relations from priority structures.

Definition 40 (Epistemic Betterness Structures Based On Priority Structures)**.** *Given an $\mathcal{L}_{\text{EL}}$-priority structure $\mathcal{G} = \langle \Phi, \prec \rangle$ and an epistemic model $M_E = \langle S, \sim_1, \cdots, \sim_n, V \rangle$, the structure $M = \langle S, \sim_1, \cdots, \sim_n, \leqslant_{\mathcal{G}}, V \rangle$ is an epistemic betterness structure based on $\mathcal{G}$ if M is M_E extended with the betterness relation $\leqslant_{\mathcal{G}}$, where $\leqslant_{\mathcal{G}}$ is defined as follows, for any two states $s, s' \in S$:*

$$s \leqslant_{\mathcal{G}} s' \iff \forall \phi \in \Phi : s \in \|\phi\|_{M_E} \Rightarrow s' \in \|\phi\|_{M_E}$$

In other words, when an epistemic model M_E and an $\mathcal{L}_{\text{EL}}$-priority structure $\mathcal{G}$ are provided, we can construct an epistemic betterness structure by adding a betterness relation based on $\mathcal{G}$ to M_E. An example of an epistemic betterness structure based on $\mathcal{G}_k$ is shown as $M_{\mathcal{G}_k}$ in Figure 4.2b. According to $\mathcal{G}_k$, the state satisfying r is the best. So s_3 is the best. The state satisfying $r \vee Kq$ is better than those not satisfying it. So s_2 is better than s_1. Since an *epistemic betterness structure based on a priority structure* is an epistemic betterness structure, we will also call them just *epistemic betterness structures* in the following parts.

It is worth noting that the betterness relation based on some priority structure no longer has the issue mentioned in Figure 4.1. It follows from Definition 40 and finiteness of priority structures. Proposition 2 shows the point.

Proposition 2. *Given a priority structure $\mathcal{G} = (\Phi, \prec)$ and an epistemic betterness structure $M_{\mathcal{G}} = (W, \sim_1, \cdots, \sim_n, \leqslant_{\mathcal{G}}, V)$ based on $\mathcal{G}$, if $T \subseteq W$ is not empty, then $\max_{\leqslant_{\mathcal{G}}} T \neq \varnothing$.*

The proof goes by contradiction. The basic strategy is to show that those states in T which satisfies the most formulas in the priority structure must be maximal states in T. The existence of the state that satisfies *the most* formulas comes from the finiteness of the priority structure.

Proof. For each $s \in T$, let $\Phi_s = \{\phi \in \Phi \mid M_{\mathcal{G}}, s \models \phi\}$. Since $|\Phi|$ is finite, there exists an $s \in T$ such that for each $t \in T$ and $t \neq s$, $|\Phi_t| \leq |\Phi_s|$. Assume that $s \notin \max_{\leqslant_{M_{\mathcal{G}}}} T$. It means that there is a $u \in T$ such that $s <_{\mathcal{G}} u$. By the definition of $<_{\mathcal{G}}$, for each $\phi \in \Phi$, we have $\phi \in \Phi_s \Rightarrow \phi \in \Phi_u$ and there is $\phi' \in \Phi$ such that $\phi' \in \Phi_u$ but $\phi' \notin \Phi_s$. This implies that $|\Phi_s| < |\Phi_u|$. Contradiction. Therefore, we have $s \in \max_{\leqslant_{M_{\mathcal{G}}}} T$. So we proved $\max_{\leqslant_{M_{\mathcal{G}}}} T \neq \varnothing$. □

In the rest of this thesis, all epistemic betterness structures are based on a given priority structure. It follows that, on a pointed epistemic betterness structure (M, s), the set $\max_{\leqslant} \|\psi\|_M \cap [s]^{\sim_i}$ is not empty if $\|\psi\|_M \cap [s]^{\sim_i}$ is not empty.

4.3 Dynamic epistemic conditional obligation

In this section, we intend to establish a dynamic extension to $\mathbb{KCDL}$. *Action models*, originally introduced in dynamic epistemic logic (DEL), can characterize not only the information changes, but also the factual changes (truth value of the propositions). The formal definitions of action models and updated epistemic models can be referred to Chapter 2.5. Then we give the language and semantics for the dynamic extension to the static logic of knowledge-based conditional obligations.

4.3.1 Language and semantics of $\mathcal{L}_{\text{DKCDL}}$

Definition 41. *The language $\mathcal{L}_{\text{DKCDL}}$ is given by the following BNF:*

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid K_i\phi \mid \odot_i(\phi|\phi) \mid [(U,e)]\phi,$$

where $p \in \mathbf{P}$, $i \in G$, and (U,e) is a pointed action model.

The language of dynamic epistemic logic is $\mathcal{L}_{\text{DEL}}$, which is identical to $\mathcal{L}_{\text{DKCDL}}$ but without the dyadic operator $\odot_i(\_|\_)$. Subsequently, we provide the semantics of the formula $[(U,e)]\phi$ over epistemic betterness structures based on priority structures. Firstly, we need to define epistemic betterness structures updated by action models.

Definition 42 (Updated epistemic betterness structures)**.** *Given an epistemic model $M_E = \langle S, \sim_1, \cdots, \sim_n, V\rangle$ and a priority structure $\mathcal{G} = \langle \Phi, \prec\rangle$, the structure $M = \langle S, \sim_1, \cdots, \sim_n, \leqslant_{\mathcal{G}}, V\rangle$ is the epistemic betterness structure based on $\mathcal{G}$. Letting $U = \langle E, R_1, \cdots, R_n, pre, post\rangle$ be an action model, the result of executing U in M is the model $M \otimes U = \langle S', \sim'_1, \cdots, \sim'_n, \leqslant', V'\rangle$ where:*

- $\langle S', \sim'_1, \cdots, \sim'_n, V'\rangle = M_E \otimes U$;
- $\leqslant' = \{((s,e),(t,f)) \in S' \times S' \mid \forall\phi \in \Phi : (s,e) \in \|\phi\|_{M_E \otimes U} \Rightarrow (t,f) \in \|\phi\|_{M_E \otimes U}\}$.

An updated epistemic betterness structure consists of its corresponding updated epistemic model and an updated betterness relation. The new betterness relation *re-orders* these new states based on the priority structure. Now we can give the truth condition of the formula $[(U,e)]\phi$.

Definition 43. *The truth conditions of atoms, Boolean formulas, epistemic formulas and dyadic conditional obligations are identical to $\mathbb{KCDL}$. Let M be an arbitrary epistemic betterness structure based on priority structure $\mathcal{G}$.*

$$M, s \models [(U,e)]\phi \quad \textit{iff} \quad M, s \models pre(e) \textit{ implies } M \otimes U, (s,e) \models \phi.$$

The symbol $\models$ is also to be used as logical consequence. It means that, for an arbitrary set of formulas Φ and an arbitrary formula ϕ, $\Phi \models \phi$ if and only if for all pointed epistemic betterness structures (M,s) such that $M, s \models \Phi$, it holds that $M, s \models \phi$.

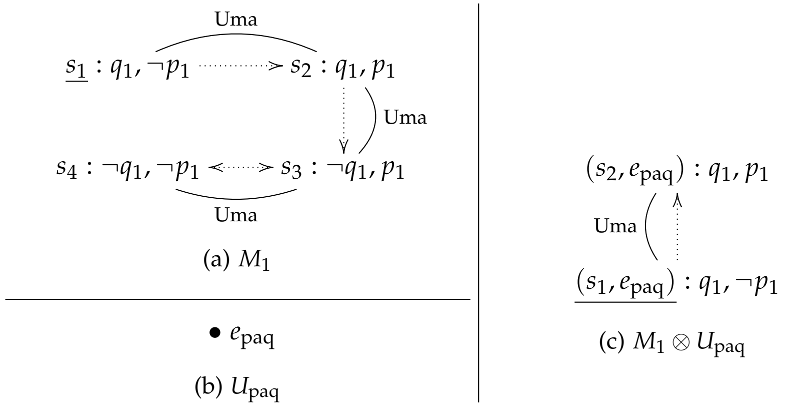


Figure 4.3: Scenario 1: Ann is shouting loudly

4.3.2 Analysis of scenarios 7, 8 and 9

We are now in a position to formalize how obligations change in response to information and factual changes. For each scenario, a priority structure is given in advance, which remains unchanged throughout the story. It specifies the betterness relations in both the initial and updated epistemic betterness structures. The information changes and factual changes are characterized by action models. After performing an action, the updated epistemic betterness structure will determine the agents' new obligations. In the following models, the transitive and reflexive closures of all types of relations are omitted in the figures.

Scenario 7: new information triggers obligations

In Figure 4.3, q_1 refers to 'Sam is ill' and p_1 refers to 'Sam is treated'. We first give the priority structure $\mathcal{G}_1$ for scenario 7. 'Sam is not ill' ($\neg q_1$) is always the best state of affairs and 'if Sam is ill, then Sam is treated' ($\neg q_1 \vee p_1$) is better than those cases where 'Sam is ill but Sam is not treated' ($q_1 \wedge \neg p_1$).

Accordingly, the initial epistemic betterness structure based on $\mathcal{G}_1$ is M_1 (see Figure 4.3a). Over M_1, we have $M_1, s_1 \models \neg \odot_{\text{Uma}}(p_1|\top) \wedge \neg K_{\text{Uma}} q_1$, which means that Uma does not know that Sam is ill and does not have an obligation to see to it that Sam is treated. Then, Sam's daughter shouts loudly outside that her dad is ill. This action can be modeled by an action model of truthful public announcements, i.e., $(U_{\text{paq}}, e_{\text{paq}})$ (see Figure 4.3b, 'paq' refers to 'public announcement that q_1'). An action model of truthful public announcement that ϕ is a singleton action model where the precondition equals to ϕ and the postconditions for all propositions are *id*. It consequently eliminates all $\neg\phi$-states

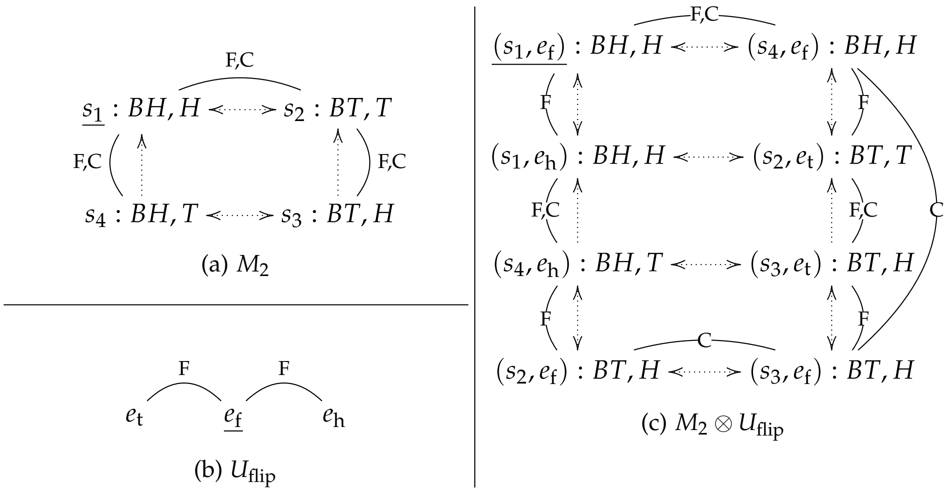


Figure 4.4: Scenario 2: Chiyo ensures that the coin lands heads up

and keeps ϕ-states. So $pre(e_{paq}) = q_1$ and postconditions for all propositions on e_{paq} are *id*.

Thereafter, shown as Figure 4.3c, the updated epistemic betterness structure $M_1 \otimes U_{paq}$ only contains the two states (s_1, e_{paq}) and (s_2, e_{paq}). We have $M_1 \otimes U_{paq}, (s_1, e_{paq}) \models \odot_{Uma}(p_1|\top)$, which means that Uma has an obligation to see to it that Sam is treated. Therefore, we have $M_1, s_1 \models \odot_{Uma}(p_1|q_1) \rightarrow [(U_{paq}, e_{paq})] \odot_{Uma}(p_1|\top)$.

Scenario 8: factual change triggers obligations

In Figure 4.4, T refers to 'the coin lands tails up', H refers to 'the coin lands heads up', BT refers to 'betting on tails', and BH refers to 'betting on heads'. First, we give the priority structure $\mathcal{G}_2$ for Scenario 8. The best state of affairs is betting correctly $((BH \wedge H) \vee (BT \wedge T))$. Any other cases are worse. Let F denote Fumio and let C denote Chiyo.

The initial epistemic betterness structure based on $\mathcal{G}_2$ is M_2 shown as Figure 4.4a. Over M_2, we have $M_2, s_1 \models \neg \odot_F(BT|\top) \wedge \neg \odot_F(BH|\top) \wedge \neg \odot_C(BT|\top) \wedge \neg \odot_C(BH|\top)$ since they cannot see the coin.

In Figure 4.4b, the action model U_{flip} describes the case where Chiyo sees the coin and ensures that the coin lands heads up but Fumio cannot see Chiyo's action. Event e_t represents that Chiyo sees the coin is tails up but does not flip. Event e_h represents that Chiyo sees the coin is heads up but does not flip. Event e_f represents that Chiyo ensures that the coin lands heads up no matter whether it was heads up or tails up. The preconditions are $pre(e_t) = T$, $pre(e_h) = H$, and

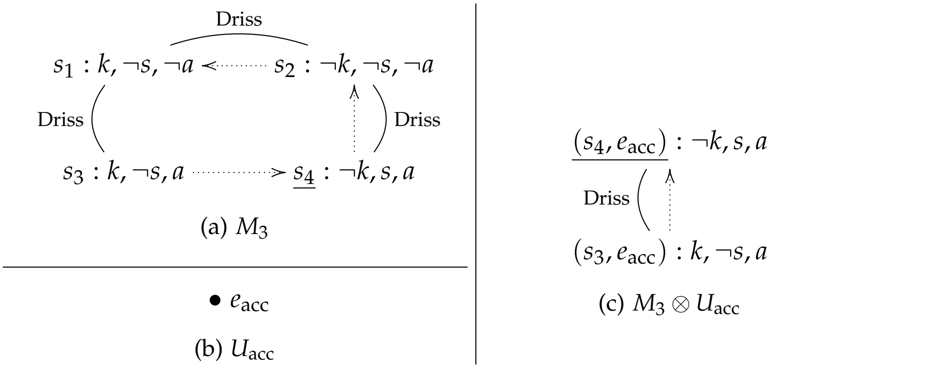


Figure 4.5: Scenario 3: A car accident is happening

$pre(e_f) = \top$. The postconditions of e_t and e_h are *id*. The postconditions of e_f are $post(e_f)(H) = \top$ and $post(e_f)(T) = \bot$.

After Chiyo performs the action, Chiyo has an obligation to bet on heads. Since Fumio does not know whether Chiyo flips the coin, Fumio still does not have an obligation to bet on heads (or on tails). These new obligations can be shown over the updated epistemic betterness structure $M_2 \otimes U_{flip}$ (see Figure 4.4c). We have $M_2 \otimes U_{flip}, (s_1, e_f) \models \odot_C(BH|\top) \wedge \neg \odot_F(BH|\top) \wedge \neg K_F \odot_C(BH|\top)$.

Scenario 9: unconditional obligations are defeasible

In Figure 4.5, k refers to 'Driss keeps promise', a refers to 'a car accident happens', and s refers to 'Driss saves the people involved in the accident'. The priority structure $\mathcal{G}_3$ for Scenario 9 would have the best state of affairs to be those where there is no car accident and Driss keeps the promise ($\neg a \wedge k$). The second best case is that no accident happens ($\neg a$). The third best case is that if an accident happens, then Driss saves the people ($\neg a \vee s$). Other cases are the worst.

Accordingly, we assume that there are only four possible situations in the initial epistemic betterness structure based on $\mathcal{G}_3$ (shown as M_3, Figure 4.5a). We have $M_3, s_4 \models \odot_{Driss}(k|\top) \wedge \neg \odot_{Driss}(s|\top)$, which means that Driss ought to keep his promise unconditionally at that moment, but does not have an unconditional obligation to save the people.

The action model (U_{acc}, e_{acc}) (Figure 4.5b) represents the event of the car accident, where $pre(e_{acc}) = a$ and $post(e_{acc})(a) = \top$. In the updated epistemic betterness structure $M_3 \otimes U_{acc}$ (Figure 4.5c), we have $M_3 \otimes U_{acc}, (s_4, e_{acc}) \models \odot_{Driss}(s|\top) \wedge \neg \odot_{Driss}(k|\top)$, which means that, after seeing the car accident, Driss's unconditional obligation to keep his promise is *overridden* by another unconditional obligation, namely, saving people.

By our analysis on Scenario 9, we would say that epistemic conditional obligations are defeasible. The defeasibility of $\odot_i(\_|\_)$ appears in two different aspects. In the static logic of epistemic conditional obligation $\mathbb{KCDL}$, it invalidates the formula $\odot_i(\phi|\psi) \rightarrow \odot_i(\phi|\psi \wedge \chi)$, which means that a stronger condition could override the old obligation (see Kraus et al. (1990), Governatori et al. (2005)). In the dynamic extension shown in the current paper, the formula $\odot_i(\phi|\top) \rightarrow [(U,e)]\neg\odot_i(\phi|\top)$ is satisfiable, which means that even an unconditional obligation could be released after taking some action. All the unconditional obligations formalized in Scenario 9 can be denoted by *prima facie* obligations, a notion that is strongly related to defeasibility. We will discuss these notions in the following section.

4.4 Information and knowledge-based obligation

In the context of conditional beliefs, van Benthem and Liu (2007) comments that "conditional beliefs *pre-encode* beliefs that we would have if we learnt certain things". Baltag and Smets (2006b) state that "conditional beliefs give descriptions of the agent's *plan* about what he will believe ... after receiving new information". Similarly, we take the view that conditional obligations pre-encode what states of affairs would be the best if specific facts were to hold. In deontic logic terminology, conditional obligations pre-encode the so-called factual detachment of obligations (see Greenspan (1975), Prakken and Sergot (1997)). And, continuing the above analogy, an epistemic conditional obligation pre-encodes what can be referred to as *epistemic detachment*: $\mathbb{KCDL} \models (\odot_i(\phi|\psi) \wedge K_i\psi) \rightarrow \odot_i(\phi|\top)$. An unconditional obligation follows from an epistemic conditional obligation and the knowledge of the antecedent.

We will be using a running example to show how the above intuitions lead to natural formalizations, within language $\mathcal{L}_{\text{DKCDL}}$, of several philosophical notions concerning obligations. In Figure 4.6, $M = \langle W, \sim_i, \leqslant, V\rangle$ where $W = \{n \in \mathbb{N} \mid 1 \leqslant n \leqslant 8\}$; relations $\sim_i$ and $\leqslant$ are as depicted in the figure; $V(P) = \{8\}$, $V(I) = \{n \mid 1 \leqslant n \leqslant 6\}$, $V(L) = \{1,2,4,5\}$, $V(S) = \{1,3,4,6\}$, $V(A) = \{4,5\}$, $V(N) = \{4,6\}$, and $V(C) = W$. Model M describes a scenario where there is a world war and i is the president of a country. i has already come to know that a world war happens but she does not know whether her country is involved in the war. In model M, proposition P refers to 'the world is peaceful', I refers to 'i's country is involved in the war', C refers to 'i protects her civilians', L refers to 'the territorial land is invaded', S refers to 'the territorial sea is invaded', A refers to 'i sends the army', and N refers to 'i sends the navy'. In order to capture different notions concerning obligation, we need to define information sets.

Definition 44 (Information set)**.** *Given a pointed epistemic betterness structure* (M,s) *and a finite set of literals* $Q = \{p_m, \neg p_m \mid 1 \leq m \leq n$ *for some* $n \in \mathbb{N}$ *and* $p_m \in \mathbf{P}\}$,

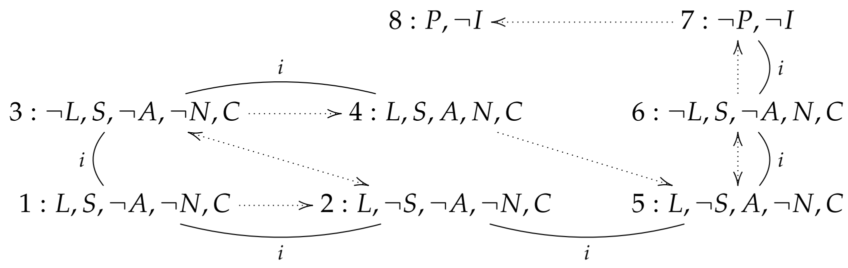


Figure 4.6: M (a president is facing a world war)

let $I \subset Q$ and for each $p_m \in Q$ (or $\neg p_m \in Q$), if $p_m \in I$ ($\neg p_m \in I$), then $\neg p_m \in I$ ($p_m \in I$). Then the information set of state s is $I_s = \{\phi \in I \mid M, s \models \phi\}$.

The set I_s consists of all true facts that have happened when s is the actual world. For the set $Q \setminus I$, it represents the state of affairs that would occur in the future. In the current example, let $I_5 = \{\neg P, I, L, \neg S\}$ be the information set in state 5 representing all states of affairs that have happened, thereby can be learnt by i. The remaining propositions $\{A, \neg N, C\}$ represent the states of affairs that would occur as a result of i's action.

Ideal conditional obligation $\bigcirc(\phi|\psi)$: Hansson's conditional obligations $\bigcirc(\phi|\psi)$ are defined over betterness structures, i.e., $M = \langle S, \leqslant, V \rangle$, where epistemic relations are absent. We call them ideal conditional obligations here to indicate that they describe the obligations regardless of agents' epistemic information. The term 'ideal' is borrowed from Jones and Pörn (1985). Formula $\bigcirc(\phi|\psi)$ can be read as: ϕ is ideally good given the situation ψ. The semantics of $\bigcirc(\phi|\psi)$ is: all best ψ-states also satisfy ϕ, which considers all ontically possible states. Moreover, $\bigcirc(\_|\_)$ is a global operator, which implies that it does not depend on the state at which you evaluate it. Formula $\bigcirc(\phi|\top)$ is a special type of ideal conditional obligations. It describes that ϕ is the ideal state of affairs over all ontically possible states.

In M, we have $M, 5 \models \bigcirc(P|\top) \wedge K_i \neg P$, which means that the president i has an ideal obligation to guarantee a peaceful world regardless of the information set I_5, although she knows that peace is no longer possible.

Epistemic unconditional obligation $\odot_i(\phi|\top)$: Formula $\odot_i(\phi|\top)$ tells the agent what ought to be the case given her current information. In M, over information set I_5, i only knows that $\neg P$. Based on her current information, we have $M, 5 \models \odot_i(\neg I|\top)$. Intuitively, she ought to guarantee that her country is not involved in the war. Arguably, epistemic unconditional obligations correspond to the notion of *absolute obligation* used by McCloskey (1963) to denote those obligations that an agent ought to comply with at a specific moment or under specific information. To be subject to an absolute obligation is "to be in a moral situation with moral commitment". These properties are reflected in the intuition of $\odot_i(\phi|\top)$. In M, we have $M, 5 \models \odot_i(A|L) \wedge \neg K_i L$, which means that i has an obligation to send an army when she knows that their territorial land is invaded, but she does not know that they are invaded. So it means that i does not have an absolute obligation to send the army. But, due to $M, 5 \models \odot_i(\neg I|\top)$, i has an absolute obligation not to have her country involved in the war.

Prima facie obligation $\odot_i^{\mathrm{P}} \phi$: The formalization of *prima facie* obligations via unconditional obligation (ideal obligation in this thesis) was first advanced by Alchourrón (1996). We expand on this tradition here, showing how our formalism also accommodates a natural formalization of this type of obligations. We use ideas from McCloskey (1963)'s analysis to justify our approach. We take as starting point McCloskey's observation that "an actual obligation does not differ 'qualitatively' from a *prima facie* obligation ..." and hence it could be captured by a formula $\odot_i(\phi|\top)$. However, we still need to distinguish unconditional obligations that can be overridden (*prima facie*) from those that cannot. The overriding phenomenon, we argue, has to do with the acquisition of new information that brings about new *prima facie* obligations that override previous ones. This is a dynamic phenomenon, and our framework is well-suited to capture it.

Suppose that we only consider the single-agent case and the actions of truthful public announcements (see Section 4.3.2). Any public announcement introduces some true information. Taking the notation in public announcement logic, given an epistemic betterness structure $M = \langle W, \sim_i, \leqslant, V \rangle$, $M|_\phi = \langle W \cap \|\phi\|_M, \sim_i', \leqslant', V' \rangle$ where $\sim_i'$, $\leqslant'$, and V' are $\sim_i$, $\leqslant$, and V restricted to the set $W \cap \|\phi\|_M$, respectively. We use a new operator $\odot_i^{\mathrm{P}} \phi$ to denote i's *prima facie* obligation to ensure ϕ. So, *prima facie* obligations can be defined as follows.

Definition 45 (*Prima facie* obligation)**.** *Given a pointed epistemic betterness structure (M, s),*

$$M, s \models \odot_i^{\mathrm{P}} \phi \quad \textit{iff} \quad \textit{there exists } \psi \in I_s \textit{ such that } M|_\psi, s \models \odot_i(\phi|\top).$$

The semantics of $\odot_i^{\mathrm{P}} \phi$ means that it is *prima facie* obligatory that ϕ for i if and only if after receiving *some* true information, i has an epistemic unconditional obligation to ensure ϕ. In our example, we have $M, 5 \models \odot_i^{\mathrm{P}} A$ since $M|_L, 5 \models \odot_i(A|\top)$. But $M, 5 \models \neg \odot_i(A|\top)$. These mean that the president has a *prima facie* obligation to send the army once she knows that the territorial land is invaded. But this *prima facie* obligation is currently not an absolute obligation. Similarly, i also has a *prima facie* obligation to send both their army and navy, $\odot_i^{\mathrm{P}}(A \wedge N)$, once she knows that the territorial land and sea are invaded. But this *prima facie* obligation will never become an absolute obligation since S is not true in 5. We argue that the above definition of *prima facie* obligation succeeds in addressing the reservations moved by Prakken and Sergot (1997) to the approach to *prima facie* obligations based on conditional obligations. Some other approaches, e.g., dynamic approach, to study *prima facie* obligations, can be seen in Willer (2016).

All-things-considered obligation $\odot_i^{\mathsf{A}}\phi$: All-things-considered obligations are usually compared with *prima facie* obligations. Prakken and Sergot (1997) state that "To find out what one's duty proper is, one should consider all things, $\cdots$ [it] can be based on any aspect of the factual circumstances and find which one is more incumbent". The statement suggests that an all-things-considered obligation should be the most ideal state of affairs when introducing all true information. It is also strongly related to van der Torre's *exact factual detachment* in the context of objective conditional obligation when all factual premises are given (see Chapter 4.1 in van der Torre (1997)). We define it as follows:

Definition 46 (All-things-considered obligation)**.** *Given a pointed epistemic betterness structure (M, s),*

$$M, s \models \odot_i^{\mathsf{A}}\phi \quad \textit{iff} \quad M|_{\bigwedge I_s}, s \models \odot_i(\phi|\top).$$

$M|_{\bigwedge I_s}$ is the model updated by introducing all information on s. In our example, $M, 5 \models \odot_i^{\mathsf{A}}(A \wedge \neg N)$, which means that the president has an all-things-considered obligation to send the army rather than the navy. However, since she does not know that their territorial land has been invaded at the moment, this all-things-considered obligation is not an absolute obligation *yet*.

Safe knowledge-based obligation $\odot_i^{\mathsf{S}}\phi$: We introduce a type of obligation that, to the best of our knowledge, has not yet been discussed in the literature, but which arises naturally in our framework. We have mentioned that absolute obligations are defeasible given different information. But it is still possible to find some obligations that cannot be defeated by the acquisition of new information. In the study on conditional beliefs, Baltag and Smets (2006b) define *safe beliefs*, where 'safe' means that they are persistent under revision with any true information. Although their definition is founded on a connected plausibility relation, we can follow their idea and define *safe knowledge-based obligations* as follows:

Definition 47 (Safe obligation)**.** *Given an epistemic betterness structure $M = \langle S, \sim_1, \leqslant, V\rangle$, $M, s \models \odot_1^{\mathsf{S}}\phi$ if and only if the following two conditions are satisfied:*

1. *$M, s \models \phi$;*
2. *for each $t, r \in [s]^{\sim_1}$, if $t \in \|\phi\|_M$ and $t \leqslant r$, then $r \in \|\phi\|_M$.*

Intuitively, if $M, s \models \odot_i^{\mathsf{S}}\phi$, then ϕ is satisfied in the actual state and $\|\phi\|_M \cap [s]^{\sim_i}$ is $\leqslant$-upward-closed. As a consequence, it is easy to check that $M, s \models$

$\odot_i^S \phi \to \odot_i(\phi|\top)$. Moreover, for any $\psi \in I_s$, we have $M|_\psi, s \models \odot_i(\phi|\top)$. Thus, $M|_{\bigwedge I_s}, s \models \odot_1(\phi|\top)$. This means that a safe obligation to ensure ϕ will never be defeated by introducing new true information. It will always be an absolute obligation as well as a *prima facie* obligation.

In our example, $\|C\|_M \cap [5]^{\sim_i}$ is $\leqslant$-upward-closed since C is satisfied over the whole set. Thus, $M, 5 \models \odot_i^S C$. This means that i has a safe knowledge-based obligation to protect her civilians, no matter what information she received.

4.5 Reduction and axiomatization

In this section, we will show that each $\mathcal{L}_{\text{DKCDL}}$-formula in the form of $\odot_i(\phi|\psi)$ can be reduced to some $\mathcal{L}_{\text{DEL}}$-formula by a Kangerian-Andersonian reduction (KA-reduction) (see Anderson (1958), Kanger (1970)). In its classical form, the reduction treats deontic operators $\bigcirc\phi$ as $\Box(Q \to \phi)$ where Q denotes a propositional ideality constant standing for 'all obligations are met'. De Lima et al. (2010) defined obligations as $\bigcirc_i\phi =_{def} \langle\langle\varnothing\rangle\rangle(\neg\phi \to vio_i)$ which means that over all outcomes (after arbitrary execution), if $\neg\phi$ is achieved, i will meet a violation (denoted by a constant vio_i). This reduction approach has been explored extensively in the literature, in a variety of settings (see Meyer (1988), De Lima et al. (2010), van Benthem et al. (2014)).

Åqvist (1997) provided KA-reduction for Hansson's dyadic operator $\bigcirc(\_|\_)$ by partitioning all states into a sequence of optimality classes $\{opt_1, \cdots, opt_m\}$, from the best class opt_1 to the worst opt_m. van Benthem et al. (2014) also gave a logically equivalent formula to Hansson's conditional obligation $\bigcirc(\phi|\psi)$ based on priority structures, which generalized Åqvist's formalization. It is shown as follows:

$$[U]((\bigvee_{\langle\phi_1,\cdots,\phi_n\rangle\in S_{\mathcal{G}}} \bigwedge_{1\leq m\leq n} (\langle U\rangle(\phi_m \wedge \psi) \to (\phi_m \wedge \psi))) \to \phi) \tag{4.1}$$

$S_{\mathcal{G}}$ is the set of longest sequences in the priority structure $\mathcal{G}$. Formula (4.1) consists of the universal modality $[U]$ and classical logical connectives. Inspired by (4.1), we first give a key formula schema for our following reduction axioms.

Given a priority structure $\mathcal{G} = \langle\Phi, \prec\rangle$ and an arbitrary formula $\chi \in \Phi \cup \{\top\}$, define $\Phi_\chi = \{\chi' \in \Phi \mid \chi' \succ \chi\}$[1]. Thus, Φ_χ consists of all the formulas in $\mathcal{G}$ better than χ. The KA-reduction of $\odot_i(\phi|\psi)$ relies on the formula:

$$\lambda_\psi^i : \quad \bigvee_{\chi\in\Phi\cup\{\top\}}((\chi \wedge \psi) \wedge K_i(\bigvee \Phi_\chi \to \neg\psi))$$

Formula λ_ψ^i says "ψ is consistent with some χ in the priority structure and agent i knows that any state of affairs that is better than χ (i.e., $\bigvee \Phi_\chi$) must falsify ψ".

[1]If $\chi \in \Phi$ and there is no $\chi' \in \Phi$ such that $\chi' \succ \chi$, let $\Phi_\chi = \varnothing$ and $\bigvee \Phi_\chi = \bot$. If $\chi = \top$, $\Phi_\chi = \Phi$.

Lemma 7. *Given a priority structure $\mathcal{G} = \langle \Phi, \prec \rangle$ and an arbitrary epistemic betterness structure (M, s) based on $\mathcal{G}$, $M, s \models \lambda^i_\psi$ iff $s \in \max_\leqslant([s]^{\sim_i} \cap \|\psi\|_M)$.*

Proof. ($\Rightarrow$) Suppose, to reach a contradiction, that $s \notin \max_\leqslant([s]^{\sim_i} \cap \|\psi\|_M)$. We split the proof into two cases: • Case 1: If $s \notin \|\psi\|_M$, then $M, s \not\models \lambda^i_\psi$. Contradiction. • Case 2: If $s \in \|\psi\|_M$ and s is not the best ψ-state, then there exists $t \in [s]^{\sim_i}$ such that $M, t \models \psi$ and $t > s$. Since there must exist $\chi \in \Phi \cup \{\top\}$ such that $M, s \models \chi \wedge \psi$, we have for any $r > s$ that there exists $\chi' \succ \chi$ (if χ is $\top$, then $\chi' \succ \top$ for each $\chi' \in \Phi$) such that $M, r \models \chi'$. Since $t \in [s]^{\sim_i}$, $M, t \models \bigvee \Phi_\chi \to \neg\psi$. By $t > s$, we have that $M, t \models \chi'$, which implies that $M, t \models \bigvee \Phi_\chi$. So $M, t \models \neg\psi$. Contradiction. Therefore, $s \in \max_\leqslant([s]^{\sim_i} \cap \|\psi\|_M)$.

($\Leftarrow$) Suppose that $s \in \max_\leqslant([s]^{\sim_i} \cap \|\psi\|_M)$. There must exist $\chi \in \Phi \cup \{\top\}$ such that $M, s \models (\chi \wedge \psi) \wedge \neg \bigvee \Phi_\chi$. We then prove by two cases: • Case 1: If there is no $t \in [s]^{\sim_i}$ such that $t > s$, this implies that $\bigvee \Phi_\chi = \bot$. It is trivial that for each $r \sim_i s$, $M, r \models \bigvee \Phi_\chi \to \neg\psi$. So $M, s \models K_i(\bigvee \Phi_\chi \to \neg\psi)$. • Case 2: If there is $t \in [s]^{\sim_i}$ such that $t > s$, there must exist $\chi' \succ \chi$ in Φ such that $M, t \models \chi' \wedge \neg\psi$, which implies that $M, t \models \bigvee \Phi_\chi \to \neg\psi$. As for each $r \sim_i s$ such that $r \not> s$, $M, r \models \neg \bigvee \Phi_\chi$, this also implies that $M, r \models \bigvee \Phi_\chi \to \neg\psi$. So for all $u \sim_i s$, $M, u \models \bigvee \Phi_\chi \to \neg\phi$. Therefore, $M, s \models K_i(\bigvee \Phi_\chi \to \neg\psi)$. □

Therefore, λ^i_ψ captures the best ψ-state among the set of epistemically indistinguishable states for agent i. The outermost operator of λ^i_ψ is not $\odot_i(\_|\_)$.

Proposition 3 (KA-reduction of $\odot_i(\_|\_)$). *Given an epistemic betterness structure (M, s) based on the priority structure $\mathcal{G}$,*

$$M, s \models \odot_i(\phi|\psi) \leftrightarrow K_i(\lambda^i_\psi \to \phi)$$

The proof involves a routine argument. The formula $K_i(\lambda_\psi \to \phi)$ can be read as 'for all states that i cannot distinguish from the real state, all best ψ-states also satisfy ϕ', which is coincident with the interpretation of $\odot_i(\phi|\psi)$. The formula λ^i_ψ can be considered as a relativized constant which plays a similar role with the Kanger's constant Q. But Q can only capture the most ideal state of affairs. λ^i_ψ goes further as it describes the ideality relativized to some specific state of affairs and a certain agent. KA-reduction above helps to reduce each formula in the form of $\odot_i(\phi|\psi)$ to a $\mathcal{L}_{\mathrm{DEL}}$-formula without any dyadic deontic operator.

Another question about the reduction axiom is: what is the reduction axiom for the formula $[(U, e)] \odot_i(\phi|\psi)$? Two different approaches are to be given, which are essentially equivalent. The first way is straightforward. It first reduces the dyadic deontic operator by (KA-reduction) shown in Proposition 3 and then reduces the dynamic operator according to dynamic epistemic logic. The second method goes conversely. Both ways are carried out by the formula λ^i_ψ.

4.5.1 Approach 1: reducing deontic operator - dynamic operator

By Proposition 3, the reduction axiom for the formula $[(U,e)]\bigodot_i(\phi|\psi)$ can be carried out by reducing the dynamic epistemic logic formula $[(U,e)]K_i(\lambda_\psi \to \phi)$.

Proposition 4 (DEL-Reduction axiom (van Ditmarsch and Kooi (2008))). *Given an epistemic model (M,s), the reduction axiom for $[(U,e)]K_i\phi$ is:*

$$M,s \models [(U,e)]K_i\phi \leftrightarrow (pre(e) \to \bigwedge_{e'R_ie} K_i[(U,e')]\phi)$$

Therefore, we can further give the reduction for $[(U,e)]K_i(\lambda^i_\psi \to \phi)$:

Proposition 5. *(Reduction Axiom I) Given an arbitrary epistemic betterness structure (M,s),*

$$M,s \models [(U,e)]\bigodot_i(\phi|\psi) \leftrightarrow (pre(e) \to \bigwedge_{e'R_ie} K_i[(U,e')](\lambda_\psi \to \phi))$$

It deserves noting that the outermost operator of $(pre(e) \to \bigwedge_{e'R_ie} K_i[(U,e')](\lambda_\psi \to \phi))$ (denoted by (𝔟)) is neither a dynamic operator nor a deontic operator. According to Kooi (2007)'s definition of reduction axiom, we find a reduction for $[(U,e)]\bigodot_i(\phi|\psi)$.

4.5.2 Approach 2: reducing dynamic operator - deontic operator

We can also reduce the dynamic operator first and then reduce the dyadic deontic operator by KA-reduction.

Proposition 6. *Given a priority structure $\mathcal{G} = \langle \Phi, \leq \rangle$, over an arbitrary epistemic betterness structure based on $\mathcal{G}$, $[(U,e)]\bigodot_i(\phi|\psi)$ is logically equivalent to the following formula:*

$$\text{(a)} \quad pre(e) \to \bigwedge_{e'R_ie} \bigwedge_{\chi'' \in \Phi} \bigodot_i([(U,e')]\phi \mid (\neg \bigvee \Phi_{\chi''} \wedge [(U,e')]\lambda_\psi))$$

Proof. ($\Leftarrow$) Suppose that $M,s \models$ (a). We then prove it by two cases:

• Case 1: If $M,s \not\models pre(e)$, then $M,s \models [(U,e)]\bigodot_i(\phi|\psi)$ trivially.

• Case 2: If $M,s \models pre(e)$, then $M,s \models \bigwedge_{\chi''\in\Phi} \bigwedge_{e'R_ie} \bigodot_i([(U,e')]\phi | \neg \bigvee \Phi_{\chi''} \wedge [(U,e')]\lambda_\psi)$. So we need to prove $M,s \models [(U,e)]\bigodot_i(\phi|\psi)$. By semantics, we need to prove $pre(e)$ implies that $M \otimes U,(s,e) \models \bigodot_i(\phi|\psi)$. We have supposed that $M,s \models pre(e)$. So we only need to check if $M \otimes U,(s,e) \models \bigodot_i(\phi|\psi)$. Assume, to reach a contradiction, that $M \otimes U,(s,e) \not\models \bigodot_i(\phi|\psi)$. This implies that there is $(s_1,e_1) \in M \otimes U$ such that $(s_1,e_1) \in \max_\leqslant([(s,e)]^{\sim_i} \cap \|\psi\|_{M\otimes U})$ and $M \otimes U,(s_1,e_1) \models \neg\phi$. By Lemma 7, we have $M \otimes U,(s_1,e_1) \models \lambda_\psi$. This means that $M,s_1 \models [(U,e_1)]\lambda_\psi$. Let χ_1 be the formula in Φ such that $M,s_1 \models \chi_1$ and for each $\chi'_1 \succ \chi_1$, $M,s_1 \not\models \chi'_1$. So we have $s_1 \in \max_\leqslant([s]^{\sim_i} \cap \|\neg \bigvee \Phi_{\chi_1} \wedge [(U,e_1)]\lambda_\psi\|_M)$. By $M,s \models \bigwedge_{\chi''\in\Phi} \bigwedge_{e'R_ie} \bigodot_i([(U,e')]\phi \mid \neg \bigvee \Phi_{\chi''} \wedge [(U,e')]\lambda_\psi)$, we have $M,s_1 \models [(U,e_1)]\phi$. So $M \otimes U,(s_1,e_1) \models \phi$. Contradiction. Thus, $M \otimes U,(s,t) \models \bigodot_i(\phi|\psi)$. Therefore, $M,s \models [(U,e)]\bigodot_i(\phi|\psi)$.

($\Rightarrow$) Suppose that $M, s \models [(U,e)] \odot_i(\phi|\psi)$. By semantics, $pre(e)$ implies that $M \otimes U, (s,e) \models \odot_i(\phi|\psi)$. Suppose that $M, s \models pre(e)$. So $M \otimes U, (s,t) \models \odot_i(\phi|\psi)$. Now we prove it by contradiction. Assume that $M, s \not\models \bigwedge_{\chi'' \in \Phi} \bigwedge_{e' R_i e} \odot_i([(U,e')]\phi \mid \neg \bigvee \Phi_{\chi''} \wedge [(U,e')]\lambda_\psi)$. This implies that there exist $\chi_1 \in \Phi$ and $e_1 \in U$ such that $M, s \not\models \odot_i([(U,e_1)]\phi \mid \neg \bigvee \Phi_{\chi_1} \wedge [(U,e_1)]\lambda_\psi)$. This means that there exists $s_1 \in [s]^{\sim_i}$ such that $s_1 \in \max_\leqslant([s]^{\sim_i} \cap \|\neg \bigvee \Phi_{\chi_1} \wedge [(U,e_1)]\lambda_\psi\|_M)$ and $M, s_1 \not\models [(U,e_1)]\phi$. So $M, s_1 \models [(U,e_1)]\lambda_\psi$, which means that $M, s_1 \models pre(e)$ implies that $M \otimes U, (s_1, e_1) \models \lambda_\psi$. If $M, s_1 \not\models pre(e_1)$, this contradicts $M, s_1 \not\models [(U,e_1)]\phi$. So $M, s_1 \models pre(e_1)$. Thus, $M \otimes U, (s_1, e_1) \models \lambda_\psi$. Since $s \sim_i s_1$ and eR_ie_1, $(s_1, e_1) \in [(s,e)]^{\sim_i}$. By Lemma 7, $(s_1, e_1) \in \max_\leqslant([(s,e)]^{\sim_i} \cap \|\psi\|_{M \otimes U})$. By $M \otimes U, (s,e) \models \odot_i(\phi|\psi)$, $M \otimes U, (s_1, e_1) \models \phi$, which contradicts to $M, s_1 \not\models [(U,e_1)]\phi$. Thus, we proved that $M, s \models \bigwedge_{\chi'' \in \Phi} \bigwedge_{e' R_i e} \odot_i([(U,e')]\phi \mid \neg \bigvee \Phi_{\chi''} \wedge [(U,e')]\lambda_\psi)$. □

Hereafter, we just need to reduce the dyadic deontic operator by *KA*-reduction as shown in Proposition 3.

Proposition 7. *Given a priority structure $\mathcal{G} = \langle \Phi, \leq \rangle$, over arbitrary epistemic betterness structure based on $\mathcal{G}$, the following two formulas are equivalent:*

(a) $pre(e) \rightarrow \bigwedge_{\chi'' \in \Phi} \bigwedge_{e' R_i e} \odot_i([(U,e')]\phi \mid \neg \bigvee \Phi_{\chi''} \wedge [(U,e')]\lambda_\psi)$.
(b) $pre(e) \rightarrow \bigwedge_{\chi'' \in \Phi} \bigwedge_{e' R_i e} K_i(\lambda_{\neg \bigvee \Phi_{\chi''} \wedge [(U,e')]\lambda_\psi} \rightarrow [(U,e')]\phi)$

Corollary 2. *(Reduction Axiom II) Given a priority structure $\mathcal{G} = \langle \Phi, \leq \rangle$, over arbitrary epistemic betterness structure based on $\mathcal{G}$, the following two formulas are equivalent:*

$$M, s \models [(U,e)] \odot_i(\phi|\psi) \leftrightarrow \text{(b)}$$

4.5.3 Harmony

We have shown two alternative reductions for the formula $[(U,e)] \odot_i(\phi|\psi)$. It is expected that they are logically equivalent.

Proposition 8. *Given a priority structure $\mathcal{G} = \langle \Phi, \leq \rangle$, over arbitrary epistemic betterness structure based on $\mathcal{G}$, the following two formulas are equivalent:*

(b) $pre(e) \rightarrow \bigwedge_{e' R_i e} K_i[(U,e')](\lambda_\psi \rightarrow \phi)$
(b) $pre(e) \rightarrow \bigwedge_{\chi'' \in \Phi} \bigwedge_{e' R_i e} K_i(\lambda_{\neg \bigvee \Phi_{\chi''} \wedge [(U,e')]\lambda_\psi} \rightarrow [(U,e')]\phi)$

To prove Proposition 8, we only need to check that $K_i[(U,e')](\lambda_\psi \rightarrow \phi)$ is equivalent to $\bigwedge_{\chi'' \in \Phi} K_i(\lambda_{\neg \bigvee \Phi_{\chi''} \wedge [(U,e')]\lambda_\psi} \rightarrow [(U,e')]\phi)$. The proof is omitted since it is not hard. A commutative diagram[2] (Figure 4.7) indicates that both approaches are in harmony. Since approach 1 is simpler and KA-reduction in

[2]A commutative diagram is a diagram such that all directed paths in the diagram with the same start and endpoints lead to the same result.

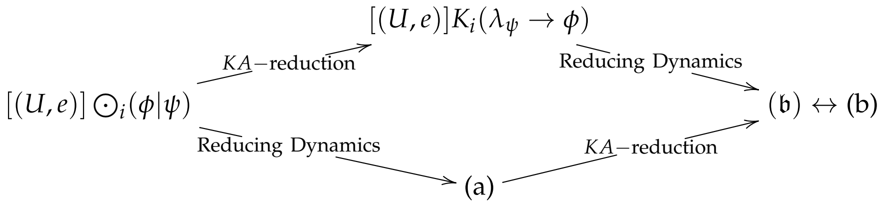


Figure 4.7: The two approaches to reduction

Proposition 3 follows the tradition of deontic logic, we will give the axiomatization based on KA-reduction and formula ($\mathfrak{b}$) will plays an important role in the completeness proof of our axiom system.

4.5.4 Axiomatization $\mathbb{DKCDL}$

It should be noted that λ^i_ψ is defined by some ψ, some $i \in G$ and a certain priority structure. Therefore, our proof system is to be established based on a fixed priority structure $\mathcal{G}$.

Definition 48. *The proof system* $\mathbb{DKCDL}$ *consists of the following axiom schemas and inference rules:*

for each $i \in G$

(TAUT)	*All instances of tautologies*
(K)	$K_i(\phi \to \psi) \to (K_i\phi \to K_i\psi)$
(T)	$K_i\phi \to \phi$
(4)	$K_i\phi \to K_iK_i\phi$
(5)	$\neg K_i\phi \to K_i\neg K_i\phi$
(U-A)	$[(U,e)]p \leftrightarrow (pre(e) \to post(e)(p))$
(U-N)	$[(U,e)]\neg\phi \leftrightarrow (pre(e) \to \neg[(U,e)]\phi)$
(U-C)	$[(U,e)](\phi \wedge \psi) \leftrightarrow ([(U,e)]\phi \wedge [(U,e)]\psi)$
(U-K)	$[(U,e)]K_i\phi \leftrightarrow (pre(e) \to \bigwedge_{e'R_ie} K_i[(U,e')]\phi)$
(KA)	$\odot_i(\phi\|\psi) \leftrightarrow K_i(\lambda^i_\psi \to \phi)$
(MP)	*From ϕ and $\phi \to \psi$, infer ψ*
(N)	*From ϕ, infer $K_i\phi$*
(RE)	*From $\phi \leftrightarrow \psi$, infer $\chi \leftrightarrow \chi[\phi/\psi]$*

$\mathbb{DKCDL}$ is given based on the proof system for dynamic epistemic logic with postconditions **UM** given by van Ditmarsch and Kooi (2008), except (KA), which is given so as to reduce the dyadic deontic operators. (RE) is the inference rule *replacement (substitution) of equivalents*, which is admissible in $\mathbb{DKCDL}$. The notation $\chi[\phi/\psi]$ denotes any formula obtained by replacing one or more occurrences of ψ in χ with ϕ.

Theorem 3. $\mathbb{DKCDL}$ *is sound with respect to the class of epistemic betterness structure.*

Soundness can be obtained straightforwardly from the soundness of DEL and the validity of (KA) (Proposition 3). The basic proof strategy of completeness refers to Chapter 7.4 in van Ditmarsch et al. (2007) and Theorem 11 in Kooi (2007). There are two key points in the completeness proof:

1. All $\mathcal{L}_{\text{DKCDL}}$-formulas are translated to $\mathcal{L}_{\text{EL}}$-formulas by a *translation* function (to be shown in Definition 49) which is induced by KA-reduction and reduction axioms for dynamic operators.

2. In the following induction proofs, some are proved by induction on the *complexity* of $\mathcal{L}_{\text{DKCDL}}$-formulas (to be shown in Definition 50) rather than the structure of formulas.

The complexity measure is used for capturing whether the translation practically reduces the outermost dynamic operators or deontic operators of an $\mathcal{L}_{\text{DKCDL}}$-formula. In this way, we reduce the completeness of $\mathbb{DKCDL}$ to the known completeness of classical epistemic logic $\mathbb{EL}$.

Definition 49. *(Translation) The translation* $t : \mathcal{L}_{\text{DKCDL}} \to \mathcal{L}_{\text{EL}}$ *is defined as follows:*

$$
\begin{array}{lll}
t(p) & = & p \\
t(\neg\phi) & = & \neg t(\phi) \\
t(\phi \wedge \psi) & = & t(\phi) \wedge t(\psi) \\
t(K_i\phi) & = & K_i t(\phi) \\
t(\bigodot_i(\phi|\psi)) & = & K_i(t(\lambda^i_\psi) \to t(\phi)) \\
t([U,e]p) & = & t(pre(e) \to post(e)(p)) \\
t([U,e]\neg\phi) & = & t(pre(e) \to \neg[U,e]\phi) \\
t([U,e](\phi \wedge \psi)) & = & t([U,e]\phi \wedge [U,e]\psi) \\
t([U,e]K_i\phi) & = & t(pre(e) \to \bigwedge_{e' R_i e} K_i[U,e']\phi) \\
t([U.e]\bigodot_i(\phi|\psi)) & = & t(pre(e) \to \bigwedge_{e' R_i e} K_i[U,e'](\lambda^i_\psi \to \phi))
\end{array}
$$

According to Definition 49, $t(\bigodot_i(\phi|\psi))$ is also equivalent to $K_i(\neg(t(\lambda^i_\psi) \wedge \neg t(\phi)))$.

Definition 50. *(Complexity of* $\mathcal{L}_{\text{DKCDL}}$*) The complexity* $c : \mathcal{L}_{\text{DKCDL}} \to \mathbb{N}$ *is defined as follows:*

$$
\begin{array}{lll}
c(p) & = & 1 \\
c(\neg\phi) & = & 1 + c(\phi) \\
c(\phi \wedge \psi) & = & 1 + \max(c(\phi), c(\psi))) \\
c(K_i\phi) & = & 1 + c(\phi) \\
c([U,e]\phi) & = & (3 + |U| + c(U)) \cdot c(\phi) \\
c(\bigodot_i(\phi|\psi)) & = & 3 + \max(c(\lambda^i_\psi), c(\neg\phi))
\end{array}
$$

where $c(U) = \max(c(pre(e_1)), \cdots, c(pre(e_{|U|})), c(post(e_1)(p_1)), \cdots, c(post(e_1)(p_k)), \cdots c(post(e_{|U|})(p_1)), \cdots, c(post(e_{|U|})(p_k)))))$.

Note that our notion of complexity is slightly different from the complexity definition used in dynamic epistemic logic or **UM** given by van Ditmarsch and Kooi (2008). The definition of $c(U)$ is worth stressing. It represents the maximal complexity among all preconditions and postconditions of each event in the action model U. The term $|U|$ represents the number of events in U. The constant 3 occurring in $c([U,e]\phi)$ is chosen for technical reasons. It is the minimal number that makes Lemma 8 work.

Then we prove that all $\phi \in \mathcal{L}_{\text{DKCDL}}$-formulas are syntactically equivalent to their translation in Lemma 8. The proof will be done by induction on the complexity of an arbitrary $\mathcal{L}_{\text{DKCDL}}$-formula ϕ. When ϕ is a propositional atom, its complexity is minimal, i.e., 1. As for the inductive step, we need to show that the translation of each formula with respect to its structure (except propositional atoms) is also syntactically equivalent to itself. The cases for negation, conjunction, epistemic operator and dyadic deontic operator are relatively easy. But when $\phi = [U,e]\psi$, according to our translation shown in Definition 49, we still need to show that the claim holds for different types of ψ respectively, i.e., ψ is a propositional atom, a negation, a conjunction, an epistemic formula or a dyadic deontic formula.

Lemma 8. *For all formulas $\phi \in \mathcal{L}_{\text{DKCDL}}$, it is the case that $\vdash \phi \leftrightarrow t(\phi)$ and $t(\phi) \in \mathcal{L}_{\text{EL}}$.*

Proof. By induction on $c(\phi)$.

• **Base case**: When $\phi = p$ for some propositional atom p, it is trivial that $\vdash p \leftrightarrow p$ and $p \in \mathcal{L}_{\text{EL}}$.

• **Induction hypothesis**: For all ϕ such that $c(\phi) < n$: we have $\vdash \phi \leftrightarrow t(\phi)$ and $t(\phi) \in \mathcal{L}_{\text{EL}}$.

• **Induction step**: If $c(\phi) = n+1$:

- When $\phi = \neg\psi$, we have $c(\neg\psi) = 1 + c(\psi)$. So $c(\psi) = n$. By induction hypothesis, we get $\vdash \psi \leftrightarrow t(\psi)$ and $t(\psi) \in \mathcal{L}_{\text{EL}}$. Thus, $\vdash \neg\psi \leftrightarrow \neg t(\psi)$. It just is $\vdash \phi \leftrightarrow t(\phi)$. And $\neg t(\psi) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = (\psi_1 \wedge \psi_2)$, we have $c(\psi_1 \wedge \psi_2) = 1 + \max(c(\psi_1, c(\psi_2)))$. So $\max(c(\psi_1, c(\psi_2))) = n$. It means that $c(\psi_1) \leq n$ and $c(\psi_2) \leq n$. By induction hypothesis, we have $\vdash \psi_1 \leftrightarrow t(\psi_1)$, $\vdash \psi_2 \leftrightarrow t(\psi_2)$, $t(\psi_1) \in \mathcal{L}_{\text{EL}}$ and $t(\psi_2) \in \mathcal{L}_{\text{EL}}$. Then we have $\vdash (\psi_1 \wedge \psi_2) \leftrightarrow (t(\psi_1) \wedge t(\psi_2))$. It is equivalent to $\vdash (\psi_1 \wedge \psi_2) \leftrightarrow t(\psi_1 \wedge \psi_2)$ by our translation. And $t(\psi_1) \wedge t(\psi_2) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = K_i\psi$: $c(K_i\psi) = 1 + c(\psi)$. So $c(\psi) = n$. By induction hypothesis, we have $\vdash \psi \leftrightarrow t(\psi)$ and $t(\psi) \in \mathcal{L}_{\text{EL}}$. By (NEC) and (K), $\vdash K_i\psi \leftrightarrow K_i(t(\psi))$. It is equivalent to $\vdash K_i\psi \leftrightarrow t(K_i\psi)$ by our translation. And we also have $K_i t(\psi) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = \odot_i(\psi_1|\psi_2)$: $c(\odot_i(\psi_1|\psi_2)) = 3 + \max(c(\lambda^i_{\psi_2}), c(\neg\psi_1))$. So $c(\lambda^i_{\psi_2}) \leq n$ and $c(\neg\psi_1) \leq n$. By induction hypothesis, we have $\vdash$

$\lambda^i_{\psi_2} \leftrightarrow t(\lambda^i_{\psi_2})$, $\vdash \neg\psi_1 \leftrightarrow t(\neg\psi_1)$, $t(\lambda^i_{\psi_1}) \in \mathcal{L}_{\text{EL}}$ and $t(\neg\psi_2) \in \mathcal{L}_{\text{EL}}$. By propositional logic, we have $\vdash \neg\psi_1 \leftrightarrow \neg t(\psi_1)$. Then it follows that $\vdash (\lambda^i_{\psi_2} \wedge \neg\psi_1) \leftrightarrow (t(\lambda^i_{\psi_2}) \wedge \neg t(\psi_1))$. So we have $\vdash \neg(\lambda^i_{\psi_2} \wedge \neg\psi_1) \leftrightarrow \neg(t(\lambda^i_{\psi_2}) \wedge \neg t(\psi_1))$. It implies that $\vdash K_i(\lambda^i_{\psi_2} \to \psi_1) \leftrightarrow K_i(t(\lambda^i_{\psi_2}) \to t(\psi_1))$ by (NEC) and (K). It is equivalent to $\vdash \odot_i(\psi_1|\psi_2) \leftrightarrow t(\odot_i(\psi_1|\psi_2))$ by (KA), (RE) and our translation. And we also have $K_i(t(\lambda^i_{\psi_2} \to \psi_1)) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = [U,e]p$: we know $c([U,e]p) = 3 + |U| + c(U)$.

  We first prove $\vdash \phi \leftrightarrow t(\phi)$. We have $c(pre(e) \to post(e)(p)) = 2 + \max(c(pre(e)), c(\neg post(e)(p)))$. If $c(pre(e)) \geq c(\neg post(e)(p))$, by $c(U) \geq c(pre(e))$, we have $c([U,e]p) > c(pre(e) \to post(e)p)$. If $c(\neg post(e)(p)) \geq c(pre(e))$, then $c(pre(e) \to post(e)(p)) = 3 + c(post(e)(p))$. By $c(U) \geq c(post(e)(p))$ and $|U| \geq 1$, we obtain $c([U,e]p) > c(pre(e) \to post(e)p)$. By induction hypothesis, we have $\vdash (pre(e) \to post(e)(p)) \leftrightarrow t(pre(e) \to post(e)(p))$. It is equivalent to $\vdash [U,e]p \leftrightarrow t([U,e]p)$ by our translation and (RE).

  Then we prove $t(\phi) \in \mathcal{L}_{\text{EL}}$. We have $t([U,e]p) = t(\neg(pre(e) \wedge \neg post(e)(p))) = \neg(t(pre(e)) \wedge t(\neg post(e)(p)))$. We know that $c(pre(e)) \leq c(U)$ and $c(\neg post(e)(p)) \leq c(U) + 1$. So we have $c(pre(e)) \leq n$ and $c(\neg post(e)(p)) \leq n$. By induction hypothesis, we have $t(pre(e)) \in \mathcal{L}_{\text{EL}}$ and $t(\neg post(e)(p)) \in \mathcal{L}_{\text{EL}}$. Thus, $t([U,e]p) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = [U,e]\neg\psi$: $c([U,e]\neg\psi) = (3 + |U| + c(U)) \cdot (1 + c(\psi)) = 3 + |U| + c(U) + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$.

  We first prove $\vdash \phi \leftrightarrow t(\phi)$. We know $c(pre(e) \to \neg[U,e]\psi) = 2 + \max(c(pre(e)), c(\neg[U,e]\psi))$. Since $c(\neg[U,e]\psi) > c(pre(e))$, we have $c(pre(e) \to \neg[U,e]\psi) = 3 + c([U,e]\psi) = 3 + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$. So $c([U,e]\neg\psi) > c(pre(e) \to \neg[U,e]\psi)$. By induction hypothesis, we obtain $\vdash (pre(e) \to \neg[U,e]\psi) \leftrightarrow t(pre(e) \to \neg[U,e]\psi)$. It follows that $\vdash [U,e]\neg\psi \leftrightarrow t([U,e]\neg\psi)$ by (RE) and our translation.

  Then we prove $t(\phi) \in \mathcal{L}_{\text{EL}}$. We know $t([U,e]\neg\psi) = t(pre(e) \to \neg[U,e]\psi) = t(\neg(t(pre(e) \wedge t(\neg post(e)(p)))))$. Since $c(pre(e)) \leq c(U)$ and $c([U,e]\psi) < c([U,e]\neg\psi)$, by induction hypothesis, we have $t(pre(e)) \in \mathcal{L}_{\text{EL}}$ and $t([U,e]\psi) \in \mathcal{L}_{\text{EL}}$. Thus, we obtain $t([U,e]\neg\psi) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = [U,e](\psi_1 \wedge \psi_2)$: $c([U,e](\psi_1 \wedge \psi_2)) = (3 + |U| + c(U)) \cdot (\max(c(\psi_1), c(\psi_2)) + 1)$.

  We first prove $\vdash \phi \leftrightarrow t(\phi)$. Assuming $c(\psi_1) \geq c(\psi_2)$, we have $c([U,e](\psi_1 \wedge \psi_2)) = 3 + |U| + c(U) + 3 \cdot c(\psi_1) + |U| \cdot c(\psi_1) + c(U) \cdot c(\psi_1)$. We also know $c([U,e]\psi_1 \wedge [U,e]\psi_2) = 1 + \max(c([U,e]\psi_1), c([U,e]\psi_2))$. By $c(\psi_1) \geq c(\psi_2)$, we have $c([U,e]\psi_1 \wedge [U,e]\psi_2) = 4 + |U| + c(U) + c(U) \cdot c(\psi_1)$. So $c([U,e](\psi_1 \wedge \psi_2)) > c([U,e]\psi_1 \wedge [U,e]\psi_2)$. By induction hypothesis, we obtain $\vdash ([U,e]\psi_1 \wedge [U,e]\psi_2) \leftrightarrow t([U,e]\psi_1 \wedge [U,e]\psi_2)$. It is equivalent to $\vdash [U,e](\psi_1 \wedge \psi_2) \leftrightarrow t([U,e](\psi_1 \wedge \psi_2))$ by (RE) and our translation.

Then we prove $t(\phi) \in \mathcal{L}_{\mathrm{EL}}$. Assuming $c(\psi_1) \geq c(\psi_2)$, we have $c([U,e](\psi_1 \wedge \psi_2)) = 3 + |U| + c(U) + 3 \cdot c(\psi_1) + |U| \cdot c(\psi_1) + c(U) \cdot c(\psi_1)$. We know $t([U,e](\psi_1 \wedge \psi_2)) = t([U,e]\psi_1) \wedge t([U,e]\psi_2)$. Since $c([U,e]\psi_1) < c([U,e](\psi_1 \wedge \psi_2))$ and $c([U,e]\psi_2) < c([U,e](\psi_1 \wedge \psi_2))$, by induction hypothesis, we have $t([U,e]\psi_1) \in \mathcal{L}_{\mathrm{EL}}$ and $t([U,e]\psi_2) \in \mathcal{L}_{\mathrm{EL}}$. Thus, we obtain $t([U,e](\psi_1 \wedge \psi_2)) \in \mathcal{L}_{\mathrm{EL}}$.

- When $\phi = [U,e]K_i\psi$, $c([U,e]K_i\psi) = 3 + |U| + c(U) + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$.

  We first prove $\vdash \phi \leftrightarrow t(\phi)$. We know $c(pre(e) \rightarrow \bigwedge_{e'R_ie} K_i[U,e']\psi) = 2 + \max(c(pre(e)), c(\neg \bigwedge_{e'R_ie} K_i[U,e']\psi))$. We also know $c(\neg \bigwedge_{e'R_ie} K_i[U,e']\psi) = 1 + c(\bigwedge_{e'R_ie} K_i[U,e']\psi) = 1 + |U| - 1 + \max(c(K_i[U,e_1]\psi), \cdots, c(K_i[U,e_{|U|}]\psi)) = 1 + |U| + \max(c([U,e_1]\psi), \cdots, c([U,e_{|U|}]\psi))$. Let $m \in \mathbb{N}$ such that $1 \leq m \leq |U|$ and $c([U,e_m]\psi) = \max(c([U,e_1]\psi), \cdots, c([U,e_{|U|}]\psi))$. Then $c(pre(e) \rightarrow \bigwedge_{e'R_ie} K_i[U,e']\psi)) = 3 + |U| + c([U,e_m]\psi) = 3 + |U| + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$. So it is easy to see that $c([U,e]K_i\psi) > c(pre(e) \rightarrow \bigwedge_{e'R_ie} K_i[U,e']\psi))$. By induction hypothesis, we obtain $\vdash (pre(e) \rightarrow \bigwedge_{e'R_ie} K_i[U,e']\psi)) \leftrightarrow t(pre(e) \rightarrow \bigwedge_{e'R_ie} K_i[U,e']\psi))$. It is equivalent to $\vdash [U,e]K_i\psi \leftrightarrow t([U,e]K_i\psi)$ by (RE) and our translation.

  Then we prove $t(\phi) \in \mathcal{L}_{\mathrm{EL}}$. We know $t([U,e]K_i\psi) = \neg t(pre(e) \rightarrow \bigwedge_{e'R_ie} K_i[U,e']\psi) = \neg(t(pre(e)) \wedge t(\neg \bigwedge_{e'R_ie} K_i[U,e']\psi))$. We know $c(pre(e)) \leq c(U)$. And $c(\neg \bigwedge_{e'R_ie} K_i[U,e']\psi) = 1 + |U| + \max(c(K_i[U,e_1]\psi), \cdots, c([U,e_m]\psi))$ where $\{e_1, \cdots, e_m\} = [e]^{\sim_i}$. Since for each e_k, $e_l \in U$, we have $c([U,e_k]\psi) = c([U,e_l]\psi)$. So $c(\neg \bigwedge_{e'R_ie} K_i[U,e']\psi) = 1 + |U| + c(K_i[U,e]\psi) = 2 + |U| + c([U,e]\psi) = 2 + |U| + c([U,e]\psi) = 2 + |U| + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$ which is strictly smaller than $c([U,e]K_i\psi)$. By induction hypothesis, $t(\neg \bigwedge_{e'R_ie} K_i[U,e']\psi) \in \mathcal{L}_{\mathrm{EL}}$. Thus, we obtain $t([U,e]K_i\psi) \in \mathcal{L}_{\mathrm{EL}}$.

- When $\phi = [U,e] \odot_i(\psi_1|\psi_2)$: By (KA) and (RE), $\vdash [U,e] \odot(\psi_1|\psi_2) \leftrightarrow [U,e]K_i(\lambda^i_{\psi_2} \rightarrow \psi_1)$.

  We first prove $\vdash \phi \leftrightarrow t(\phi)$. By the above case of $\phi = [U,e]K_i\psi$, we can easily prove that $\vdash \phi \leftrightarrow t(\phi)$.

  Then we prove $t(\phi) \in \mathcal{L}_{\mathrm{EL}}$. By Lemma 8, we have $\vdash [U,e] \odot_i(\psi_1|\psi_2) \leftrightarrow t([U,e] \odot_i(\psi_1|\psi_2))$ and $\vdash [U,e]K_i(\lambda^i_{\psi_2} \rightarrow \psi_1) \leftrightarrow t([U,e]K_i(\lambda^i_{\psi_2} \rightarrow \psi_1))$. So we have $\vdash t([U,e] \odot_i(\psi_1|\psi_2) \leftrightarrow t([U,e]K_i(\lambda^i_{\psi_2} \rightarrow \psi_1)))$. We know $c([U,e] \odot_i(\psi_1|\psi_2)) = 9 + 3 \cdot |U| + 3 \cdot c(U) + 3 \cdot \max(c(\lambda^i_{\psi_2}), c(\neg\psi_1)) + |U| \cdot \max(c(\lambda^i_{\psi_2}), c(\neg\psi_1)) + c(U) \cdot \max(c(\lambda^i_{\psi_2}), c(\neg\psi_1))$. We know $t([U,e]\bigodot_i(\psi_1|\psi_2)) = t(pre(e) \to \bigwedge_{e'R_ie} K_i[U,e'](\lambda^i_{\psi_2} \to \psi_1)) = \neg(t(pre(e)) \wedge t(\neg \bigwedge_{e'R_ie} K_i[U,e'](\lambda^i_{\psi_2} \to \psi_1)))$. We know $c(pre(e)) \leq c(U) \leq n$. And $t(\neg \bigwedge_{e'R_ie} K_i[U,e'](\lambda^i_{\psi_2} \to \psi_1)) = 1 + |U| + 1 + c([U,e](\lambda^i_{\psi_2} \to \psi_1)) = 2 + |U| + (3 + |U| + c(U)) \cdot c(\lambda^i_{\psi_2} \to \psi_1) = 2 + |U| + (3 + |U| + c(U)) \cdot (1 + \max(c(\lambda^i_{\psi_2}), c(\neg\psi_1))) = 2 + |U| + 3 + |U| + c(U) + 3 \cdot \max(c(\lambda^i_{\psi_2}), c(\neg\psi_1))) + |U| \cdot \max(c(\lambda^i_{\psi_2}), c(\neg\psi_1))) + c(U) \cdot \max(c(\lambda^i_{\psi_2}), c(\neg\psi_1)))$ which is strictly smaller than $c([U,e]\bigodot_i(\psi_1|\psi_2))$. By induction hypothesis, $t(\neg \bigwedge_{e'R_ie} K_i[U,e'](\lambda^i_{\psi_2} \to \psi_1)) \in \mathcal{L}_{\text{EL}}$. Thus, $t([U,e]\bigodot_i(\psi_1|\psi_2)) \in \mathcal{L}_{\text{EL}}$.

Therefore, we proved that for each formula $\phi \in \mathcal{L}_{\text{DKCDL}}$, it is the case that $\vdash \phi \leftrightarrow t(\phi)$ and $t(\phi) \in \mathcal{L}_{\text{EL}}$. □

Let Γ be a set of $\mathcal{L}_{\text{DKCDL}}$-formulas and let $t(\Gamma) = \{t(\phi) \mid \phi \in \Gamma)\}$. We need the following lemma to give the final completeness proof.

Lemma 9. *For every set of formulas $\Gamma \cup \{\phi\} \subseteq \mathcal{L}_{\text{DKCDL}}$, $\Gamma \models \phi$ implies that $t(\Gamma) \models_{S5} t(\phi)$.*

Proof. Let $M = \langle S, \sim_1, \cdots, \sim_n, \leqslant, V\rangle$ be an arbitrary epistemic betterness structure and let $M_E = \langle S, \sim_1, \cdots, \sim_n, V\rangle$ be the epistemic model removed the betterness relation from M. We first show that for each $\mathcal{L}_{\text{EL}}$-formula ψ and arbitrary $s \in S$, $M, s \models \psi \Leftrightarrow M_E, s \models_{S5} \psi$.

By induction on the structure of ψ:

- When $\psi = p$ for some propositional atom p: by semantics of $\mathcal{L}_{\text{DKCDL}}$, $M, s \models p$ if and only if $s \in V(p)$. Then, by semantics of $\mathcal{L}_{\text{EL}}$, $M_E, s \models_{S5} p$ if and only if $s \in V(p)$. Thus, we have $M, s \models p \Leftrightarrow M_E, s \models_{S5} p$.

- When $\psi = \neg\chi$: Suppose that $M, s \models \neg\chi$. By semantics of $\mathcal{L}_{\text{DKCDL}}$, we have $M, s \not\models \chi$. By induction hypothesis, we have $M_E, s \not\models_{S5} \chi$. So we have $M_E, s \models_{S5} \neg\chi$. Suppose that $M_E, s \models_{S5} \neg\chi$. By semantics of $\mathcal{L}_{\text{EL}}$, we have $M_E, s \not\models_{S5} \chi$. By induction hypothesis, we have $M, s \not\models \chi$. So we have $M, s \models \neg\chi$.

- When $\psi = (\chi_1 \wedge \chi_2)$: Suppose that $M, s \models (\chi_1 \wedge \chi_2)$. By semantics of $\mathcal{L}_{\text{DKCDL}}$, we have $M, s \models \chi_1$ and $M, s \models \chi_2$. By induction hypothesis, we have $M_E, s \models_{S5} \chi_1$ and $M_E, s \models_{S5} \chi_2$. So we have $M_E, s \models_{S5} (\chi_1 \wedge \chi_2)$. Suppose that $M_E, s \models_{S5} (\chi_1 \wedge \chi_2)$. By semantics of $\mathcal{L}_{\text{EL}}$, we have $M_E, s \models_{S5} \chi_1$ and $M_E, s \models_{S5} \chi_2$. By induction hypothesis, we have $M, s \models \chi_1$ and $M, s \models \chi_2$. So we have $M, s \models (\chi_1 \wedge \chi_2)$.

- When $\psi = K_i\chi$: Suppose that $M, s \models K_i\chi$. By semantics of $\mathcal{L}_{\text{DKCDL}}$, we have $M, t \models \chi$ for all $t \in M$ such that $s \sim_i t$. By inductive hypothesis, we have $M_E, t \models_{S5} \chi$. Since we know that $[s]^{\sim_i}_M = [s]^{\sim_i}_{M_E}$, we have $M_E, s \models_{S5} K_i\chi$. Suppose that $M_E, s \models_{S5} K_i\chi$. By semantics of $\mathcal{L}_{\text{EL}}$, we have $M_E, t \models_{S5} \chi$ for all $t \in M_E$ such that $s \sim_i t$. By inductive hypothesis, we have $M, t \models \chi$. Since we know that $[s]^{\sim_i}_M = [s]^{\sim_i}_{M_E}$, we have $M, s \models K_i\chi$.

Now we proved that for an arbitrary $\mathcal{L}_{\text{EL}}$-formula ψ and for an arbitrary state $s \in S$, $M, s \models \psi \Leftrightarrow M_E, s \models_{S5} \psi$.

Let (N_E, r) be an arbitrary pointed epistemic model such that $N_E, r \models_{S5} t(\Gamma)$. So for each $\theta \in t(\Gamma)$, we have $N, r \models \theta$ by what we proved above where N is the epistemic betterness structure derived from N_E by the betterness relation based on our priority structure. Let $\theta = t(\theta')$ and thus $\theta' \in \Gamma$. According to Lemma 8 and soundness, we have $N, r \models \theta'$. Similarly, for each $\theta'' \in \Gamma$, we have $N, r \models \theta''$. So we have $N, r \models \Gamma$. By $\Gamma \models \phi$, we have $N, r \models \phi$. By Lemma 8 and soundness, we have $N, r \models t(\phi)$. By what we proved above, $N_E, r \models t(\phi)$. Therefore, we proved $t(\Gamma) \models_{S5} \phi$. □

Now we are fully prepared for the final strong completeness proof.

Theorem 4. *(Strong completeness) For every set of formulas* $\Gamma \cup \{\phi\} \subseteq \mathcal{L}_{\text{DKCDL}}$, $\Gamma \models \phi$ *implies* $\Gamma \vdash \phi$.

Proof. Suppose that $\Gamma \models \phi$. By Theorem 3 and $\mathbb{DKCDL} \vdash \phi \leftrightarrow t(\phi)$ (Lemma 8), we have $\Gamma \models t(\phi)$. By Lemma 9, we have $t(\Gamma) \models_{S5} t(\phi)$. By strong completeness of $\mathbb{EL}$ with respect to $S5$, we have $t(\Gamma) \vdash_{\mathbb{EL}} t(\phi)$, which means that there is a syntactic proof S where we can derive $t(\phi)$ from a finite set $\Lambda \subseteq t(\Gamma)$ by $\mathbb{EL}$. Let $\Lambda = \langle \psi_1, \psi_2, \cdots, \psi_m \rangle$. For each $\psi_n \in \Lambda$, we have $\vdash \psi_n \leftrightarrow t(\psi_n)$. We know that S is a sequence of formulas. Then we can give a $\mathbb{DKCDL}$-syntactic proof which can derive ϕ from Λ as follows:

(1)	$\vdash \psi_1 \leftrightarrow t(\psi_1)$
(2)	$\vdash \psi_2 \leftrightarrow t(\psi_2)$
$\vdots$	
(m)	$\vdash \psi_m \leftrightarrow t(\psi_m)$
$\vdots$	S
(m+$\|S\|$+1)	$\vdash \phi$
(m+$\|S\|$+2)	$\vdash \phi \leftrightarrow t(\phi)$
(m+$\|S\|$+3)	$\vdash \phi$

Therefore, we conclude that $\Gamma \vdash \phi$. □

Corollary 3. $\mathbb{DKCDL}$ *is sound and strongly complete with respect to the class of epistemic betterness structures.*

4.6 Discussion and conclusion

In this chapter, we provide a logic used for capturing obligation changes due to information or factual changes. Then we give a list of key points of the article and some possible future research.

Static normative system We introduced priority structures as linguistic resources for referring to the betterness ordering on states of affairs. Priority structures exist independently from epistemic betterness structures, which provide us meta-level criterion on goodness. In the context of deontic logic, they can be regarded as normative systems in the sense that every priority structure prescribes some norms which instructs agents what states of affairs ought to be achieved. In semantics of $\mathbb{DKCDL}$, a priority structure is given and never changes itself. So priority structure, as normative systems, are static.

Dynamic obligations Epistemic betterness structures are used for describing agents' obligations. The dynamics investigated in this chapter are only performed over epistemic betterness structure. We extended the static logic of epistemic conditional obligation $\mathbb{KCDL}$ with a dynamic operator. Accordingly, when an agent's epistemic conditional obligation is triggered (to an unconditional obligation) by getting new information or coming to know that some facts changed, the updated epistemic betterness structure shows the new information and the new betterness relation. Therefore, $\mathbb{DKCDL}$ can explicitly capture obligation change. We showed how this logic can naturally accommodate, in an original way, several key deontic notions such as ideal obligation, absolute obligation, *prima facie* obligation, all-things-considered obligation, and safe knowledge-based obligation.

Axiomatization We established the sound and strongly complete axiom system $\mathbb{DKCDL}$ with respect to epistemic betterness structures. By giving a Kangerian-Andersonian reduction for the deontic operator and reduction axioms for the dynamic operator, we can translate all $\mathcal{L}_{\text{DKCDL}}$-formula to a syntactically equivalent $\mathcal{L}_{\text{EL}}$-formula. Therefore, we can derive the completeness of $\mathbb{DKCDL}$ from the completeness of $\mathbb{EL}$.

Following research As shown in this chapter, agents' knowledge-based obligations are strongly affected by their information. However, in most cases in real life, the obligations are also updates by receiving new commands or by following new rules. For instance, a mother asks her children to open the door for their father. At the moment, the children get a new obligation to open the door due to their mother's command. The command prescribes a new norm that 'the door is open' is a better state of affairs than 'the door is not open'. The mother updates the normative system, which then affects the children's obligation. It inspires us that normative systems can also be dynamified by means of updating priority structures. The issues will be discussed in the upcoming Chapter 5.

Amount of previous research defined obligations with respect to propositions. This chapter also follows the classical approach. But there are growing appealing for action-based obligations where obligations are represented by actions rather than propositions. It is also perfect in line with our intuitions that when you ought to 'do' something, there is an *action* which you should perform. Although there have been a lot of papers formalizing action-based obligations, dynamic epistemic logic would become a brand new framework to represent action-based obligations. It is very natural to formalize obligations as these action models which are able to practically 'improve' the epistemic models. In other words, if an action model leads to an updated epistemic model where each state is better than its counterpart in the previous epistemic model, the action is an obligation. Chapter 6 is going to investigated action-based obligations.

Chapter 5

Making Norms and Following Norms

5.1 Introduction

A normative sentence contains information which is used for either describing some deontic state of affairs or prescribing a new norm. The descriptive use is normally shown in an indicative mood, e.g., 'Su ought to keep his promise.' The prescriptive use usually appears in an imperative mood, e.g., 'Su, keep your promise!'. These different uses of normative sentences involve different deontic logics.

In the *descriptive* sense, normative propositions describe agents' deontic states, thereby having truth values. For example, 'Pieter ought to drive on the right' has a certain truth value under some circumstances. In the Netherlands, it is true, but it is false in England. Therefore, it is natural to construct a logic of normative propositions. Varieties of deontic logic were developed following this approach, such as standard deontic logic (see Chapter 1.1), dyadic deontic logics (see Chapter 1.2), deontic logic in normative systems (see Ågotnes et al. (2007)), dynamic deontic logic (see Meyer (1988)), and a number of deontic Stit logic (see Chapter 1.2), etc.

In the *prescriptive* sense, we pay more attention to norms rather than normative propositions since they norms are not propositions and hence do not have truth values. Norms can be 'commands' (in the pragmatic sense) or can change agents' deontic states. Infinitives or imperatives are generally used for prescribing new norms, e.g., 'It it obligatory for drivers to keep right on the roads.' or 'Boys, come to help your Dad!'. Since imperatives or norms are unable to bear truth values, as a consequence, a puzzling question arises: is there a logic of imperatives? The question is known as *Jørgensen's dilemma* (see Jørgensen (1937)), which is also in line with Poincaré's question: can imperatives be part of logical inferences (see Poincaré (1913))? Let us think about the following inference from Hansen (2008):

John, open the door!
The door cannot be opened unless it is first unlocked.

John, unlock the door!

The above inference is considered to be valid at the first glance. It suggests that one imperative, by some means, can entail another imperative. However, according to Hansen's insightful argument (see Chapter 1 in Hansen (2008)), there is *no* logic of imperatives. In this sense, it seems impossible to establish logical inferences of imperatives. But Hansen also said, "imperatives still can be meaningfully used to determine what obligations arise in a certain situation", which leads us to a logic *about* imperatives.

Several logics earlier than Hansen's argument, in effect, were developed in a similar way to Hansen (2008). The basic strategy is introducing an independent set of norms which can be updated by adding new norms or subtracting norms from it (see Wright (1991), Makinson (1999)). Extensive work can be witnessed in propositional dynamic logic given by Meyer (1988), input/output logic established by Makinson and van der Torre (2007a), update semantics for deontic reasoning studied by van der Torre and Tan (1998) and Willer (2016), and dynamic epistemic deontic logic given by Aucher et al. (2009). There is also a series of studies about changing legal systems in the framework of defeasible logic given by Governatori and Rotolo (2004, 2010), Governatori et al. (2007).

Following the strategy mentioned above, this chapter also splits off normative systems from deontic models. However, our normative systems will be characterized by ideality sequences where norms are not independent with each other. Rather, they are structural. Moreover, the updates on ideality sequences would make effects on conditional obligations by changing betterness relations in deontic models. Accordingly, we give the notion of successful updates and the Jørgensen's dilemma can be resolved.

Outline of this chapter We first review Hansson's conditional obligations as the technical background. Then we show how to induce a betterness structure based on a bare structure and an ideality sequence. Furthermore, prescribing norms can be reflected on the updates on ideality sequences. Then we provide a sound and strongly axiomatization of conditional obligations relativized to normative systems. The notion of successful updates is given afterwards so as to resolve the Jørgensen's dilemma. The comparison with the related work is shown in the last section.

5.2 Preliminaries

We use the language $\mathcal{L}_{\text{DDL}}$ for Hansson's conditional obligations as a basis. Necessary technical background to be used in this chapter will be illustrated. Let **P** be a countable set of propositional atoms. We give the definition of $\mathcal{L}_{\text{DDL}}$ hereby as a reminder.

Definition 51 (Language $\mathcal{L}_{\text{DDL}}$)**.** *The language $\mathcal{L}_{\text{DDL}}$ is given by the following BNF:*

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid U\phi \mid \bigcirc(\phi|\phi)$$

where $p \in \boldsymbol{P}$.

Intuitively, $U\phi$ stands for "ϕ is necessary"[1]; $\bigcirc(\phi|\psi)$ can be read as "if ψ is the case, ϕ ought to be the case". $\hat{U}\phi = \neg U \neg \phi$ which means that ϕ is possible. In Hansson's style, the conditional obligations do not involve any epistemic elements.

5.2.1 Priority sequence as normative system

Priority sequences are a special type of priority structures in the sense that they are a strict linear order. As mentioned in Chapter 2.4, the idea relates to Hansen (2006)'s observations on the logic about imperatives where obligations are decided by a set of prioritized imperatives. Liu (2008) also employed priority sequences to derive preference relations in the context of preference logic.

Definition 52 (Priority sequences)**.** *A priority sequence is a tuple $\mathcal{G} = \langle \Phi, \prec \rangle$ such that:*

- *$\Phi \subset \mathcal{L}_{\text{PL}}$ and Φ is finite (Φ could be empty);*
- *$\prec$ is a strict linear order[2] on Φ such that, for all formulas $\phi, \psi \in \Phi$, if $\phi \prec \psi$ then ψ logically implies ϕ.*

The term $\phi \prec \psi$ represents ψ is strictly better than ϕ.

A priority sequence can be regarded as a normative system to capture which states of affairs are better.

[1] In Definition 1, we use the term $\Box$ to denote the universal operator since $\Box$ is the original term used by Hansson (1969). Now we replace $\Box$ with U to show the meaning of 'universal' behind the operator.

[2] If $\langle A, < \rangle$ is a strict linear order, then $<$ is an antisymmetric, transitive, and trichotomous binary relation over A.

5.2.2 Betterness structure for description

We have already provided a method to derive a betterness relation over states from a given priority structure in Definition 40. Similarly, we hereby define the betterness structure based on a priority sequence. It is used for modelling conditional obligations in Hansson's approach.

Definition 53 (Betterness structures based on priority sequence)**.** *Given a priority sequence $\mathcal{G} = \langle \Phi, \prec \rangle$, $M_{\mathcal{G}} = \langle S, \leqslant_{\mathcal{G}}, V \rangle$ is a betterness structure based on $\mathcal{G}$ where*

- *S is a nonempty set of states;*
- *$V : \mathbf{P} \to \mathcal{P}(S)$;*
- *$\leqslant_{\mathcal{G}}$: $S \times S$ such that $s \leqslant_{\mathcal{G}} s' \iff \forall \phi \in \Phi : s \in \|\phi\|_V \Rightarrow s' \in \|\phi\|_V$.*

Here, $\|\phi\|_V = \{s \in S \mid M, s \models \phi$ and $\phi \in \Phi\}$.

According to Definition 53, $\leqslant_{\mathcal{G}}$ is a total preorder. The notation $M_{\mathcal{G}}$ indicates that the betterness structure $M_{\mathcal{G}}$ is based on the priority sequence $\mathcal{G}$. For example, if $\mathcal{G} = \langle \{p\}, \varnothing \rangle$, then the betterness relation $\leqslant_{\mathcal{G}}$ makes all p-states strictly better than any $\neg p$-states. We will also call them just *betterness structures* if the priority sequence needs not be specified.

According to Proposition 2, we know that if a subset T of the domain of a betterness structure is non-empty, then $\max_{\leqslant_{\mathcal{G}}} \neq \varnothing$. And we know that the semantics of Hansson's conditional obligation $\bigcirc(\phi|\psi)$ means that the obligation addressee has an obligation $\bigcirc(\phi|\psi)$ in s if and only if all the best ψ-states with respect to $\leqslant_{\mathcal{G}}$ in the model also satisfy ϕ. Thus $\bigcirc(\_|\_)$ is a global operator whose semantics are not related to which state we are evaluating on. It *describes* the ideal state of affairs (ϕ) under some certain circumstances (ψ).

According to the definition of betterness structures and the semantics of conditional obligations, a priority sequence determines the obligations of the agents. It is also straightforward to see that any update on a priority sequence can change the betterness relations in the betterness structures based on it, yielding different conditional obligations.

In the following section, we will replace priority sequence with a simpler and more intuitive notion: ideality sequence. It plays similar roles as priority sequence to order states in models.

5.3 Ideality sequence and betterness structure

In this section, we will introduce the notion of ideality sequence. It describes which states of affair are strictly better than another in a more straightforward way. In the aspect of inducing betterness structures, they are essentially equivalent. However, ideality sequences facilitate defining the language of relativized conditional obligations in Section 5.5 in the sense that they do not require any further restrictions on the formulas in it.

Definition 54 (Ideality sequence). *A pair $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ is an ideality sequence where*

- $I \subset \mathcal{L}_{\mathrm{PL}}$ *is a finite (can be empty) set of propositional formulas;*
- $\ll$: $I \times (I \cup \{\epsilon\})$ *is a strict linear order such that for each $\phi \in I$, $\phi \ll \epsilon$.*

An ideality sequence provides us an ordering on several states of affairs denoted by propositional formulas. For example, if $\phi \ll \psi$, it means that ψ is better than ϕ. The idea of it is more straightforward than that of priority sequence. In this chapter, we treat ideality sequences as synonyms for 'normative systems'.

Notations: Given an ideality sequence $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ and $I \neq \varnothing$, for each $\phi \in I$,

- $\mathcal{I}_\phi = \{\phi' \in I \mid \phi \ll \phi' \text{ or } \phi' = \phi\}$;
- if ϕ is not the maximal element in $\mathcal{I}$, $\bigvee I^+_\phi = \bigvee\{\phi' \in I \mid \phi \ll \phi'\}$;
- if ϕ is the maximal element in $\mathcal{I}$, $\bigvee I^+_\phi = \bot$.

The semantic apparatus to be used in the logic of relativized conditional obligations is bare structure.

Definition 55 (Bare structure). *A pair $M = (W, V)$ is a bare structure where*

- *W is a set of states and $W \neq \varnothing$;*
- *$V : \boldsymbol{P} \to \mathcal{P}(W)$ is a valuation.*

A bare structure consists of a domain of states and a valuation. It is equivalent to a betterness structure without the betterness relation. A bare structure provides a factual background. But given an ideality sequence, a betterness relation can be derived from a bare structure and thereafter, a betterness structure is constructed. The way to derive the betterness relation from an ideality sequence on a bare structure is shown as follows.

Definition 56 (Betterness structure based on ideality sequence). *Given an ideality sequence $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ and a bare structure $M = (W, V)$, the betterness structure based on $\mathcal{I}$ is $M_{\mathcal{I}} = (W, \leqslant_{\mathcal{I}}, V)$ where $\leqslant_{\mathcal{I}}$: $W \times W$ is the betterness relation between states satisfying the following condition:*

$$
\begin{array}{lll}
s \leqslant_{\mathcal{I}} t & \iff & \textit{either (i) or (ii), where} \\
 & & \textit{(i) } \exists \phi \in I : (M, t \models \phi \textit{ and } \forall \psi \in I(\phi \ll \psi \to M, s \not\models \psi)) \\
 & & \textit{(ii) } \neg \exists \psi \in I : (M, s \not\models \psi \textit{ or } M, t \not\models \psi)
\end{array}
$$

The definition of $s \leqslant_{\mathcal{I}} t$ intuitively means that the best formula that t satisfies is not worse than the best formula that s satisfies. The relation $\leqslant_{\mathcal{I}}$ is also a total preorder (transitive and strongly connected).

Then we will show that a priority sequence is equivalent to an ideality sequence, up to the betterness structures induced by them, and vice versa.

Definition 57 (Priority sequence induced by ideality sequence)**.** *Given an ideality sequence $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$, the priority sequence induced by $\mathcal{I}$ is $\mathcal{G} = (\Phi, \prec)$ where*

- $\Phi = \{\bigvee I_\phi \mid \phi \in I\}$;
- $\prec: \Phi \times \Phi$. *It satisfies the following condition:* $\bigvee I_\phi \prec \bigvee I_\psi \iff \phi \ll \psi$.

Theorem 5. *Given an ideality sequence $\mathcal{I}$ and the priority sequence $\mathcal{G}_{\mathcal{I}}$ induced by $\mathcal{I}$, we have $\leqslant_{\mathcal{I}} = \leqslant_{\mathcal{G}}$ based on arbitrary bare structure M.*

Proof. ($\Rightarrow$) Suppose s and t are two arbitrary states in M such that $s \leqslant_{\mathcal{I}} t$. By Definition 56, there exists ϕ such that $M, t \models \phi$ and for each $\psi \gg \phi$, we have $M, s \not\models \psi$. Let $\bigvee I_\theta$ be an formula in $\mathcal{G}_{\mathcal{I}}$ such that $M, s \models \bigvee I_\theta$. By $s \leqslant_{\mathcal{I}} t$, we have $\theta \ll \phi$. Since $M, t \models \phi$, it holds that $M, t \models \bigvee I_\phi$. By the definition of $\mathcal{G}_{\mathcal{I}}$, we have $M, t \models \bigvee I_\phi \to \bigvee I_\theta$. So $M, t \models \bigvee I_\theta$. Thus, $s \leqslant_{\mathcal{G}_{\mathcal{I}}} t$.

($\Leftarrow$) Suppose s and t are two arbitrary states in M such that $s \leqslant_{\mathcal{G}_{\mathcal{I}}} t$. This means that for each $\bigvee I_\phi \in \mathcal{G}_{\mathcal{I}}$ such that $M, s \models \bigvee I_\phi$, we have $M, t \models \bigvee I_\phi$. Let $\bigvee I_\psi$ be the maximal element in $\mathcal{G}_{\mathcal{I}}$ such that $M, s \models \bigvee I_\psi$. This also implies that $M, s \models \psi$. Then we have $M, t \models \bigvee I_\psi$. If $M, t \models \psi$, we find ψ such that $M, t \models \psi$ and for each $\theta \gg \psi$, it holds that $M, s \not\models \theta$. If $M, t \not\models \psi$, by $M, t \models \bigvee I_\psi$, there must exist $\chi \gg \psi$ such that $M, t \models \chi$. So we also find χ such that $M, t \models \chi$ and for each $\theta \gg \chi$, we have $M, s \not\models \theta$. Thus, we proved that $s \leqslant_{\mathcal{I}} t$. □

Theorem 5 shows that the betterness structure based on the ideality sequence $\mathcal{I}$ is equals to the betterness structure based on the priority sequence induced by $\mathcal{I}$.

Definition 58 (Ideality sequence induced by priority sequence)**.** *Given an priority sequence $\mathcal{G} = (\Phi, \prec)$, the ideality sequence induced by $\mathcal{G}$ is $\mathcal{I}_{\mathcal{G}} = (I_{\mathcal{G}} \cup \{\epsilon\}, \ll)$ where*

- $I_{\mathcal{G}} = \Phi$;
- *for each* $\phi, \psi \in I$, $\phi \ll \psi$ *iff* $\phi \prec \psi$.

Theorem 6. *Given a priority sequence $\mathcal{G} = (\Phi, \prec)$ and the ideality sequence $\mathcal{I}_{\mathcal{G}} = (I_{\mathcal{G}} \cup \{\epsilon\}$ induced by $\mathcal{G}$, there holds that $\leqslant_{\mathcal{G}} = \leqslant_{\mathcal{I}_{\mathcal{G}}}$ based on arbitrary bare strcture M.*

Proof. ($\Rightarrow$) Suppose that $s \leqslant_{\mathcal{I}_{\mathcal{G}}} t$. By Definition 56, there exists $\phi \in I_{\mathcal{G}}$ such that $M, t \models \phi$ and for each $\psi \in I_{\mathcal{G}}$ with $\psi \gg \phi$, $M, s \not\models \psi$. Let θ be an arbitrary formula in Φ and $M, s \models \theta$. Since $I_{\mathcal{G}} = \Phi$, we then prove it by three cases. When

$\theta \gg \phi$, it contradicts to $s \leqslant_{\mathcal{I}_\mathcal{G}} t$. When $\theta = \phi$, we have $M, t \models \theta$ by $M, t \models \phi$. When $\theta \ll \phi$, by Definition 58, $\theta \prec \phi$. By Definition 52, $\models \phi \to \theta$. So we have $M, t \models \theta$. Thus, we proved that $s \leqslant_\mathcal{G} t$.

($\Leftarrow$) Suppose that $s \leqslant_\mathcal{G} t$. By Definition 53, for each $\phi \in \Phi$, $M, s \models \phi$ implies that $M, t \models \phi$. Let $\psi \in \Phi$ be the formula such that $M, t \models \psi$ and for each θ such that $\psi \prec \theta$, $M, t \not\models \theta$. Assume that $M, s \models \theta$. By $M, s \models \phi \Longrightarrow M, t \models \phi$ for each $\phi \in \Phi$, we have $M, t \models \theta$. Contradiction. Since $\Phi = I_\mathcal{G}$, we find a formula $\psi \in I_\mathcal{G}$ such that $M, t \models \psi$ and for each $\theta \in I_\mathcal{G}$ with $\psi \ll \theta$, $M, s \not\models \theta$. Thus, we proved that $s \leqslant_{\mathcal{I}_\mathcal{G}} t$. □

Now we know that ideality sequence and priority sequence are actually equivalent up to the betterness relation induced by them given a certain bare structure. And since different ideality sequences would induce different betterness relations on a given bare structure, we do not set betterness structures as the semantics apparatus in this chapter. Rather, bare structures are used as the models in this chapter for characterizing conditional obligations relativized to different ideality sequences.

5.4 Making norms: generating normative systems

There have witnessed amount of research about the updates on obligations in deontic logic (see van der Torre and Tan (1998), Mastop (2011), Yamada (2006, 2008, 2011)). In this section, we will show several ways of updating a given ideality sequence. It can be considered as an action performed by a commander. In the remaining part, we treat ideality sequences as synonyms for 'normative systems'.

We subsequently introduce four types of updates among which two types of updates can derive all the other updates. Given an ideality sequence $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$, we use $\mathcal{I} = \langle \phi_1, \phi_2, \cdots, \phi_n, \epsilon \rangle$ to denote the sequence of formulas in $\mathcal{I}$ with respect to $\ll$, where $\phi_i \ll \phi_j$ if $i \leq j$.

5.4.1 Four possible updates

The first fundamental update is *deletion*. As the name suggests, 'deleting a norm' from a normative system means removing a formula from a given ideality sequence. The definition is given as follows:

Definition 59 (Deletion). *Given an ideality sequence $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ and ϕ_1 is the least formula in $\mathcal{I}$, deleting ϕ_1 from $\mathcal{I}$ brings about the ideality sequence $\mathcal{I} - \phi_1$ where ϕ_1 in $\mathcal{I}$ is removed and the betterness order $\ll$ for the remaining formulas is preserved.*

The deletion update captures abolishing a norm. It is related to the notion of repeal or annulment in law. For example, in 1997, the Chinese government

$$\phi_1 \dashrightarrow \phi_2 \dashrightarrow \phi_3 \xrightarrow{\text{delete } \phi_1} \phi_2 \dashrightarrow \phi_3$$

Figure 5.1: An example of deletion

$$\phi_1 \dashrightarrow \phi_2 \xrightarrow{\text{postfix } \phi_0} \phi_0 \dashrightarrow \phi_1 \dashrightarrow \phi_2$$

Figure 5.2: An example of postfixing

decriminalized speculation. Figure 5.1 is an example of deletion, where the dashed arrows from ϕ_i to ϕ_j represent $\phi_i \ll \phi_j$.

The second update is *postfixing* which was originally proposed by Liu (2008) to add a least norm to the tail of a priority sequence. In the current paper, the definition is modified to adapt to ideality sequence. When the original ideality sequence is $(\varnothing \cup \{\epsilon\}, \ll)$, we just put the new formula into the empty set, making an ideality sequence consisting of a single norm and ϵ. The definition of postfixing is given as follows.

Definition 60 (Postfixing (Liu (2008))). *Let $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ and $\phi \notin I$.*

- *If $I \neq \varnothing$ and $\mathcal{I} = \langle \phi_1, \cdots, \phi_n, \epsilon \rangle$, then postfixing ϕ to $\mathcal{I}$ is $\mathcal{I} \lhd \phi = \langle \phi, \phi_1, \cdots, \phi_n, \epsilon \rangle$.*
- *If $I = \varnothing$, then postfixing ϕ to $\mathcal{I}$ is $\mathcal{I} \lhd \phi = \langle \phi, \epsilon \rangle$.*

Postfixing introduces a sub-ideal state of affairs to the original normative system. It is relevant to the notion of 'derogation' in law. For example, the law says no killing people. So not killing people is better than killing people. Then a new provision is added that if you kill a person out of self-defense, you are exonerated. It implies that killing a person out of self-defense is better than murder. An example of postfixing is shown in Figure 5.2.

The third update is *prefixing* which can be originally found in van Benthem et al. (2014). It is used for adding a better state of affairs than the original best state of affairs. For example, according to some law, the best state of affairs is not killing people. If there is a law which introduces a stronger requirement that 'it is obligatory to not kill people and to not be aggressive against others', then the best state of affairs becomes not killing people *and* not being aggressive. We also propose an adaptive modification to ideality sequences.

Definition 61 (Prefixing (van Benthem et al. (2014))). *Given are an ideality sequence $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ and $\phi \notin I$. Let $\mathcal{I} = \langle \phi_1, \cdots, \phi_n, \epsilon \rangle$. Prefixing of $\mathcal{I}$ by ϕ yields the ideality sequence $\mathcal{I} \rhd \phi = \langle \phi_1, \cdots, \phi_n, \phi, \epsilon \rangle$.*

The last update is insertion. It is employed to refine our ideality sequence so as to make our normative system include more norms. For example, when measuring the penalty for the crime of corruption, different amounts of money that a person accepts will lead to different sentencing. The initial law says 20 years in prison for embezzling €100,000 and 5 years in prison for embezzling €10,000. Then a new provision is inserted that 10 years in prison for embezzling €50,000. The measuring on the penalty is refined after the insertion.

Definition 62 (Insertion)**.** *Given an ideality sequence* $\mathcal{I} = \langle \phi_1, \cdots, \phi_{i-1}, \phi_i, \cdots, \phi_n, \epsilon \rangle$*, then inserting a norm* ϕ *into* $\mathcal{I}$ *at position* i *yields the ideality structure* $\mathcal{I}^{\frown \phi, i} = \langle \phi_1, \cdots, \phi_{i-1}, \phi, \phi_i, \cdots, \phi_n, \epsilon \rangle$*.*

In the technical level, not all types of updates are elementary since we can construct any ideality sequence from a given ideality sequence by only deletion and postfixing. The idea is straightforward: delete all the formulas in the original one and then postfix the formulas one by one according to the order of the target sequence.

Fact 6. *Let* $\mathcal{I} = \langle \phi_1, \cdots, \phi_n, \epsilon \rangle$ *and* $\mathcal{I}' = \langle \psi_1, \cdots, \psi_n, \epsilon \rangle$ *be two arbitrary ideality sequences.* $\mathcal{I}'$ *can be built from* $\mathcal{I}$ *by deletion and postfixing.*

Proof. • **Step 1**: Delete each formula in $\mathcal{I}$.

$$\begin{array}{ll} (1) & \text{delete } \phi_0 : \ \mathcal{I} - \phi_0 = \langle \phi_1, \phi_2, \cdots, \phi_n, \epsilon \rangle \\ (2) & \text{delete } \phi_1 : \ (\mathcal{I} - \phi_0) - \phi_1 = \langle \phi_2, \cdots, \phi_n, \epsilon \rangle \\ \vdots & \\ (n+1) & \text{delete } \phi_n : \ ((\mathcal{I} - \phi_0) - \phi_1) - \cdots \phi_n = \langle \epsilon \rangle \end{array}$$

• **Step 2**: Build up $\mathcal{I}'$ from ϵ by postfixing.

$$\begin{array}{ll} (1) & \text{postfix } \psi_m : \ \text{by Definition 60, postfixing } \psi_m \text{ to } \langle \epsilon \rangle \text{ yields } \langle \epsilon \rangle \lhd \psi_m = \langle \psi_m, \epsilon \rangle. \\ (2) & \text{postfix } \psi_{m-1} : \ (\langle \epsilon \rangle \lhd \psi_m) \lhd \psi_{m-1} = \langle \psi_{m-1}, \psi_m, \epsilon \rangle. \\ \vdots & \\ (m+1) & \text{postfix } \psi_0 : \ (((\langle \epsilon \rangle \lhd \psi_m) \lhd \psi_{m-1}) \lhd \cdots \psi_0 = \langle \psi_0, \psi_1, \cdots, \psi_{m-1}, \psi_m, \epsilon \rangle. \end{array}$$

Since $(((\langle \epsilon \rangle \lhd \psi_m) \lhd \psi_{m-1}) \lhd \cdots \psi_0 = \mathcal{I}'$, we therefore build up $\mathcal{I}'$ from $\mathcal{I}$ by deletion and postfixing. When $\mathcal{I} = \langle \epsilon \rangle$, we only need to do Step 2. When $\mathcal{I}' = \langle \epsilon \rangle$, we only need to do Step 1. □

Consequently, prefixing or inserting a norm on a given ideality sequence can also be expressed with several updates only by deletion and postfixing.

5.4.2 A case study

In this section, we show how different updates on an ideality sequence result in obligation changes by using a case study.

A

(a) $(\epsilon, \ll)$ (b) $(\epsilon, \ll) \lhd A$

Figure 5.3: Creating the first norm A

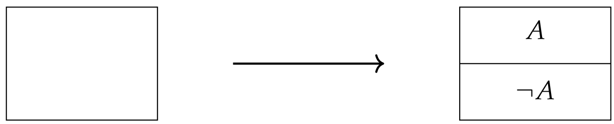


Figure 5.4: M_1 (left) and M_2 (right)

Scenario The company ADD plans to bid for some tenders, but they have not received any information on opening tenders yet. There are two firms A and B which would open tenders in several days. Since ADD does not know anything about tenders A and B, they currently do not have any preferences on these tenders. So we assume that their initial ideality sequence is $(\epsilon, \ll)$.

We will expand on the scenario discussing some more concrete possible events. Let proposition A refer to 'ADD prepares for tender A', let B refer to 'ADD prepares for tender B', and let Q refer to 'ADD supplies products in good quality'.

Event 1. *The firm A announces that they are opening a tender. ADD gets the information and now it is obligatory for them to bid for A.*

Update 1: creating an initial norm The first update on $(\epsilon, \ll)$ is establishing the first norm in it. The update can be expressed by postfixing a norm to $(\epsilon, \ll)$.

In Event 1, the firm A opens their tender, which makes 'preparing for tender A' more preferred than any other choice. Accordingly, the update on ideality sequence in Event 1 is shown in Figure 5.3. Then, the betterness relation in the original betterness structure will change by the update on the ideality sequence (see Figure 5.4). M_1 is derived based on $(\epsilon, \ll)$ and M_2 is derived based on $(\epsilon, \ll) \lhd A$. In Figure 5.4, the top layer in M_2 is the set of A-states and the bottom layer is the set of $\neg A$-states, which means that all A-states are better than all $\neg A$-states. In all of the following figures of betterness structures, the higher layers are strictly better than the lower layers.

Event 2 (Continuing Event 1)**.** *After the firm A opens their tender, they add a clause that if any company can supply products in very good quality, the company will get an additional remuneration.*

Update 2: updating the best norm The firm A announces the new clause which creates a better state of affairs than 'preparing for tender A'. So, currently,

A

(a) $(\epsilon, \ll) \triangleleft A$

$A \dashrightarrow A \wedge Q$

(b) $((\epsilon, \ll) \triangleleft A) \triangleright Q$

Figure 5.5: Prefixing the norm Q (the dashed arrow means $A \ll (A \wedge Q)$)

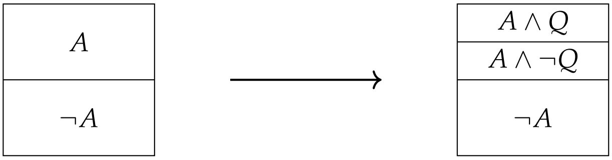


Figure 5.6: M_2 (left) and M_3 (right)

the most preferred state of affairs is 'preparing for tender A *and* supplying products in good quality'. The update on the ideality sequence is shown in Figure 5.5. In Figure 5.5b, the one-way dashed arrow from A to $A \wedge Q$ denotes that $A \ll (A \wedge Q)$. The update on the betterness structure is shown in Figure 5.6. M_3 is derived based on $((\epsilon, \ll) \triangleleft A) \triangleright Q$. Prefixing the norm Q divides the set of A-states into two layers in M_3: $(A \wedge Q)$-states and $(A \wedge \neg Q)$-states. $\neg A$ is still the worst state of affairs.

Event 3 (Continuing Event 1)**.** *After the firm A opens their tender, ADD receives new information that the firm B also opens a tender. But the remuneration paid by B is less than A's. So ADD still considers 'preparing for tender A' their most preferred and 'preparing for tender B' as the second preferred.*

Update 3: adding a least sub-ideality As mentioned above, postfixing introduces a worse state of affairs than the original worst state of affairs. In other words, postfixing a norm prescribes a new sub-ideal state of affairs. In Event 3, the firm B opens their tender after A did. But bidding for B will get less profit than bidding for A. In this case, the best state of affairs is still 'preparing for tender A'. The second best state of affairs is to bid for B. The update on the ideality sequence is shown in Figure 5.7. The update on the betterness structure is shown in Figure 5.8. M_4 is derived based on $((\epsilon, \ll) \triangleleft A) \triangleleft B$. Postfixing a norm $(A \vee B)$ layers the set of $\neg A$-states in M_2, forming two separated layers in M_4: $(\neg A \wedge B)$-states and $(\neg A \wedge \neg B)$-states. A is still the best state of affairs.

Event 4 (Continuing Event 3)**.** *After the firm A and B open their tenders, the firm C also announces that they will open their tender. Comparing the remuneration provided by A, B, and C, ADD finds that A is still the best, C is the second best and B is the worst.*

Update 4: refining the sequence In this case, we insert a new norm into the ideality sequence. The current best state of affairs for ADD is still 'preparing

(a) $(\epsilon, \ll) \lhd A$ (b) $((\epsilon, \ll) \lhd A) \lhd B$

Figure 5.7: Postfixing the norm B

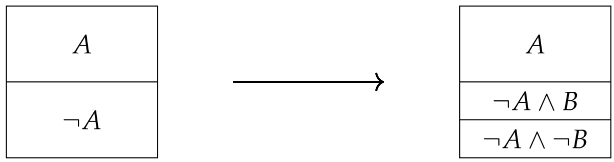


Figure 5.8: M_2 (left) and M_4 (right)

for tender A'. The second best should be 'preparing for tender C if bidding for A is failed'. The third best is 'preparing for B if neither of bids for A nor C successes'. The worst case is none of bids is successful. The updates on the priority sequence and the betterness structure are shown in Figure 5.9 and Figure 5.10 respectively. M_5 is derived based on $(((\epsilon, \ll) \lhd A) \lhd B)^{\curvearrowleft C}$. The set of $\neg A$-states (including $(\neg A \wedge B)$-states and $(\neg A \wedge \neg B)$-states) in M_4 will be refined by the new norms $(A \vee C)$ and $(A \vee C \vee B)$.

Event 5 (Continuing Event 4)**.** *After the firm A, B, and C open their tenders, C announces that they have to quit their tender for some reasons. Consequently, bidding for B if bidding for A is unsuccessful becomes second preferred.*

Update 5: removing a norm The firm C quits their tender, which leads to the case that bidding for C will not bring any profit for ADD. Thus, we just need to delete the norm C from the ideality sequence $(((\epsilon, \ll) \lhd A) \lhd B)^{\curvearrowleft C}$ (see

(a) $((\epsilon, \ll) \lhd A) \lhd B$ (b) $(((\epsilon, \ll) \lhd A) \lhd B)^{\curvearrowleft C}$

Figure 5.9: Inserting the norm C

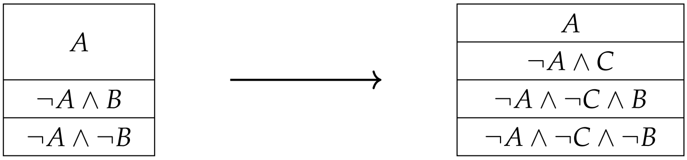


Figure 5.10: M_4 (left) and M_5 (right)

(a) $(((\epsilon, \ll) \lhd A) \lhd B)^{\frown C}$ (b) $((((\epsilon, \ll) \lhd A) \lhd B)^{\frown C}) - C$

Figure 5.11: Removing a norm

Figure 5.11). The update on the betterness structure is shown in Figure 5.12. M_6 is derived based on $(((\epsilon, \ll) \lhd A) \lhd B) - C$. The set of $(\neg A \wedge C)$-states in M_5 merges with the set of $(\neg A \wedge \neg C \wedge B)$-states. M_6 shows that 'preparing for B' is as good as 'preparing for C'.

5.5 The logic ℙℂ𝔻𝕃

In this section, we will establish the logic of relativized conditional obligations ℙℂ𝔻𝕃. The subscript PCDL is an acronym of 'Prescriptive Conditional Deontic Logic'.

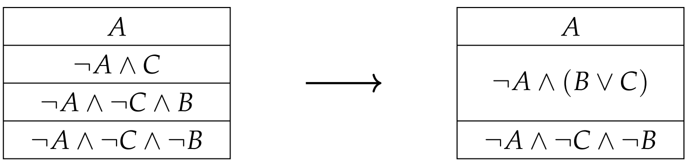


Figure 5.12: M_5 (left) and M_6 (right)

5.5.1 Language and semantics

Definition 63. *The language $\mathcal{L}_{\text{PCDL}}$ is given by the following BNF:*

$$\begin{array}{l}\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid U\phi \mid \bigcirc_{\mathcal{I}}(\phi|\phi)\\ \mathcal{I} ::= \epsilon \mid \chi;\mathcal{I}\end{array}$$

where $p \in \mathbf{P}$ and $\chi \in \mathcal{L}_{\text{PL}}$.

The formula $\bigcirc_{\mathcal{I}}(\phi|\psi)$ represents that based on the ideality sequence $\mathcal{I}$, it ought to be ϕ given ψ.

Definition 64 (Semantics of $\mathcal{L}_{\text{PCDL}}$). *Let $M = \langle W, V\rangle$ be an arbitrary bare structure. The semantics of $\mathcal{L}_{\text{PCDL}}$ is given as follows (only nontrivial cases):*

$$\begin{array}{lll}M,s \models U\phi & \textit{iff} & \|\phi\|_M = W;\\ M,s \models \bigcirc_{\mathcal{I}}(\phi|\psi) & \textit{iff} & \max_{\leqslant_{\mathcal{I}}} \|\psi\|_M \subseteq \|\phi\|_M;\end{array}$$

The semantics of $\bigcirc_{\mathcal{I}}(\phi|\psi)$ is also equivalent to $M_{\mathcal{I}}, s \models \bigcirc(\phi|\psi)$ where $\bigcirc(\phi|\psi)$ is Hansson's conditional obligations and $M_{\mathcal{I}}$ is the betterness structure induced by the bare structure M and the ideality sequence $\mathcal{I}$. The truth values of $\bigcirc_{\mathcal{I}}(\phi|\psi)$ is decided by the ideality sequence $\mathcal{I}$.

Based on the semantics given above, we show two valid formulas involving deletion and postfixing updates. Both formulas are related to the concept of successful updates to be introduced in Section 5.6.2

Fact 7. *The following two formulas are valid:*

$$\begin{array}{ll}(1) & \bigcirc_{\phi_1;\mathcal{I}}(\phi_1|\neg\bigvee I) \wedge \hat{U}(\neg\bigvee I \wedge \neg\phi_1) \to \neg\bigcirc_{\mathcal{I}}(\phi_1|\neg\bigvee I);\\ (2) & \neg\bigcirc_{\mathcal{I}}(\phi_1|\neg\bigvee I) \wedge \hat{U}(\neg\bigvee I \wedge \phi_1) \to \bigcirc_{\mathcal{I}\triangleleft\phi_1}(\phi_1|\neg\bigvee I);\end{array}$$

Proof. (1) Let (M,s) be an arbitrary bare structure such that $M,s \models \bigcirc_{\phi_1;\mathcal{I}}(\phi_1|\neg\bigvee I) \wedge \hat{U}(\neg\bigvee I \wedge \neg\phi_1)$. By $M,s \models \hat{U}(\neg\bigvee I \wedge \neg\phi_1)$, we have $\|\neg\bigvee I\|_M \cap \|\neg\phi_1\|_M \neq \varnothing$. By the definition of $\leqslant_{\mathcal{I}}$, for each $r,t \in \|\neg\bigvee I\|_M$, we have $r \leqslant_{\mathcal{I}} t$ and $t \leqslant_{\mathcal{I}} r$. This means that $\|\neg\bigvee I\|_M = \max_{\leqslant_{\mathcal{I}}} \|\neg\bigvee I\|_M$. So we have $M,s \models \neg\bigcirc_{\mathcal{I}}(\phi_1|\neg\bigvee I)$

(2) Let (M,s) be an arbitrary bare structure such that $M,s \models \neg\bigcirc_{\mathcal{I}}(\phi_1|\neg\bigvee I) \wedge \hat{U}(\neg\bigvee I \wedge \phi_1)$. By $M,s \models \hat{U}(\neg\bigvee I \wedge \phi_1)$, we have $\|\neg\bigvee I\|_M \cap \|\phi_1\|_M \neq \varnothing$. By Proposition 2, we have $\max_{\leqslant_{\mathcal{I}\triangleleft\phi_1}} \|\neg\bigvee I\|_M \neq \varnothing$. Let t be an arbitrary state in $\max_{\leqslant_{\mathcal{I}\triangleleft\phi_1}} \|\neg\bigvee I\|_M$. Assume that $M,t \not\models \phi_1$. Let $r \in \|\neg\bigvee I\|_M \cap \|\phi_1\|_M$. By the definition of $\leqslant_{\mathcal{I}\triangle\phi_1}$, we have $t < r$. This contradicts to $t \in \max_{\leqslant_{\mathcal{I}\triangleleft\phi_1}} \|\neg\bigvee I\|_M$. Thus, $M,t \models \phi_1$. Therefore, we proved that $M,s \models \bigcirc_{\mathcal{I}\triangleleft\phi_1}(\phi_1|\neg\bigvee I)$ □

Formula (1) represents that deleting the norm ϕ_1 defeats the obligation to see to it that ϕ_1 when $\neg\phi_1$ is possible under the certain condition $\neg\bigvee I$. Formula (2) means that postfixing the norm ϕ_1 brings about the new obligation to see to it that ϕ_1 when ϕ_1 is possible under the certain condition $\neg\bigvee I$.

5.5.2 Axiomatization

Our axiom system will rely on the following formula schema: given an ideality sequence $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ and an arbitrary formula ψ,

$$\theta_\psi^{\mathcal{I}}: \quad \bigvee_{\chi \in I}((\chi \wedge \psi) \wedge U(\bigvee I_\chi^+ \to \neg\psi)) \vee (\psi \wedge U(\bigvee I \to \neg\psi))$$

Lemma 10. *Given an ideality sequence $\mathcal{I} = \langle I \cup \{\epsilon\}, \ll \rangle$ and a bare structure (M, s) based on $\mathcal{I}$,*

$$M, s \models \theta_\psi^{\mathcal{I}} \quad \textit{iff} \quad s \in \max_{\leqslant_{\mathcal{I}}} \|\psi\|_M$$

Proof. ($\Rightarrow$) Suppose, to reach a contradiction, that $s \notin \max_{\leqslant_{\mathcal{I}}} \|\psi\|_M$. We split the proof into two cases:

• Case 1: If $s \notin \|\psi\|_M$, then $M, s \not\models \theta_\psi^{\mathcal{I}}$. Contradiction.

• Case 2: If $s \in \|\psi\|_M$ and s is not the best ψ-state, then there exists $t \in M$ such that $M, t \models \psi$ and $t >_{\mathcal{I}} s$. Since there must exist $\chi \in \Phi \cup \{\top\}$ such that $M, s \models \chi \wedge \psi$, we have for any $r > s$, there exists $\chi' \gg \chi$ (if χ is $\top$, we let $\chi' \gg \top$ for each $\chi' \in \Phi$) such that $M, r \models \chi'$. Since $t \in M$, $M, t \models \bigvee I_\chi^+ \to \neg\psi$. By $t >_{\mathcal{I}} s$, we have that $M, t \models \chi'$ for some $\chi' \gg \chi$, which implies that $M, t \models \bigvee I_\chi^+$. So $M, t \models \neg\psi$. Contradiction. Therefore, we proved that $s \in \max_{\leqslant_{\mathcal{I}}} \|\psi\|_M$.

($\Leftarrow$) Suppose that $s \in \max_{\leqslant_{\mathcal{I}}} \|\psi\|_M$. It must be either the case where there exists $\chi \in \Phi$ such that $M, s \models (\chi \wedge \psi) \wedge \neg \bigvee I_\chi$ or the case where $M, s \models \psi \wedge \neg \bigvee I$. We then prove by two cases:

• Case 1: There exists $\chi \in \Phi$ such that $M, s \models (\chi \wedge \psi) \wedge \neg \bigvee I_\chi$:

- Case 1.1: If there is no $t \in M$ such that $t >_{\mathcal{I}} s$, this implies that $\bigvee I_\chi^+ = \bot$. It is trivial that for each $r \in M$, $M, r \models \bigvee \Phi_\chi \to \neg\psi$. So $M, s \models U(\bigvee I_\chi^+ \to \neg\psi)$.

- Case 1.2: If there is $t \in M$ such that $t >_{\mathcal{I}} s$, there must exist $\chi' \gg \chi$ in Φ such that $M, t \models \chi' \wedge \neg\psi$, which implies that $M, t \models \bigvee I_\chi^+ \to \neg\psi$. As for each $r \in M$ such that $r \not>_{\mathcal{I}} s$, we have $M, r \models \neg \bigvee I_\chi^+$ which also implies that $M, r \models \bigvee I_\chi^+ \to \neg\psi$. Thus, for all $u \in M$, it holds that $M, u \models \bigvee I_\chi^+ \to \neg\phi$. Therefore, $M, s \models U(\bigvee I_\chi^+ \to \neg\psi)$.

• Case 2: $M, s \models \psi \wedge \neg \bigvee I$: Let t be an arbitrary state in M:

- Case 2.1: If $M, t \models \neg \bigvee I$, it trivially holds that $M, t \models \bigvee I \to \neg\psi$.

- Case 2.2: If $M, t \models \bigvee I$, this implies that there must exist $\chi \in I$ such that $M, t \models \chi$. By the definition of $\leqslant_{\mathcal{I}}$, we have $s <_{\mathcal{I}} t$. By our supposition that $s \in \max_{\leqslant_{\mathcal{I}}} \|\psi\|_M$, we have $M, t \not\models \psi$. Thus, we obtain that $M, t \models \bigvee I \to \neg\psi$.

Therefore, we can conclude that $M, s \models U(\bigvee I \to \neg\psi)$. And we can also obtain that $M, s \models \theta_\psi^{\mathcal{I}}$. □

So Lemma 10 indicates that $\theta^{\mathcal{G}}_{\psi}$ captures the best ψ-states with respect to $\leqslant_{\mathcal{I}}$. In the light of $\theta^{\mathcal{I}}_{\psi}$, we can give the Kanger-Anderson reduction for the formula $\bigcirc_{\mathcal{I}}(\phi|\chi)$ and therefore the proof system of the logic $\mathbb{PCDL}$.

Proposition 9 (KA-reduction of $\bigcirc_{\mathcal{I}}(\_|\_)$)**.** *Given a bare structure (M, s) and an ideality sequence $\mathcal{I}$,*

$$M, s \models \bigcirc_{\mathcal{I}}(\phi|\psi) \leftrightarrow U(\theta^{\mathcal{I}}_{\psi} \rightarrow \phi)$$

Proof. ($\Rightarrow$) Suppose that $M, s \models \bigcirc_{\mathcal{I}}(\phi|\psi)$. By semantics, $\max_{\leqslant_{\mathcal{I}}} \|\psi\|_M \subseteq \|\phi\|_M$. Assume, to reach a contradiction, that there exist a state $t \in M$ such that $M, t \models \theta^{\mathcal{I}}_{\psi} \wedge \neg\phi$. By Lemma 10, $t \in \max_{\leqslant_{\mathcal{I}}} \|\psi\|_M$. So we have $M, t \models \phi$. Contradiction.

($\Leftarrow$) Suppose that $M, s \models U(\theta^{\mathcal{I}}_{\psi} \rightarrow \phi)$. Let $t \in \max_{\leqslant_{\mathcal{I}}} \|\psi\|_M$. Assume, to reach a contradiction, that $M, t \models \neg\phi$. By Lemma 10, we have $M, t \models \theta^{\mathcal{I}}_{\psi}$. Thus, we have $M, t \models \phi$. Contradiction. □

The KA-reduction above helps to reduce each formula in the form of $\bigcirc_{\mathcal{I}}(\phi|\psi)$ to a $\mathcal{L}_{\text{EL}}$-formula without any dyadic deontic operator. Therefore we can provide the proof system of the logic $\mathbb{PCDL}$

Definition 65. *The proof system $\mathbb{PCDL}$ consists of the following axiom schemas and inference rules:*

(TAUT)	*All instances of tautologies*
(K)	$U(\phi \rightarrow \psi) \rightarrow (U\phi \rightarrow U\psi)$
(T)	$U\phi \rightarrow \phi$
(4)	$U\phi \rightarrow UU\phi$
(5)	$\neg U\phi \rightarrow U\neg U\phi$
(KA)	$\bigcirc_{\mathcal{I}}(\phi|\psi) \leftrightarrow U(\theta^{\mathcal{I}}_{\psi} \rightarrow \phi)$
(MP)	*From ϕ and $\phi \rightarrow \psi$, infer ψ*
(N)	*From ϕ, infer $U\phi$*
(RE)	*From $\phi \leftrightarrow \psi$, infer $\chi \leftrightarrow \chi[\phi/\psi]$*

The soundness of axiom (KA) has been given in Proposition 3. And it is easy to prove that the inference rule (RE) is sound since we can prove that 'from $\vdash \theta \leftrightarrow \chi$, infer $\vdash \bigcirc_{\mathcal{I}}(\theta|\psi) \leftrightarrow \bigcirc_{\mathcal{I}}(\chi|\psi)$ and $\vdash \bigcirc_{\mathcal{I}}(\phi|\theta) \leftrightarrow \bigcirc_{\mathcal{I}}(\phi|\chi)$' is valid. The remaining axioms and inference rules are also sound since they are classical axioms or rules from modal logic $S5$.

By the axiom (KA), we can reduce every formula with the operator $\bigcirc_{\mathcal{I}}(\_|\_)$ to a formula without the operator. Therefore, completeness of the axiom system $\mathbb{PCDL}$ with respect to the semantics can be proved by translating $\mathcal{L}_{\text{PCDL}}$-formulas to $\mathcal{L}_{\text{EL}}$-formulas via reduction axioms and induction on the complexity of the formulas (see Chapter 7.4 in van Ditmarsch et al. (2007) and Theorem 11 in Kooi (2007)).

Theorem 7. *The logic $\mathbb{PCDL}$ is sound and strongly complete with respect to the class of bare structures.*

5.6 Successful updates and Jørgensen's dilemma

In Section 5.4, we showed several ways of making norms which are to update the normative systems (ideality sequences). However, making a norm does not always bring the corresponding obligation to addressees. We call it 'Moorean phenomena' in deontic context. In this section, we first provide some philosophical investigations on the *addressee's* conditional obligations. They *describe* what an obligation addressee ought to 'see to it that' given certain circumstances.

Notations: If $\mathcal{I}_1 = \phi_i;\mathcal{I}_2$, let $\mathcal{I}_1 - \phi_i = \mathcal{I}_2$ and let $\mathcal{I}_2 \lhd \phi_i = \mathcal{I}_1$.

5.6.1 Descriptive obligations: following norms or not

As shown beforehand, the formula $\bigcirc_{\mathcal{I}}(\phi|\psi)$ is a normative proposition which is satisfied or not in a bare structure. To be more precise, its satisfaction is equivalent to the satisfaction of $\bigcirc(\phi|\psi)$ on a betterness structure based on $\mathcal{I}$. Normative propositions sometimes do not perform in line with norms from the normative system which it is based on. In other words, in some cases, an agent does not have an obligation to achieve ϕ even though ϕ has already been a norm in the concerning normative system. This happens since norms exist *outside* the betterness structures. They indeed affect what obligations the addressee would have, however, the addressee's obligations are not only decided by normative systems, but also by these facts embedded in the bare structures.

In order to clarify the discrepancy between norms and obligations, we should first elaborate, in our context, on the well-known Kantian principle that "*ought implies can*". The principle states that if an agent has an obligation, then (s)he must be able to achieve that. Hence it is easy to see that the principle is attributed to obligation addressees. The 'ought' denotes the addressees' obligations which is decided on the betterness structures. We can therefore say that 'ought implies can' should be rephrased more precisely as 'normative propositions imply can'.

How should we interpret the term 'can'? Many deontic logicians have been attempting to introduce the ability or agency into the deontic logics. These contributions can mostly be referred to deontic stit logic as shown in Chapter 1.2. But in our framework, we do not go as far as Stit logic. We propose to interpret the 'can' in a more straightforward sense – possibility. Several similar interpretations can also be found in Hansson (1969)'s original discussion on the dyadic deontic operator and in work by other deontic logicians like Feldman (1986). We say ϕ is possible in a bare structure (M, s) if $M, s \models \hat{U}\phi$. Accordingly, 'ought implies can' can be formulated by following valid formulas.

Fact 8.
(a) $\models \bigcirc_{\mathcal{I}}(\phi|\top) \to \hat{U}\phi$
(b) $\models (\bigcirc_{\mathcal{I}}(\phi|\psi) \wedge \hat{U}\psi) \to \hat{U}\phi$

The formula (a) means that 'it ought to be ϕ unconditionally based on $\mathcal{I}$' implies 'ϕ is possible'; (b) shows that 'it ought to be ϕ given ψ based on $\mathcal{I}$' and 'the condition ψ is possible' imply 'ϕ is possible'. The two validities illustrate that if some state of affairs ought to be achieved, then it must be possible. In other words, they 'can' be done.

5.6.2 Successful updates

Definition 66 (Successful updates)**.** *Let $\mathcal{I}_1 = (I_1 \cup \{\epsilon\}, \ll)$ be an ideality sequence and let $\mathcal{I}_1 = \phi_1; \mathcal{I}_2$.*

$$-\phi_1 \text{ is a successful update on } \mathcal{I}_1 \text{ in } (M,s) \quad \text{iff} \quad M,s \models \neg \bigcirc_{\mathcal{I}_1 - \phi_1} (\phi_1 | \neg \bigvee I_2)$$
$$\lhd\phi_1 \text{ is a successful update on } \mathcal{I}_2 \text{ in } (M,s) \quad \text{iff} \quad M,s \models \bigcirc_{\mathcal{I}_2 \lhd \phi_1} (\phi_1 | \neg \bigvee I^{+}_{1\phi_1})$$

Deleting the worst state of affairs from the ideality sequence $\mathcal{I}_1$ is successful in (M,s) if and only if the formula $\neg \bigcirc_{\mathcal{I}_1 - \phi_1} (\phi_1 | \neg \bigvee I^{+}_{1\phi_1})$ is true in (M,s). The formula means that the agent no longer has the obligation to see to it that ϕ_1 when any better state of affairs is not the case with respect to the updated ideality sequence $\mathcal{I}_1 - \phi_1$. Postfixing a new state of affairs to the ideality sequence $\mathcal{I}_2$ is successful if and only if the agent gets a new obligation to see to it that ϕ_1 when any better state of affairs is not the case with respect to the updated ideality sequence $\mathcal{I}_2 \lhd \phi_1$. Definition 66 is in line with our intuition. A successful command to delete a norm is supposed to release some obligations and a successful command to add a norm is meant to assign some new obligations.

The issues of successful updates are closely related to CUGO principle (an acronym of 'Commands Usually Generate Obligations') put forward by Yamada (2008):

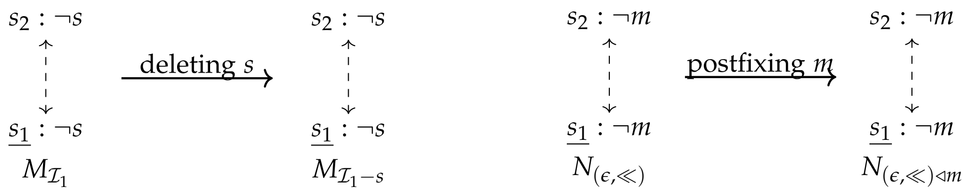
$$[!_{(i,j)}\phi] \bigcirc_{(i,j)} \phi$$

It can be read as "after a command to i given by an authority j to see to it that ϕ, i has an obligation to j to see to it that ϕ". In our framework, we can distinguish between the commands which do generate obligations and those commands which do not. We therefore can rename the CUGO principle as SCGO which means that 'Successful Commands Generate Obligations'.

5.6.3 Possible failure of prescription

A bare structure M provides a model capturing possible worlds, which are supposed to respect logical, ontic, physical or epistemic laws. For example, the contradiction is by no means satisfied in any possible world. Furthermore, a betterness structure based on some ideality sequence $M_{\mathcal{I}}$ shows the betterness ordering between each two possible worlds.

However, giving norms is not constrained by these laws. It is still possible for a law-giver to command you to open the door and close the door simultaneously (maybe when the commander is drunk), which can never be carried out by any obligation addressees. The law-giver might also command the obligation



(a) Deleting s is failed

(b) Postfixing m is failed

Figure 5.13: Deletion Failure and Postfixing Failure

addressee to do something impossible for the addressee, e.g., "Go to the Moon!". Hence, it is easy to see that some updates on normative systems fail to bring about corresponding obligations into betterness structures. We next give two examples where deletion and postfixing are failed, respectively.

Example 1: deletion failure In 1991, an American man requested his son, Tom, not go to the Soviet Union. But during that year, they got the news of the dissolution of the Soviet Union. So the man revoked his request. Since the Soviet Union no longer exists, Tom would never go there.

Let proposition s stand for "Tom is in the Soviet Union". The initial ideality sequence is $\mathcal{I}_1 = \langle \neg s, \epsilon \rangle$, which means that not being in the Soviet Union is the best state of affairs, otherwise is worse. After deleting s, $\mathcal{I}_1$ becomes $(\lambda, \ll)$. The corresponding change on the betterness structure is shown in Figure 5.13a.

Since $\neg s$ is true everywhere, it is impossible for Tom to be in the Soviet Union. The only state of affairs he sees to it that is $\neg s$. So $M_{\mathcal{I}_1}$ is identical to the updated betterness structure $M_{\mathcal{I}_1 - s}$. Therefore we have $M, s_1 \not\models \neg \bigcirc_{\mathcal{I}_1} (\neg s | \top)$ and $M, s_1 \not\models \neg \bigcirc_{\mathcal{I}_1 - s} (\neg s | \top)$. That means that the command of deleting $\neg s$ is failed.

Example 2: postfixing failure Tom's father asks Tom to go to the moon (denoted by m). We assume the initial ideality sequence for Tom is $(\epsilon, \ll)$. After receiving the command from his father, the ideality sequence becomes $(\epsilon, \ll) \lhd m$. The corresponding change of the betterness structure is shown in Figure 5.13b.

After Tom receives the command that m, a new normative system is established: $(\epsilon, \ll) \lhd m$ which will make all m-states in the betterness structure better than those $\neg m$-states. But there is no m-states in N. So $N, s_1 \models \neg \bigcirc_{(\epsilon, \ll) \lhd m} (m | \top)$, which means that commanding m fails in assigning the obligation $\bigcirc (m | \top)$ to Tom.

Conditions for success The two valid formulas in Fact 8 provides clues to conditions for successes of deletion and postfixing respectively.

$$\begin{array}{ll}(1) & \bigcirc_{\phi_1;\mathcal{I}}(\phi_1|\neg\bigvee I)\wedge\hat{U}(\neg\bigvee I\wedge\neg\phi_1)\rightarrow\neg\bigcirc_{\mathcal{I}}(\phi_1|\neg\bigvee I);\\ (2) & \neg\bigcirc_{\mathcal{I}}(\phi_1|\neg\bigvee I)\wedge\hat{U}(\neg\bigvee I\wedge\phi_1)\rightarrow\bigcirc_{\mathcal{I}\triangleleft\phi_1}(\phi_1|\neg\bigvee I);\end{array}$$

We can treat ϕ_1 as the content of the norm updated. Formula (1) shows that if the negation of content ($\neg\phi_1$) is consistent with the condition that all the norms in the updated ideality sequence ($\neg\bigvee I$), then deleting ϕ_1 will be successful. Formula (2) indicates that if the content is consistent with the condition that all the norms in the updated ideality sequence ($\neg\bigvee I$), then postfixing ϕ_1 will be successful.

5.6.4 Resolving Jørgensen's dilemma

We will explore two examples of Jørgensen's dilemma with the concept of successful updates.

(1) Open the door!
(2) The door cannot be opened unless it is unlocked.
―――――――――――――――――――
(3) Unlock the door!

The inference above shows that we can derive a norm or an imperative from an imperative and some declarative information. Sentences (1) and (3) are essentially commands given by some speech acts of a commander, which renders some states of affairs better than others. To comply with the speech-act reading and avoid any ambiguity, it would be better to rephrase the Example above as follows:

Example 1.
(1) Let the door be open!
(2) It is impossible that the door is open but it is not unlocked.
―――――――――――――――――――
(3) Let the door be unlocked!

We subsequently give two ways to interpret the dilemma and they provide different resolutions.

From norms to obligations

In the first approach, sentence (3) is interpreted with a normative proposition $\bigcirc(u|\top)$ which means that John ought to unlock the door unconditionally. Therefore, the inference is essentially not an inference between imperatives. It expresses a process where a new conditional obligation is assigned to John after prescribing a new norm (1), i.e., $M_\epsilon, s \models (\hat{U}o\wedge U(o\rightarrow u))\rightarrow\bigcirc_{(\epsilon,\ll)\triangleleft o}(u|\top)$. The formula can be read as 'if opening the door is possible and it must be the case that if the door is open, then it is unlocked, then after postfixing the norm of opening the door, it is obligatory to be unlocking the door unconditionally'.

Proposition 10. *Let M be an arbitrary bare structure. If $\mathcal{I} = (\epsilon, \ll)$, then $M, s \models (\hat{U}o \wedge U(o \to u)) \to \bigcirc_{(\epsilon,\ll)\lhd o}(u|\top)$*

Proposition 10 explains why we intuitively think that the above inference is 'valid'.

Proposition 11. *Let $\mathcal{I} = \langle \Phi, \ll \rangle$ be an arbitrary ideality sequence and let (M, s) be an arbitrary bare structure. We have $M, s \models (\hat{U}(\psi \wedge \neg \bigvee \Phi) \wedge U(\psi \to \chi)) \to \bigcirc_{(\epsilon,\ll)\lhd\psi}(\chi|\neg \bigvee \Phi)$.*

Proposition 11 is a generalization of Proposition 10. The proof is very straightforward, so we do not elaborate it here.

From norms to norms

In terms of the second approach, we interpret sentence (1) and (3) to norms rather than obligations. Sentence (2) suggests that the content of norm (3) is implied by the content of norm (1). In other words, the set of obligations brought about by (1) includes the obligations brought about by (3). Following this intuition, we give an interpretation on Example 1.

Given an ideality sequence $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ and a bare structure (M, s), let $O_\psi^{\mathcal{I}} = \{\phi \mid M, s \models \bigcirc_{\mathcal{I}}(\phi|\psi)\}$. In the following part, proposition o represents that the door is open and u represents that the door is unlocked.

Assume that, in Example 1, the original ideality sequence is $\mathcal{I} = (\epsilon, \ll)$ and the bare structure is (M, s). Sentence (1) is a speech act postfixing a new norm o to $\mathcal{I}$. It yields a new ideality sequence $\mathcal{I} \lhd o$. Sentence (2) can be interpreted as the formula $\neg\hat{U}(o \wedge \neg u)$ which is logically equivalent to $U(o \to u)$. Sentence (3) is a different speech act which postfixes a new norm u to $\mathcal{I}$ yielding $\mathcal{I} \lhd u$. The set of obligations brought about by (1) is $O_{\neg\bigvee I}^{\mathcal{I}\lhd o} = \{\phi \mid M, s \models \bigcirc_{\mathcal{I}\lhd o}(\phi|\neg \bigvee I)\}$. The set of obligations brought about by (3) is $O_{\neg\bigvee I}^{\mathcal{I}\lhd u} = \{\phi \mid M, s \models \bigcirc_{\mathcal{I}\lhd u}(\phi|\neg \bigvee I)\}$. According to our intuition, it should be the case that $O_{\neg\bigvee I}^{\mathcal{I}\lhd u} \subseteq O_{\neg\bigvee I}^{\mathcal{I}\lhd o}$.

Proposition 12. *Let (M, s) be an arbitrary bare structure and $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ be an arbitrary ideality sequence. If $\lhd\psi$ is a successful update on $\mathcal{I}$ in (M, s) and $M, s \models U(\psi \to \chi)$, then $\lhd\chi$ is also a successful update on $\mathcal{I}$ in (M, s) and $O_{\neg\bigvee I}^{\mathcal{I}\lhd\chi} \subseteq O_{\neg\bigvee I}^{\mathcal{I}\lhd\psi}$.*

Proof. Let ϕ be an arbitrary formula in $O_{\neg\bigvee I}^{\chi;\mathcal{I}}$. So $M, s \models \bigcirc_{\chi;\mathcal{I}}(\phi|\neg \bigvee I)$. We know that $M, s \models \bigcirc_{\psi;\mathcal{I}}(\psi|\neg \bigvee I)$. If $\|\neg \bigvee I\|_M = \varnothing$, we have $M, s \models \bigcirc_{\psi;\mathcal{I}}(\phi|\neg \bigvee I)$ trivially. If $\|\neg \bigvee I\|_M \neq \varnothing$, we have $\|\psi\|_M \cap \|\neg \bigvee I\|_M \neq \varnothing$. By $M, s \models U(\psi \to \chi)$, we have $\|\psi\|_M \subseteq \|\chi\|_M$ and thus $\|\chi\|_M \cap \|\neg \bigvee I\|_M \neq \varnothing$. By the semantics of $M, s \models \bigcirc_{\chi;\mathcal{I}}(\phi|\neg \bigvee I)$, we have $\max_{\leqslant_{\chi;\mathcal{I}}} \|\neg \bigvee \|_M \subseteq \|\phi\|_M$. By the definition of $M_{\chi;\mathcal{I}}$, it holds that $\max_{\leqslant_{\chi;\mathcal{I}}} \|\neg \bigvee I\|_M = \|\neg \bigvee I\|_M \cap \|\chi\|_M$. So $\|\neg \bigvee I\|_M \cap \|\chi\|_M \subset \|\phi\|_M$. By $\|\psi\|_M \subseteq \|\chi\|_M$, we have $\|\psi\|_M \cap \|\neg \bigvee I\|_M \subseteq \|\phi\|_M$. By the definition of $M_{\psi;\mathcal{I}}$, it holds that $\max_{\leqslant_{\psi;\mathcal{I}}} \|\neg \bigvee I\|_M = \|\neg \bigvee I\|_M \cap$

$\|\psi\|_M$. Thus, we have $\max_{\leqslant_{\psi;\mathcal{I}}} \|\neg \bigvee I\|_M \subseteq \|\phi\|_M$. This is equivalent to $M, s \models \bigcirc_{\psi;\mathcal{I}}(\phi|\neg \bigvee I)$. □

Proposition 12 indicates that the success of (1) and information (2) implies that (3) is successful and the obligations triggered by (3) are also triggered by (1).

Example 2.

(1) There is no longer need to let the door be open!*
(2) The door cannot be opened unless it is unlocked.
———
(3) There is no longer need to let the door be unlocked!*

Example 2 involves different updates on ideality sequences. Sentence (1*) deletes a norm o from the ideality sequence $o;\mathcal{I}$. Sentence (3*) deletes u from the ideality sequence $u;\mathcal{I}$. We consider it as a valid inference since we intuitively think deleting o defeats more obligations than deleting u.

Proposition 13. *Let (M, s) be an arbitrary bare structure and $\mathcal{I} = (I \cup \{\epsilon\}, \ll)$ be an arbitrary ideality sequence. If $-\psi$ is a successful updates on $\psi;\mathcal{I}$ in (M, s) and $M, s \models U(\psi \to \chi)$, then $-\chi$ is also a successful updates on $\chi;\mathcal{I}$ in (M, s) and $(O^{\chi;\mathcal{I}}_{\neg \bigvee I} - O^{(\chi;\mathcal{I})-\chi}_{\neg \bigvee I}) \subseteq (O^{\psi;\mathcal{I}}_{\neg \bigvee I} - O^{(\psi;\mathcal{I})-\psi}_{\neg \bigvee I})$.*

Proposition 13 indicates that the success of (1*) and information (2) implies that (3*) is successful and the obligations defeated by (3*) are also defeated by (1*). It worth noting that $\chi;\mathcal{I} - \chi = \psi;\mathcal{I} - \psi = \mathcal{I}$. So the inclusion in the end of Proposition 13 is equivalent to $(O^{\chi;\mathcal{I}}_{\neg \bigvee I} - O^{\mathcal{I}}_{\neg \bigvee I}) \subseteq (O^{\psi;\mathcal{I}}_{\neg \bigvee I} - O^{\mathcal{I}}_{\neg \bigvee I})$. This is equivalent to $O^{\chi;\mathcal{I}}_{\neg \bigvee I} \subseteq O^{\psi;\mathcal{I}}_{\neg \bigvee I}$ which is same as the inclusion in the end of Proposition 12.

5.7 Discussion and conclusion

The research in this chapter was established on the distinction between obligations for describing agent's deontic states and normative systems for prescribing norms that assess what states of affairs are better or worse. There have been many deontic logics concerning this prominent topic. Let us discuss some related work to our research.

Dynamifying input/output logic framework The relation between description and prescription can be represented by a widely accepted framework provided in input/output logic by Makinson and Van Der Torre (2000), Makinson and van der Torre (2007a). We will show that our logic $\mathbb{PCDL}$ practically dynamifies the framework of input/output logic. Let us first see Figure 5.14 showing the framework of input/output logic:

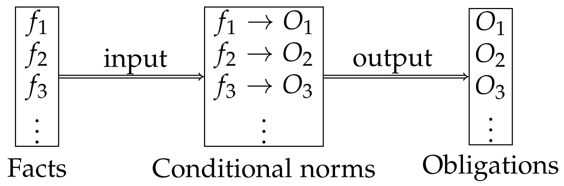


Figure 5.14: The framework of input/output logic

The core element in input/output logic is the set of conditional norms. The conditions are facts and consequences are obligations. So, given a set of facts, these norms can produce a set of obligations based on their own mechanism.

The basic framework used in this thesis can also be simplified to the pattern of input/output logic. A bare structure in this chapter provides a factual basis for describing a possible situation. It plays a similar role as 'facts' in input/output logic. An ideality sequence, as a normative system, is a criterion for assessing which states in the bare structure are better and therefore, it decides what obligations the agent has. So it can be thought of as the 'mechanism', i.e., conditional norms in input/output logic. Finally, the obligations over the betterness structure are the output of the bare structure and the ideality sequence.

This chapter developed a pattern similar to input/output logic since we can change the 'mechanism' and we can also update the obligations accordingly. In other words, we dynamified the static framework of input/output logic.

Updates on independent normative systems There has been a good amount of research about updates on obligations in deontic logic. As mentioned in Section 4.2, priority structures or ideality sequences studied in this thesis serve a similar purpose to norms given by van der Torre and Tan (1998). They also investigated updates on deontic models (deontic states in their terminology). However, van der Torre and Tan (1998) defined updates directly on the models where the betterness relations can be changed by new norms. Similarly, Mastop (2011) also established a dynamic deontic logic where norm systems are substantially deontic models. These norm systems can be updated by different types of formulas, including deontic formulas. An update by a norm is a restriction on the original deontic relation, which leads to a new deontic relation complying with the norm. Also reviewing the series work on speech acts affecting obligations given by Yamada (see Yamada (2006, 2008, 2011)), we found that all these updates on obligations are defined on the deontic relations of the models.

In contrast, this chapter defines updates on the normative system rather than models (bare structures in this chapter). The normative systems (ideality sequences) and bare structures are separated based on the distinction between prescription and description. Any change on the normative system would bring about changes on the truth values of obligations. In other words, the updates studied in this chapter change obligations indirectly.

Comparisons with the updates on legal systems In this chapter, we investigated several types of updates on normative systems. Similarly, Governatori and Rotolo and their colleagues have studied updates on legal systems by defeasible logic. A defeasible theory is a structure $(F, R, \succ)$ where F is a set of facts, R is a set of rules which represent relationships between premises and conclusions, and $\succ$ is an acyclic superiority relation on R. Governatori et al. (2007) studied norm modifications in defeasible logic. They mentioned four types of modifications: substitution, derogation, annulment and abrogation. These modifications change the set of rules in a defeasible structure. Substitution can replace some rules with other rules. Derogation limits the effects of the derogated rule. It is related to postfixing since both updates characterize what should follow if there are exceptions. Annulment and abrogation are used for cancelling some rules. Annulment requires removing all effects/consequences of the provision that has been annulled. But abrogation does not cancel the effects/consequences that were obtained before the modification. In this sense, the deletion introduced in this chapter is more similar to annulment since deleting a norm successfully directly cancels the corresponding obligations. Governatori and Rotolo (2010) also studied updates resulting from adding new norms into the legal systems. It plays a similar role as postfixing or insertion defined in this chapter.

Conclusion Considering the differences between the prescriptive reading and the descriptive reading in normative sentences, we introduced the notion of relativized conditional obligations based on ideality sequences. Each ideality sequence can be treated as a normative system which provides a criterion on the relative ideality of states of affairs. Once that is done, every betterness structure based on a given ideality sequence describes the conditional obligations. Deletion and postfixing are two updates on the normative system which can bring about their corresponding obligations successfully or not. Jørgensen's dilemma can be conceptualized by using the notion of successful updates. Furthermore, a sound and strongly complete axiom system for the logic of relativized conditional obligations $\mathbb{PCDL}$ has been established in this chapter.

Chapter 6

What I Ought to Do Based on What I Know

6.1 Introduction

As mentioned in Section 1.1, Von Wright (1951) is generally regarded as the person who initiated the field of deontic logic. His first system did interpret the objects of obligations as act types instead of propositions, which indicated that he meant to characterize obligation with respect to actions. However, he applied classical logical connectives, such as negation and material implication, to act types, which causes problems on how to understand these compound act types. For example, the formula $\neg P\alpha \to \bigodot(\alpha \to \beta)$ is valid in his system, which means that if an action α is forbidden, then doing action α commits us to do anything (β is an arbitrary action). This is counter-intuitive and how to interpret the implication between act types is questionable. Thereupon Von Wright himself accepted the proposal that deontic operators are used to be applied to propositions and he rebuilt his standard deontic logic following Prior (1955), Anderson and Moore (1957), and Kanger (1970)'s approaches. We call this type of obligations 'ought-to-be' obligations. As a consequence, from the 1960s onward, deontic logicians could merely formalize which states of affairs are good rather than expressing which actions are good. 'Ought-to-be' obligations are normally interpreted as 'it ought to be the case that $\cdots$' or 'it is obligatory that $\cdots$'.

However, obligations in natural language usually appear with respect to verbs, or we can say, to actions. We generally make claims such as 'students ought to go to school on time', 'you ought to keep promises' and 'I ought to open the window now since my mother asks me to do so'. Only actions can be performed and hence, our obligations are with respect to actions that we should perform.

There have been several deontic logics established based on actions after Von Wright's first attempt. Meyer (1988) treated deontic logic as a variant of dynamic logic where actions are formalized as binary relations between states to represent transitions between situations. An action is forbidden if all the states that can be reached by performing the action are bad. A similar characterization can also be found in Segerberg (1982). Meyer's approach was extended by van der Meyden (1996), where actions are assigned to sequences of states (multiple transitions between states). Broersen (2004) investigated the problem of 'action negation' in dynamic deontic logic. Kulicki and Trypuz (2017) introduced both actions and states into their deontic logic where an action ought to be done if it satisfies both *a*-norms (for actions) and *s*-norms (for states). These approaches have been developed based on the idea that actions are represented as transitions between states and an action which leads to the optimal consequences ought to be done.

Furthermore, as Chapter 3 and 4 showed, what an agent ought to do at some moment also depends on his/her knowledge. So this chapter will treat agents' obligations as actions that they ought to perform based on their knowledge.

Dynamic epistemic logic seems to be an alternative framework for modelling actions based on knowledge since action models are capable of expressing how actions update situations according to agents' information. Following the idea of Meyer (1988), an action ought to be done if and only if the action leads from the initial situation to the most ideal situations. In the context of dynamic epistemic logic, it seems that we should suggest analogously that an action model ought to be done if and only if the action model updates the initial epistemic model to the optimal epistemic models. However, this chapter will replace the 'optimality' with 'improvement' to define 'ought-to-do' obligations. Roughly speaking, an action ought to be done if and only if the action always improves the initial situation to *better* situations. We will explain why this approach makes sense in Section 6.3.

The following scenarios motivate us to treat obligations as actions and suggest why we should involve epistemic elements into obligations. Scenario 12 has been given by Horty (2019) and Scenarios 10, 11 and 13 are variants of the scenarios studied by Pacuit et al. (2006).

Scenario 10. Ann is a nurse and Uma is a doctor. Sam is a patient at Ann and Uma's practice. Uma just checks the test sheet and diagnoses Sam with diabetes. Ann is taking care of Sam in the sickroom but she did not yet get the information that Sam has diabetes. So Ann does not have an obligation to inject insulin for Sam at that time.

Scenario 11. Ann is a nurse and Uma is a doctor. Sam is a patient at Ann and Uma's practice. Uma diagnoses Sam with diabetes and tells Ann the fact. Now Ann has an obligation to inject insulin for Sam.

Scenario 12. Tao places a coin on the table but Chiyo cannot see whether it is heads up or tails up. Chiyo must risk five euros for the opportunity to bet on heads or tails, with ten euros to win if Chiyo bets correctly (if the coin lands heads up and Chiyo bets on heads, or if the coin lands tails up and Chiyo bets on tails) and cannot get the five euros back if she bets incorrectly; or Chiyo can choose not to gamble, without any profit or loss.

Scenario 13. Uma is a doctor working in a hospital. Sam is a patient at Uma's practice and he lies in a sickroom of the hospital. Suddenly, Sam is having a heart attack in the hospital. Uma has an obligation to know the health state of the patients at her practice.

Outline of the chapter In this chapter, we extend the action models introduced in Section 2.5 to an action algebra in order to discuss compound actions, e.g., 'to open the door or to open the window' and 'to open the door and then to open the window'. Then a method on comparing states and different epistemic models is provided based on priority structures. In Section 6.4, we establish the logic of knowledge-based 'ought-to-do' obligations and the axiomatization is proved to be sound and strongly complete. Section 6.5 reformulates several famous paradoxes in deontic logic using the newly proposed framework and resolves them. In the end, we show how to model the scenarios mentioned above by our framework.

6.2 Compound actions and transitions between models

Propositional dynamic logic captures actions as binary relations between states. Each action represents a transition from one state to another. On page 165-166 in Harel et al. (2000), propositional dynamic logic defines four main operators on actions: sequential composition, non-deterministic choice, iteration and test. A sequential composition of two actions represents doing the first action and then doing the second. A non-deterministic choice between two actions means choosing one of the two actions non-deterministically and doing it. An iteration of one action means doing the action a non-deterministically chosen finite number of times. Testing a formula means proceeding if the formula is true; fail if it is false.

Analogous to propositional dynamic logic, dynamic epistemic logic also treats actions as transitions between two situations. Section 2.5 shows the way an action model transforms (updates) an epistemic model to a new epistemic model. Each action model can be regarded as a counterpart of a relation between two states in propositional dynamic logic since both of them represent a single action.

Section 2.5 only showed pointed action models in dynamic epistemic logic. As a matter of fact, in some more extended versions of the logic, sequential composition and non-deterministic choice have been introduced for capturing more complex actions. This chapter will also focus on these two types of action operators and they can update epistemic models in different ways. They make our framework more expressive in describing different types of actions and accommodate the paradoxes in deontic logic.

In the remaining part of this section, several transition types between epistemic models caused by different types of actions are defined. If α is an action, the term $[\![\alpha]\!]$ represents the transition relation between epistemic models. For example, if (M, s) and (M', s') are two epistemic models, the notation $(M, s)[\![\alpha]\!](M', s')$ represents that (M, s) is updated by the action α and (M', s') is one of the updated models. Before we introduce the notion of compound actions, the transition by a single pointed action model is defined as follows:

Definition 67 (Transition by pointed action models)**.** *Let (U, e) be an arbitrary pointed action model.*

$$(M, s)[\![U, e]\!](M', s') \quad \textit{iff} \quad M, s \models pre(e) \textit{ and } (M', s') = (M \otimes U, (s, e))$$

The model $(M \otimes U, (s, e))$ is the update model generated by executing (U, e) on (M, s) (see Definition 19). A transition caused by a pointed action model from an initial pointed epistemic model leads to the updated epistemic model. In the following, we introduce two operators on actions: non-deterministic choice and sequential composition.

6.2.1 Non-deterministic choice

The first type of compound action is *non-deterministic choice*. 'Do A or B' is the general form of it. In terms of permission, 'I am allowed to do sports or play the piano' means that I am permitted to do sports *and* I am also permitted to play the piano, which is also known as a 'free-choice' sentence in deontic logic.

In the context of obligation, the sentence 'you ought to call the ambulance or save her by yourself' implies that you ought to perform at least one of the two actions. Moreover, it also indicates that both actions are obligatory, although it might be the case that you can only do one of them at that moment.

According to our intuition in the context of permission and obligation, non-deterministic choice between two actions seems to suggest the meaning of conjunction, instead of disjunction. This interesting linguistic phenomenon is the root of several paradoxes, like Ross's paradox and the paradox of derived obligation (to be introduced in Section 6.5) in many proposition-based deontic logics. These issues motivate us to take non-deterministic choice into account. We use the standard notation from dynamic epistemic logic (see page 112-113 in van Ditmarsch et al. (2007)):

If α_1 and α_2 are actions, $(\alpha_1 \cup \alpha_2)$ is a non-deterministic choice action.

It deserves noting that α_1 and α_2 may themselves be pointed action models or compound actions. We will define all actions inductively in Definition 71.

The transitions between epistemic models caused by non-deterministic choices are defined as follows:

Definition 68 (Transition of non-deterministic choice)**.** *Let α_1 and α_2 be two arbitrary actions.*

$$(M,s)[\![\alpha_1 \cup \alpha_2]\!](M',s') \quad \textit{iff} \quad (M,s)[\![\alpha_1]\!](M',s') \textit{ or } (M,s)[\![\alpha_2]\!](M',s');$$

Definition 68 implies that the number of updated epistemic models that are generated from one epistemic model by a non-deterministic choice could be multiple. In contrast, the updated model generated by a pointed action model is unique.

6.2.2 Sequential composition

Two actions can also be sequentially composed together forming a new action. Specifically, we can concatenate two actions performed in sequence as a whole. For example, 'put the letter into an envelop and then mail it' is a sequential composition of 'put the letter into an envelop' and 'mail the letter'. In the deontic context, 'you ought to put the letter into an envelop and then mail it' implies that the two actions should be done in sequence and if only the first is finished, it would bring about a violation. We also use the standard notation for sequential composition as used in dynamic epistemic logic (see page 150 in van Ditmarsch et al. (2007)):

If α_1 and α_2 are actions, $\alpha_1;\alpha_2$ denotes the composition of α_1 and α_2.

The transitions between epistemic models caused by sequential compositions are defined as follows:

Definition 69 (Transition by sequential composition)**.** *Let α_1 and α_2 be two arbitrary actions.*

$$(M,s)[\![\alpha_1;\alpha_2]\!](M',s') \quad \textit{iff} \quad \textit{there exists an epistemic model } (M'',s'') \textit{ such that } (M,s)[\![\alpha_1]\!](M'',s'') \textit{ and } (M'',s'')[\![\alpha_2]\!](M',s');$$

Definition 69 means that performing action α_1 leads (M,s) to (M'',s''), and then performing action α_2 leads to (M',s'). When two actions are pointed action models, the sequential composition of them is still a pointed action model. This new action model is constructed in the following way:

Definition 70 (Sequential composition of pointed action models (Def 2.8, van Ditmarsch and Kooi (2008)))**.** *Let $(U,e) = (E,R_1,\cdots,R_n,pre,post)$ and $(U',e') = (E',R'_1,\cdots,R'_n,pre',post')$ be two pointed action models. Then their sequential composition $((U,e);(U',e'))$ is the action model $((U \otimes U'),(e,e')) = ((E'',R''_1,\cdots,R''_n,pre'',post''),(e,e'))$ where*

- $E'' = E \times E'$;
- *for each* $i \in G$, $(e_1, e'_1)R''_i(e_2, e'_2)$ *iff* $(e_1, e_2) \in R_i$ *and* $(e'_1, e'_2) \in R'_i$;
- *for each* $(e_1, e'_1) \in E''$, $pre''(e_1, e'_1) = \langle U, e_1 \rangle pre'(e'_1)$;
- $dom(post''(e_1, e'_1)) = dom(post(e)) \cup dom(post'(e'))$ *and* *if* $p \in dom(post''(e_1, e'_1))$, *then*

  $$post''(e_1, e'_1)(p) = \begin{cases} post(e)(p) & \text{if } p \notin dom(post'(e')), \\ [U, e]post'(e')(p) & \text{otherwise} \end{cases}$$

  For $p \in \mathbf{P}$ *and each* $(e_1, e'_1) \in E''$, $post''(p)(e_1, e'_1) = \langle U, e_1 \rangle post'(p)(e'_1)$.

Definition 70 shows that two pointed action models can be composed to one action model. So $((E'', R''_1, \cdots, R''_n, pre'', post''), (e, e'))$ itself is still a pointed action model. Apart from composing two pointed action models, two compound actions can also be composed. The following part in this section will show that any compound action consisting of non-deterministic choice and sequential composition equals an action in a normal form.

6.2.3 Normal form of compound actions

The following propositions have already been proved in van Ditmarsch et al. (2007) and they tell us how to reduce a sequential composition to non-deterministic choice between sequential compositions.

Proposition 14 (Sequential composition of compound actions (Prop 6.10, van Ditmarsch et al. (2007))). *Let* α_1, α_2 *and* α_3 *be arbitrary actions.*

(1) *The compound action* $((\alpha_1 \cup \alpha_2); \alpha_3)$ *equals* $((\alpha_1; \alpha_3) \cup (\alpha_2; \alpha_3))$.
(2) *The compound action* $(\alpha_1; (\alpha_2 \cup \alpha_3))$ *equals* $((\alpha_1; \alpha_2) \cup (\alpha_1; \alpha_3))$.

In item (1) in Proposition 14, the word 'equals' means that, given an arbitrary epistemic model (M, s), the set of those updated epistemic models that generated from (M, s) by performing $((\alpha_1 \cup \alpha_2); \alpha_3)$ is the same as the set of those updated epistemic models that derived from (M, s) by performing $((\alpha_1; \alpha_3) \cup (\alpha_2; \alpha_3))$. Similarly, item (2) represents that the set of those updated epistemic models that derived from (M, s) by $(\alpha_1 \cup (\alpha_2; \alpha_3))$ is the same as the set of those updated epistemic models that derived from (M, s) by $((\alpha_1; \alpha_2) \cup (\alpha_1; \alpha_3))$. So Proposition 14 can be reformulated as follows:

(1*) $\{(M', s') \mid (M, s)[\![(\alpha_1 \cup \alpha_2); \alpha_3]\!](M', s')\} = \{(M', s') \mid (M, s)[\![(\alpha_1; \alpha_3) \cup (\alpha_2; \alpha_3)]\!](M', s')\}$

(2*) $\{(M', s') \mid (M, s)[\![\alpha_1; (\alpha_2 \cup \alpha_3)]\!](M', s')\} = \{(M', s') \mid (M, s)[\![(\alpha_1; \alpha_2) \cup (\alpha_1; \alpha_3)]\!](M', s')\}$

Since we only consider these two types of compound actions, i.e., non-deterministic choice and sequential composition, it follows immediately that each compound action equals a non-deterministic choice between several pointed action models or sequential compositions according to Proposition 14.

Proposition 15. *Each compound action α equals an action in the form of $((U_1, e_1) \cup \cdots \cup (U_n, e_n))$ for some $n \geq 1$ where $(U_1, e_1), \cdots, (U_n, e_n)$ are pointed action models or sequential compositions of pointed action models.*

Proposition 15 follows immediately from Proposition 14 and Definition 70. Since each sequential composition of two pointed action models is still a pointed action model, we can say each (U_m, e_m) from $(U_1, e_1), \cdots, (U_n, e_n)$ is a pointed action model. We call $((U_1, e_1) \cup \cdots \cup (U_n, e_n))$ the normal form of compound actions.

Therefore, for simplicity, our language for actions only expresses pointed action models and non-deterministic choice.

Definition 71. *The language $\mathcal{L}^{\mathrm{Act}}_{\mathrm{AKDL}}$ is defined as follows:*

$$\alpha \quad ::= \quad (U, e) \mid (\alpha \cup \alpha)$$

where (U, e) are pointed action models with finite domains.

6.3 Comparing epistemic models

In this section, we will define knowledge-based 'ought-to-do' obligations over the framework of dynamic epistemic logic. An epistemic model can be thought of as a set of possible worlds for agents. An action is an update on some epistemic model which can lead to new epistemic models. Meyer (1988)'s idea is that an action ought to be done if and only if it updates the initial state to the optimal states. However, this chapter follows a different approach. We suggest that, in the context of dynamic epistemic logic, an action ought to be done if and only if it always improves the initial model to better models.

This alternative interpretation deserves a further discussion here. Meyer requires an obligation to make the current state achieve *the best* states. Reviewing all deontic logics introduced in previous chapters, we found that optimality is necessary for defining obligations. In standard deontic logic, being obligatory requires all *ideal* states satisfies something. In Hansson's dyadic deontic logic, a conditional obligation is a proposition that holds in all the *best* states under some condition. In deontic stit logic, an agent ought to see to it that ϕ if and only if the *optimal* available action for the agent guarantees ϕ. However, in our settings, the notion of 'better' takes the place of 'best' to determine whether an action ought to be done. This idea makes senses as well. Let us recall the notion of safe knowledge-based obligation defined in Section 4.4. Safe knowledge-based obligation means that no matter what information the agent knows, the obligation always holds. In other words, fulfilling a safe obligation can never be wrong. In this Chapter, performing an action which always leads to better situations in the current state is safe in the sense that it never makes the situation worse. So we also suggest that the obligation to be defined in this chapter is a kind of safe obligation with respect to the agents' current situation. In addition, the betterness structures introduced in Section 2.1 originate from preference logic. In preference logic, the characterizations of 'an agent i prefers a proposition ϕ to another proposition ψ' are various. It can be given based on optimality, like 'for each ψ-state, the best states satisfy ϕ'. And it can also be defined based on 'better': for each ψ-state, there exists at least one better state satisfying ϕ (see van Benthem et al. (2006)). Looking back at our idea of knowledge-based 'ought-to-do' obligation, we can say that an action ought to be done if it always leads to better situations. Maybe the consequences are not the optimal ones. But the action is safe enough to bring about better places.

In the current step, we need to find a method for assessing whether the updated epistemic models become better.

6.3.1 Betterness between states and models

The first notion to be defined is the betterness between two states which can be from different epistemic models. It takes a given priority structure as the criterion to compare these two states.

Definition 72. *(Betterness between states) Let M and M' be two arbitrary epistemic models. Given an $\mathcal{L}_{\mathrm{EL}}$-priority structure $\mathcal{G} = \langle \Phi, \prec \rangle$, the betterness relation between states $\leqslant_{\mathcal{G}}$ is defined as follows: for any two states $s \in M$, $s' \in M'$,*

$$s \leqslant_{\mathcal{G}} s' \iff \forall \phi \in \Phi : M, s \models \phi \Rightarrow M', s' \models \phi$$

If $s \leqslant_{\mathcal{G}} s'$ but $s' \nleqslant_{\mathcal{G}} s$, we have $s <_{\mathcal{G}} s'$.

We have introduced the idea of $\leqslant_{\mathcal{G}}$ in previous chapters. The idea stems from van Benthem et al. (2014). But all the betterness relations shown before are only defined over the states of *one* model. In the current chapter, we release the restriction and render states from different models comparable based on a certain priority structure.

Accordingly, we can evaluate whether an action improves an epistemic model relativized to some agent i by the *sure-thing* principle. In decision theory, the sure-thing principle states that a decision maker who would take a certain action if he knew that event E has occurred, and also if he knew that the negation of E has occurred, should also take that same action if he knows nothing about E (see Page 21, Savage (1972)). This idea also comes from Horty (2019) when he compares two act types.

Let $G = \{1, \cdots, n\}$ be the set of agents. Let (M, s) be a pointed epistemic model where $M = \langle W, \sim_1, \cdots, \sim_n, \leqslant, V \rangle$. Let (U, e) be a pointed action model.

The notations below are going to be used in the remaining part of this chapter: let $i \in G$,

- If t is a state of M, then $[t]_M^{\sim_i} = \{r \in M \mid t \sim_i r\}$.
- $(M,s)|_{\sim_i}$ represents the epistemic model M restricted to the subset $[s]^{\sim_i}$.

In the light of the sure-thing principle, if for each state in $(M,s)|_{\sim_i}$, its updated states in the updated model by action α are not worse than it, and there is at least one state from $(M,s)|_{\sim_i}$ which is updated to a strictly better state, then we can say that the action α practically improves the initial epistemic model relativized to the agent i. Then the following definition shows how to compare two models based on Definition 72.

Definition 73 (Betterness between Models). *Let $\mathcal{G}$ be an $\mathcal{L}_{\mathrm{EL}}$-priority structure. Given an arbitrary epistemic model $M = \langle S, \sim_1, \sim_2, \cdots, \sim_n, V\rangle$ and an arbitrary action α which equals $((U_1, e_1) \cup \cdots \cup (U_n, e_n))$. For each $i \in G$, if $(M,s)[\![\alpha]\!](M',s')$,*

$(M,s) <^i_{\mathcal{G}} (M',s')$	*iff*	*there is $(U,e) \in \{(U_1,e_1), \cdots, (U_n,e_n)\}$ such that $(M,s)[\![U,e]\!](M',s')$ and $(M,s) <^i_{\mathcal{G}} (M \otimes U, (s,e))$;*
$(M',s') <^i_{\mathcal{G}} (M,s)$	*iff*	*there is $(U,e) \in \{(U_1,e_1), \cdots, (U_n,e_n)\}$ such that $(M,s)[\![U,e]\!](M',s')$ and $(M \otimes U, (s,e)) <^i_{\mathcal{G}} (M,s)$;*
where		
$(M,s) <^i_{\mathcal{G}} (M \otimes U, (s,e))$	*iff*	*(1) $M,s \models pre(e)$;* *(2) for each $t \in [s]_M^{\sim_i}$, if there exists $e' \in U$ such that $(t,e') \in [(s,e)]_{M\otimes U}^{\sim_i'}$, then $t \leqslant_{\mathcal{G}} (t,e')$;* *(3) there exists $t \in [s]_M^{\sim_i}$ and there exists $e' \in U$ such that $(t,e') \in [(s,e)]_{M\otimes U}^{\sim_i'}$ and $t <_{\mathcal{G}} (t,e')$.*
$(M \otimes U, (s,e)) <^i_{\mathcal{G}} (M,s)$	*iff*	*(1) $M,s \models pre(e)$;* *(2) for each $t \in [s]_M^{\sim_i}$, if there exists $e' \in U$ such that $(t,e') \in [(s,e)]_{M\otimes U}^{\sim_i'}$, then $(t,e') \leqslant_{\mathcal{G}} t$;* *(3) there exists $t \in [s]_M^{\sim_i}$ and there exists $e' \in U$ such that $(t,e') \in [(s,e)]_{M\otimes U}^{\sim_i'}$ and $(t,e') <_{\mathcal{G}} t$.*

Definition 73 shows how to compare a pointed epistemic model with its updated epistemic model relativized to the agent i's epistemic states. In terms of

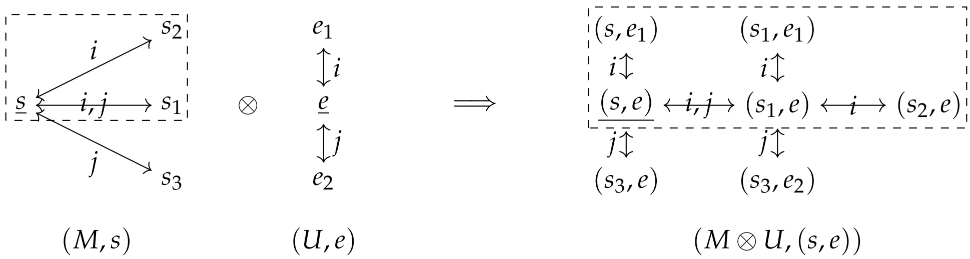


Figure 6.1: An example of updating an epistemic model

$(M, s) \leqslant^i_{\mathcal{G}} (M \otimes U, (s, e))$, we explain the three conditions one by one. Condition (1) requires that the atomic action model is executable on the pointed epistemic model. Condition (2) means that for each state that i cannot distinguish from the current state, if its updated counterpart still cannot be distinguished by i, then the updated state is not worse than its original state. Condition (3) illustrates that there exists at least one original state that i cannot distinguish from the current state such that its updated counterpart is still not be distinguished from the factual state after the update by i and it is strictly worse than its updated counterpart.

The intuition behind Definition 73 is straightforward. Assume s to be the factual world. $(M, s)|_{\sim_i}$ denotes the set of epistemically indistinguishable worlds from s of i. After performing some action model (U, e), the model $(M, s)|_{\sim_i}$ yields an updated model. But we only need to consider these updated worlds that i cannot distinguish from (s, e) since (s, e) is the factual state after the action. When comparing $(M, s)|_{\sim_i}$ and $(M \otimes U, (s, e))|_{\sim'_i}$, all updated worlds should not be worse than their original ones and at least one of them is strictly better than its original counterpart. In terms of $(M \otimes U, (s, e)) <^i_{\mathcal{G}} (M, s)$, it can be understood in a similar way. It represents that update epistemic model is worse than the original one. This intuition is also in line with the well-known concept of *Pareto dominance* in game theory. Given two strategy profiles s and s', the profile s Pareto dominates the profile s' if in the profile s, some agents are better off without making any other agent worse off (see Definition 2.1.1 from Leyton-Brown and Shoham (2008)). In our context, the updated model is better than the initial one if some states become strictly better and no state gets worse.

Figure 6.1 shows which parts of the two models we need to compare when we focus on the agent i (the reflexive and transitive relations are omitted). In the initial epistemic model M shown in Figure 6.1, s is the factual world and we only need to focus on the subset $\{s, s_1, s_2\} \subseteq M$ since i cannot distinguish them. After executing the action model (U, e), we get the model $(M \otimes U, (s, e))$ where we only need to check these states that are derived from $\{s, s_1, s_2\}$ and are connected to the updated factual world (s, e). In other words, we need to compare the set $\{s, s_1, s_2\}$ with the set $\{(s, e), (s, e_1), (s_1, e), (s_1, e_1), (s_2, e)\}$.

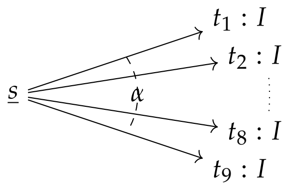


Figure 6.2: α is obligatory on s in Meyer's logic

6.3.2 Building on Meyer's approach

Meyer (1988) proposed a new setting of deontic logic which is viewed as a variant of dynamic logic. As mentioned in Section 6.1, Meyer treats obligations as actions which are transitions between point-style states. His logic is established based on propositional dynamic logic (PDL) models, eg. $(S, R_{\alpha_1}, \cdots, R_{\alpha_n}, V)$ where $\{\alpha_1, \cdots, \alpha_n\}$ is a set of atomic actions.

Let α be an action and let I be a proposition constant which denotes 'ideality'. A formal definition of obligation in Meyer's style can be defined accordingly. The semantics of formula $O\alpha$ is as follows:

$$M, s \models O\alpha \quad \text{iff} \quad \text{for all } t \text{ such that } sR_\alpha t,\ M, t \models I.$$

The semantics means that the action α ought to be done if and only if after performing the actions, all consequence states are ideal. In other words, performing an obligatory action always brings about ideal states. Figure 6.2 gives an example of an action α that is obligatory in Meyer's style.

However, in our real life, it is very unlikely that we can know all available actions when facing some situations and therefore cannot judge which action can bring about the optimal situations. Generally, we perform an action only because it improves the current situation. Following this idea and our notion of comparison between epistemic models, we give our idea about obligation: an action is obligatory if and only if performing the action always leads to *better* situations. Accordingly, the definitions of obligation and prohibition in our framework would be like:

$$M, s \models \odot_i \alpha \quad \text{iff} \quad \text{for all } (M', s') \text{ such that } (M, s)[\![\alpha]\!](M', s'),\ (M, s) <^i_{\mathcal{G}} (M', s')$$
$$M, s \models F_i\alpha \quad \text{iff} \quad \text{for all } (M', s') \text{ such that } (M, s)[\![\alpha]\!](M', s'),\ (M', s') <^i_{\mathcal{G}} (M, s)$$

Figure 6.3 gives an example of the action α, which is obligatory according to our definition. The similarities and differences between Meyer's and our approach are listed below.

Similarities

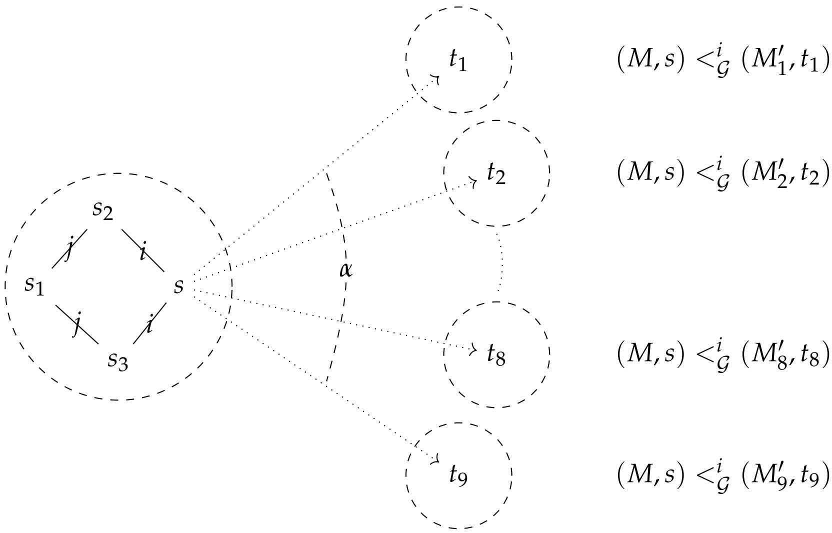


Figure 6.3: α is obligatory on s in our approach

- Actions are represented as transitions between states or situations.

- Obligations are interpreted with respect to actions.

Differences

- Meyer's approach takes PDL models as the basis. Each state/situation is a point in a PDL-model. Our approach uses the DEL framework where each state/situation is an epistemic model. Therefore, a state in our approach encodes more information, especially agents' knowledge.

- In PDL, the actions are transitions between states in *one* PDL-model. However, in DEL, actions are structured as action models and they are transitions between epistemic models. DEL characterizes how an action changes the situation in both epistemic and factual aspects.

- Our approach leaves out the notion of violation or ideality to specify which state is the worst or best. According to our definition, an obligation is an action that always leads to better situations, instead of the most ideal situations.

6.4 A Logic of knowledge-based ‘ought-to-do’ obligation

In this section, we will give a formal characterization of knowledge-based ‘ought-to-do’ obligation. A sound and complete axiom system for the logic will be presented and investigated.

6.4.1 Language and semantics

Let **P** be a countable set of propositional variables and let G be a finite set of agents.

Definition 74 (Language $\mathcal{L}_{\text{AKDL}}$[1]). *The language $\mathcal{L}_{\text{AKDL}}$ is given by BNF:*

$$\mathcal{L}_{\text{AKDL}} \quad \phi \quad ::= \quad p \mid \neg\phi \mid (\phi \wedge \phi) \mid K_i\phi \mid [\alpha]\phi \mid \odot_i \alpha \mid F_i\alpha$$

$$\mathcal{L}^{\text{Act}}_{\text{AKDL}} \quad \alpha \quad ::= \quad (U, e) \mid (\alpha \cup \alpha)$$

where $p \in \boldsymbol{P}$, $i \in G$.

The formula $K_i\phi$ represents the knowledge of agent i that ϕ is the case. The formula $[\alpha]\phi$ can be read as ‘after performing action α, ϕ is the case’. The formula $\odot_i \alpha$ represents the knowledge-based obligation of agent i that he/she ought to do α. It can be read as: i ought to perform action α based on i’s information. The formula $F_i\alpha$ paraphrases ‘it is forbidden for agent i to perform action α’. Since we have an action algebra shown in an inductive manner, we can read obligations (respectively, prohibitions) in different ways with respect to different types of actions.

- $\odot_i(U, e)$: i ought to perform the action (U, e);
- $\odot_i(\alpha_1 \cup \alpha_2)$: i ought to perform action α_1 or α_2.
- $F_i(U, e)$: i is forbidden to perform the action (U, e);
- $F_i(\alpha_1 \cup \alpha_2)$: i is forbidden to perform action α_1 or α_2.

If we consider the obligation to perform a sequential composition of two actions, i.e., $\odot_i(\alpha_1; \alpha_2)$, we can read it as: i ought to perform α_1 and then perform α_2. Similarly, the formula $F_i(\alpha_1; \alpha_2)$ can be read as: i is forbidden to first perform α_1 and then perform α_2.

The truth conditions for $\mathcal{L}_{\text{AKDL}}$-formulas are shown as follows.

Definition 75 (Semantics of $\mathcal{L}_{\text{AKDL}}$). *Given an arbitrary epistemic model $M = \langle S, \sim_1, \sim_2, \cdots, \sim_n, V\rangle$ and an $\mathcal{L}_{\text{EL}}$-priority structure $\mathcal{G}$, the truth conditions of $\mathcal{L}_{\text{AKDL}}$ is defined as follows:*

[1]The abbreviation AKDL refers to the expression ‘action and knowledge-based deontic logic’

$M, s \models p$	*iff*	$s \in V(p)$.
$M, s \models \neg\phi$	*iff*	$M, s \not\models \phi$,
$M, s \models (\phi \wedge \psi)$	*iff*	$M, s \models \phi$ *and* $M, s \models \psi$.
$M, s \models K_i\phi$	*iff*	$M, t \models \phi$ *for each* $t \in [s]^{\sim_i}$.
$M, s \models [\alpha]\phi$	*iff*	*for all* (M', s'): *if* $(M, s)[\![\alpha]\!](M', s')$, *then* $M', s' \models \phi$.
$M, s \models \bigodot_i \alpha$	*iff*	*for all* (M', s'): *if* $(M, s)[\![\alpha]\!](M', s')$, *then* $(M, s) <^i_{\mathcal{G}} (M', s')$.
$M, s \models F_i\alpha$	*iff*	*for all* (M', s'): *if* $(M, s)[\![\alpha]\!](M', s')$, *then* $(M', s') <^i_{\mathcal{G}} (M, s)$.

The intuition of obligations and prohibitions discussed in Section 6.3.2 are reflected in the semantics of $\bigodot_i \alpha$ and $F_i\alpha$. The formula $\bigodot_i \alpha$ is true if and only if performing the action α always causes better situations. In contrast, the semantics of $F_i\alpha$ means that performing the action α always results in worse situations. Therefore, agent i ought to refrain from doing α. In other words, the action α is forbidden for i. It deserves noting that, in $\mathbb{KCDL}$, the agent i knows every knowledge-based conditional obligation that he/she has. But in this chapter, the formula $\bigodot_i \alpha \to K_i \bigodot_i \alpha$ is not valid, which means that an agent might not know his/her own knowledge-based obligations.

6.4.2 Reduction and axiom system $\mathbb{AKDL}$

In this section, we will show that each $\mathcal{L}_{\text{AKDL}}$-formula in the form of $\bigodot_i \alpha$ and $F_i\alpha$ can be reduced to some $\mathcal{L}_{\text{DEL}}$-formula by two reduction axioms, respectively. However, the reduction axioms that we need here are not in Kangerian-Andersonian style anymore. See the following two formulas:

Let $\mathcal{G} = \langle \Phi, \prec \rangle$ be an $\mathcal{L}_{\text{EL}}$-priority structure and let (U, e) be a pointed action model.

$$(\text{RD}\bigodot_i) \quad \bigodot_i(U, e) \leftrightarrow \quad (1) \quad \bigwedge_{\phi \in \Phi \cup \{\top\}} K_i(\phi \to \bigwedge_{e' R_i e}[U, e']\phi) \quad \wedge \quad (2) \quad \neg K_i \neg (\bigvee_{\psi \in \Phi} \bigvee_{e' R_i e}(\langle U, e' \rangle \psi \wedge \neg\psi)).$$

$$(\text{RD}F_i) \quad F_i(U, e) \leftrightarrow \quad (1') \quad \bigwedge_{\phi \in \Phi \cup \{\top\}} K_i(\bigwedge_{e' R_i e}[U, e']\phi \to \phi) \quad \wedge \quad (2) \quad \neg K_i \neg (\bigvee_{\psi \in \Phi} \bigvee_{e' R_i e}(\langle U, e' \rangle \neg\psi \wedge \psi)).$$

The formula $(\text{RD}\bigodot_i)$ represents that $\bigodot_i(U, e)$ is logically equivalent to the conjunction of formulas (1) and (2). Formula (1) intuitively means that for an arbitrary formula $\phi \in \Phi \cup \{\top\}$, for all the states that i cannot distinguish from the current state, if the state satisfies ϕ, then, after updating the model by action (U, e), the updated state satisfies ϕ as well. In other words, the updated state is at least as good as the original state by Definition 72. Formula (2) represents that there exists one state whose updated state satisfies some formula $\psi \in \Phi$, but it does not satisfy ψ itself. Alternatively speaking, formula (2) means that there exists one updated state that i cannot distinguish from the factual state after the update such that it is strictly better than its original state. So it says that i ought to perform the action (U, e).

In contrast, the formula (RDF_i) represents that each original state is at least as good as its updated state. What is more, there exists one updated state that i cannot distinguish such that it is strictly worse than its original state. Therefore, the atomic action (U, e) is forbidden for i.

Proposition 16. *The formula (RD$\bigodot_i$) and (RDF_i) are valid over the class of epistemic models.*

Proof. Let $M = \langle S, \sim_1, \cdots, \sim_n, V \rangle$ be an arbitrary epistemic model. We first prove that (RD$\bigodot_i$) is valid.

($\Rightarrow$) Suppose that $M, s \models \bigodot_i(U, e)$. By semantics, for all (M', s'), if $(M, s)[\![U, e]\!](M', s')$, then $(M, s) <^i_{\mathcal{G}} (M', s')$.

- Let t be an arbitrary state in M such that $s \sim_i t$. Let $\phi \in \Phi \cup \{\top\}$. If $M, t \not\models \phi$, then $M, t \models \phi \to \bigwedge_{e' R_i e}[U, e']\phi$ holds trivially. If $M, t \models \phi$, we need to show that $M, t \models \bigwedge_{e' R_i e}[U, e']\phi$. Let $e' \in U$ such that $e' R_i e$. If $M, t \not\models pre(e')$, then $M, t \models [U, e']\phi$. If $M, t \models pre(e')$, then $(t, e') \in M'$ and $(s, e) R'_i (t, e')$. By Definition 72 and $M, s \models \bigodot_i(U, e)$, we have that $t \leqslant_{\mathcal{G}} (t, e')$. Thus, it holds that $M', (t, e') \models \phi$. So we have $M, t \models [U, e']\phi$. Thus, we have $M, t \models \phi \to \bigwedge_{e' R_i e}[U, e']\phi$. Then we have $M, s \models \bigwedge_{\phi \in \Phi \cup \{\top\}}(\phi \to \bigwedge_{e' R_i e}[U, e']\phi)$.
- By the definition of $(M, s) <^i_{\mathcal{G}} (M', s')$, there exists $t \in [s]^{\sim_i}_M$ and $e' \in U$ such that $(t, e') \in [(s, e)]^{\sim'_i}_{M'}$ and $t <_{\mathcal{G}} (t, e')$. This implies that there exists $\psi \in \Phi$ such that $M, t \not\models \psi$ and $M', (t, e') \models \psi$. Thus, it follows that $M, s \models \neg K_i \neg(\bigvee_{\psi \in \Phi} \bigvee_{e' R_i e}(\langle U, e' \rangle \psi \wedge \neg \psi))$.

($\Leftarrow$) Suppose that $M, s \models \bigwedge_{\phi \in \Phi \cup \{\top\}}(\phi \to \bigwedge_{e' R_i e}[U, e']\phi) \wedge \neg K_i \neg(\bigvee_{\psi \in \Phi} \bigvee_{e' R_i e}(\langle U, e' \rangle \psi \wedge \neg \psi))$.

- By the semantics, for each t such that $t \in [s]^{\sim_i}_M$, $M, t \models \phi \to \bigwedge_{e' R_i e}[U, e']\phi$ for each $\phi \in \Phi$. Let ϕ be an arbitrary formula in Φ such that $M, t \models \phi$. Then $M, t \models \bigwedge_{e' R_i e}[U, e']\phi$. Let e' be an arbitrary event in U. If $M, t \not\models pre'(e')$, then $(t, e') \notin M'$. If $M, t \models pre(e')$, then $M', (t, e') \models \phi$. So we have $t \leqslant_{\mathcal{G}} (t, e')$.
- By $M, s \models \neg K_i \neg(\bigvee_{\psi \in \Phi} \bigvee_{e' R_i e}(\langle U, e' \rangle \psi \wedge \neg \psi))$, there exists $t \in M$ such that $s \sim_i t$ and let $M, t \models \bigvee_{\psi \in \Phi} \bigvee_{e' R_i e}(\langle U, e' \rangle \psi \wedge \neg \psi)$. Let $\psi \in \Phi$ and $e' R_i e$ be the formula and event making $M, t \models \bigvee_{\psi \in \Phi} \bigvee_{e' R_i e}(\langle U, e' \rangle \psi \wedge \neg \psi)$

hold. So we have $M, t \models \langle U, e' \rangle \psi \land \neg\psi$. This implies that $M, t \models \neg\psi$ and $M', (t, e') \models \psi$. Then we have $t <_{\mathcal{G}} (t, e')$.

Therefore, we proved that $M, s \models \bigodot_i (U, e)$.
The validity of the formula (RDF_i) can be proved in a similar way. □

Now we can give the axiom system for the logic of knowledge-based 'ought-to-do' obligation, namely $\mathbb{AKDL}$.

Definition 76. *The proof system $\mathbb{AKDL}$ consists of the following axiom schemas and inference rules:*

(TAUT)	*All instances of tautologies*
(K)	$K_i(\phi \to \psi) \to (K_i\phi \to K_i\psi)$
(T)	$K_i\phi \to \phi$
(4)	$K_i\phi \to K_iK_i\phi$
(5)	$\neg K_i\phi \to K_i\neg K_i\phi$
(U-A)	$[U, e]p \leftrightarrow (pre(e) \to post(e)(p))$
(U-N)	$[U, e]\neg\phi \leftrightarrow (pre(e) \to \neg[U, e]\phi)$
(U-C)	$[U, e](\phi \land \psi) \leftrightarrow ([U, e]\phi \land [U, e]\psi)$
(U-K)	$[U, e]K_i\phi \leftrightarrow (pre(e) \to \bigwedge_{e'R_ie} K_i[(U, e')]\phi)$
(NC)	$[\alpha \cup \alpha']\phi \leftrightarrow ([\alpha]\phi \land [\alpha']\phi)$
(RD$\bigodot_i$)	$\bigodot_i(U, e) \leftrightarrow (\bigwedge_{\phi \in \Phi \cup \{\top\}} K_i(\phi \to \bigwedge_{e'R_ie}[U, e']\phi) \land$ $\neg K_i \neg(\bigvee_{\psi \in \Phi} \bigvee_{e'R_ie}(\langle U, e' \rangle \psi \land \neg\psi)))$
(RDF_i)	$F_i(U, e) \leftrightarrow (\bigwedge_{\phi \in \Phi \cup \{\top\}} K_i(\bigwedge_{e'R_ie}[U, e']\phi \to \phi) \land$ $\neg K_i \neg(\bigvee_{\psi \in \Phi} \bigvee_{e'R_ie}(\langle U, e' \rangle \neg\psi \land \psi)))$
(RNC$\bigodot_i$)	$\bigodot_i(\alpha \cup \alpha') \leftrightarrow (\bigodot_i \alpha \land \bigodot_i \alpha')$
(RNCF_i)	$F_i(\alpha \cup \alpha') \leftrightarrow (F_i\alpha \land F_i\alpha')$
(MP)	*From ϕ and $\phi \to \psi$, infer ψ*
(N)	*From ϕ, infer $K_i\phi$*
(RE)	*From $\phi \leftrightarrow \psi$, infer $\chi \leftrightarrow \chi[\phi/\psi]$*

The axiomatization $\mathbb{AKDL}$ consists of dynamic epistemic logic with post-conditions **UM** established by van Ditmarsch and Kooi (2008), the reduction axiom for non-deterministic choice (NC), and the reduction axioms for obligations in all types of actions, ie. (RD$\bigodot_i$), (RDF_i), (RNC$\bigodot_i$) and (RNCF_i). Moreover, a non-deterministic choice is obligatory for i if and only if both choices (actions) are obligatory. This is also in line with the axiom (NC) which represents that 'after performing a non-deterministic choice, ϕ is the case' is logically equivalent to 'ϕ is the case after performing either of the two actions'.

Theorem 8 (Soundness of $\mathbb{AKDL}$). $\mathbb{AKDL}$ *is sound with respect to the class of epistemic models.*

All axioms and inference rules from **UM** are valid, which have been given by van Ditmarsch and Kooi (2008). The validity of axiom (RC) and (NC) can be referred to Page 152-152 in van Ditmarsch et al. (2007). Proposition 16 proves that (RD$\bigodot_i$) and (RDF_i) are valid. The validity of (RNC$\bigodot_i$) and (RNCF_i) can be given straightforwardly.

Completeness can be proved by translating $\mathcal{L}_{\mathrm{AKDL}}$-formulas to $\mathcal{L}_{\mathrm{EL}}$-formulas via reduction axioms for dynamic operators and deontic operators, and induction on the complexity of the formulas. Let $\mathcal{G} = \langle \Phi, \prec \rangle$ be the priority structure that our system $\mathbb{AKDL}$ is based on. We give the translation and the complexity of $\mathcal{L}_{\mathrm{AKDL}}$-formulas as follows.

Definition 77. *(Translation) The translation $t : \mathcal{L}_{\mathrm{AKDL}} \to \mathcal{L}_{\mathrm{EL}}$ based on $G = \{1, \cdots, n\}$ is defined as follows. For each $i \in G$*

$$
\begin{array}{lll}
t(p) & = & p \\
t(\neg\phi) & = & \neg t(\phi) \\
t(\phi \wedge \psi) & = & t(\phi) \wedge t(\psi) \\
t(K_i\phi) & = & K_i t(\phi) \\
t(\bigodot_i(U,e)) & = & \bigwedge_{\phi\in\Phi\cup\{\top\}} K_i(\phi \to \bigwedge_{e'R_ie} t([U,e']\phi)) \wedge \\
 & & \neg K_i \neg(\bigvee_{\psi\in\Phi} \bigvee_{e'R_ie}(t(\langle U,e'\rangle\psi) \wedge \neg\psi)) \\
t(F_i(U,e)) & = & \bigwedge_{\phi\in\Phi\cup\{\top\}} K_i(\bigwedge_{e'R_ie} t([U,e']\phi) \to \phi) \wedge \\
 & & \neg K_i \neg(\bigvee_{\psi\in\Phi} \bigvee_{e'R_ie}(t(\langle U,e'\rangle\neg\psi) \wedge \psi)) \\
t(\bigodot_i(\alpha_1 \cup \alpha_2)) & = & t(\bigodot_i \alpha_1 \wedge \bigodot_i \alpha_2) \\
t(F_i(\alpha_1 \cup \alpha_2)) & = & t(F_i\alpha_1 \wedge F_i\alpha_2) \\
t([U,e]p) & = & t(pre(e) \to post(e)(p)) \\
t([U,e]\neg\phi) & = & t(pre(e) \to \neg[U,e]\phi) \\
t([U,e](\phi \wedge \psi)) & = & t([U,e]\phi \wedge [U,e]\psi) \\
t([U,e]K_i\phi) & = & t(pre(e) \to \bigwedge_{e'R_ie} K_i[U,e']\phi) \\
t([U_1,e_1]\bigodot_i(U_2,e_2)) & = & \bigwedge_{\phi\in\Phi\cup\{\top\}} t(pre(e_1) \to \\
 & & \bigwedge_{e_1'R_ie_1} K_i[U_1,e_1'](\phi \to \wedge_{e_2'R_ie_2}[U_2,e_2']\phi)) \wedge t(pre(e_1) \to \\
 & & \neg \bigwedge_{e_1'R_ie_1} K_i[U_1,e_1'] \bigwedge_{\psi\in\Phi} \bigwedge_{e_2'R_ie_2}([U_2,e_2']\psi \vee \neg\psi)) \\
t([U_1,e_1]F_i(U_2,e_2)) & = & \bigwedge_{\phi\in\Phi\cup\{\top\}} t(pre(e_1) \to \\
 & & \bigwedge_{e_1'R_ie_1} K_i[U_1,e_1'](\wedge_{e_2'R_ie_2}[U_2,e_2']\phi) \to \phi) \wedge t(pre(e_1) \to \\
 & & \neg \bigwedge_{e_1'R_ie_1} K_i[U_1,e_1'] \bigwedge_{\psi\in\Phi} \bigwedge_{e_2'R_ie_2}([U_2,e_2']\neg\psi \vee \psi)) \\
t([U,e]\bigodot_i(\alpha_1 \cup \alpha_2)) & = & t([U,e]\bigodot_i \alpha_1 \wedge [U,e]\bigodot_i \alpha_2) \\
t([U,e]F_i(\alpha_1 \cup \alpha_2)) & = & t([U,e]F_i\alpha_1 \wedge [U,e]F_i\alpha_2)
\end{array}
$$

Let $\theta = \max\{c(\phi) \mid \phi \in \Phi\}$. In other words, θ is the formula in the domain of the priority structure Φ such that the complexity of θ is the maximal among all formulas in Φ.

Definition 78. *(Complexity of $\mathcal{L}_{\text{AKDL}}$) The complexity $c : \mathcal{L}_{\text{AKDL}} \to \mathbb{N}$ based on $G = \{1, \cdots, n\}$ is defined as follows: for each $i \in G$,*

$$\begin{array}{lcl} c(p) & = & 1 \\ c(\neg\phi) & = & 1 + c(\phi) \\ c(\phi \wedge \psi) & = & 1 + \max(c(\phi), c(\psi))) \\ c(K_i\phi) & = & 1 + c(\phi) \\ c([U,e]\phi) & = & (3 + |U| + c(U)) \cdot c(\phi) \\ c(\bigodot_i(U,e)) & = & 5 + 2|\Phi| + 2|U| + c([U,e]\theta) \\ c(F_i(U,e)) & = & 5 + 2|\Phi| + 2|U| + c([U,e]\theta) \\ c(\bigodot_i(\alpha_1 \cup \alpha_2)) & = & 1 + \max(c(\bigodot_i \alpha_1), c(\bigodot_i \alpha_2)) \\ c(F_i(\alpha_1 \cup \alpha_2)) & = & 1 + \max(c(F_i\alpha_1), c(F_i\alpha_2)) \end{array}$$

where $c(U) = \max(c(pre(e_1)), \cdots, c(pre(e_{|U|})), c(post(e_1)(p_1)), \cdots, c(post(e_1)(p_k)), \cdots c(post(e_{|U|})(p_1)), \cdots, c(post(e_{|U|})(p_k)))))$.

It is worth noting that the complexity of $\mathcal{L}_{\text{AKDL}}$ is different from that of $\mathcal{L}_{\text{DKCDL}}$ in the aspect of the deontic formulas. And the complexity of obligation and prohibition are the same.

The completeness proof is analogous to Theorem 4 for the system $\mathbb{DKCDL}$. The constant 5 is the least number that can make the proof go through when we do induction on the complexity of $\mathcal{L}_{\text{AKDL}}$-formulas. The proof details are shown as follows.

Lemma 11. *For all $\mathcal{L}_{\text{AKDL}}$-formula ϕ, it holds that $\vdash_{\mathbb{AKDL}} \phi \leftrightarrow t(\phi)$ and $t(\phi) \in \mathcal{L}_{\text{EL}}$.*

Proof. By induction on $c(\phi)$.

• **Base case**: When $\phi = p$ for some propositional atom p, it is trivial that $\vdash_{\mathbb{AKDL}} p \leftrightarrow p$ and $p \in \mathcal{L}_{\text{EL}}$.

• **Induction hypothesis**: For all ϕ such that $c(\phi) < n$: we have $\vdash \phi \leftrightarrow t(\phi)$ and $t(\phi) \in \mathcal{L}_{\text{EL}}$.

• **Induction step**: If $c(\phi) = n + 1$:

- When $\phi = \neg\psi$, we have $c(\neg\psi) = 1 + c(\psi)$. So $c(\psi) = n$. By induction hypothesis, we get $\vdash_{\mathbb{AKDL}} \psi \leftrightarrow t(\psi)$ and $t(\psi) \in \mathcal{L}_{\text{EL}}$. Thus, $\vdash_{\mathbb{AKDL}} \neg\psi \leftrightarrow \neg t(\psi)$. It just is $\vdash_{\mathbb{AKDL}} \phi \leftrightarrow t(\phi)$. And $\neg t(\psi) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = (\psi_1 \wedge \psi_2)$, we have $c(\psi_1 \wedge \psi_2) = 1 + \max(c(\psi_1, c(\psi_2)))$. So $\max(c(\psi_1), c(\psi_2)) = n$. It means that $c(\psi_1) \leq n$ and $c(\psi_2) \leq n$. By induction hypothesis, we have $\vdash_{\mathbb{AKDL}} \psi_1 \leftrightarrow t(\psi_1)$, $\vdash_{\mathbb{AKDL}} \psi_2 \leftrightarrow t(\psi_2)$, $t(\psi_1) \in \mathcal{L}_{\text{EL}}$ and $t(\psi_2) \in \mathcal{L}_{\text{EL}}$. Then we have $\vdash_{\mathbb{AKDL}} (\psi_1 \wedge \psi_2) \leftrightarrow (t(\psi_1) \wedge t(\psi_2))$. It is equivalent to $\vdash_{\mathbb{AKDL}} (\psi_1 \wedge \psi_2) \leftrightarrow t(\psi_1 \wedge \psi_2)$ by our translation. And $t(\psi_1) \wedge t(\psi_2) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = K_i\psi$: $c(K_i\psi) = 1 + c(\psi)$. So $c(\psi) = n$. By induction hypothesis, we have $\vdash_{\mathbb{AKDL}} \psi \leftrightarrow t(\psi)$ and $t(\psi) \in \mathcal{L}_{\text{EL}}$. By (NEC) and (K), $\vdash_{\mathbb{AKDL}} K_i\psi \leftrightarrow K_i(t(\psi))$. It is equivalent to $\vdash_{\mathbb{AKDL}} K_i\psi \leftrightarrow t(K_i\psi)$ by our translation. And we also have $K_i t(\psi) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = \bigodot_i(U,e)$: $c(\bigodot_i(U,e)) = 5 + 2|\Phi| + 2|U| + c([U,e]\theta)$. According to Definition 77, we have $t(\bigodot_i(U,e)) = \bigwedge_{\chi\in\Phi\cup\{\top\}} K_i(\chi \rightarrow \bigwedge_{e'R_ie} t([U,e']\chi)) \wedge \neg K_i\neg(\bigvee_{\psi\in\Phi}\bigvee_{e'R_ie}(t(\langle U,e'\rangle\psi) \wedge \neg\psi))$. Since $t([U,e']\chi) = (3 + |U| + c(U)) \cdot c(\chi)$ and $c(\chi) < c(\theta)$, it is obtained that $c([U,e'])\chi \leq c([U,e']\theta)$. So $c([U,e']\chi) < c(\bigodot_i(U,e))$. By induction hypothesis, it holds that $\vdash_{\mathbb{AKDL}} [U,e']\chi \leftrightarrow t([U,e']\chi)$. Then, by the axiom (RD$\bigodot_i$) and the inference rule (RE), we have $\vdash_{\mathbb{AKDL}} \bigodot_i(U,e) \leftrightarrow t(\bigodot_i(U,e))$. Since $c([U,e']\chi) \leq n$, by induction hypothesis, we have $t([U,e']\chi) \in \mathcal{L}_{\text{EL}}$. So it holds that $t(\bigodot_i(U,e)) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = F_i(U,e)$, the proof is similar to the above case.

- When $\phi = \bigodot_i(\alpha_1 \cup \alpha_2)$, we know $t(\bigodot_i(\alpha_1 \cup \alpha_2)) = t(\bigodot_i \alpha_1 \wedge \bigodot_i \alpha_2)$. By Definition 78, we have $c(\bigodot_i(\alpha_1 \cup \alpha_2)) = 1 + \max(c(\bigodot_i \alpha_1), c(\bigodot_i \alpha_2))$. So we have $c(\bigodot_i \alpha_1) < n$ and $c(\bigodot_i \alpha_2) < n$. By induction hypothesis, $\vdash_{\mathbb{AKDL}} t(\bigodot_i \alpha_1) \leftrightarrow \bigodot_i \alpha_1$ and $\vdash_{\mathbb{AKDL}} t(\bigodot_i \alpha_2) \leftrightarrow \bigodot_i \alpha_2$. So we have $\vdash_{\mathbb{AKDL}} (\bigodot_i \alpha_1 \wedge \bigodot_i \alpha_2) \leftrightarrow (t(\bigodot_i \alpha_1) \wedge t(\bigodot_i \alpha_2))$. By the axiom (RNC$\bigodot_i$) and Definition 77, we have $\vdash_{\mathbb{AKDL}} \bigodot_i(\alpha_1 \cup \alpha_2) \leftrightarrow t(\bigodot_i(\alpha_1 \cup \alpha_2))$.

  Now we show $t(\bigodot_i(\alpha_1 \cup \alpha_2)) \in \mathcal{L}_{\text{EL}}$. Since $t(\bigodot_i(\alpha_1 \cup \alpha_2)) = t(\bigodot_i \alpha_1 \wedge \bigodot_i \alpha_2) = t(\bigodot_i \alpha_1) \wedge t(\bigodot_i \alpha_2)$. By induction hypothesis, we have $t(\bigodot_i \alpha_1) \in \mathcal{L}_{\text{EL}}$ and $t(\bigodot_i \alpha_2) \in \mathcal{L}_{\text{EL}}$. Therefore, it follows that $t(\bigodot_i(\alpha_1 \cup \alpha_2)) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = F_i(\alpha_1 \cup \alpha_2)$, the proof is similar to the above case.

- When $\phi = [U,e]p$, we know $c([U,e]p) = 3 + |U| + c(U)$.

  We first prove $\vdash_{\mathbb{AKDL}} \phi \leftrightarrow t(\phi)$. We have $c(pre(e) \rightarrow post(e)(p)) = 2 + \max(c(pre(e)), c(\neg post(e)(p)))$. If $c(pre(e)) \geq c(\neg post(e)(p))$, by $c(U) \geq c(pre(e))$, we have $c([U,e]p) > c(pre(e) \rightarrow post(e)p)$. If $c(\neg post(e)(p)) \geq c(pre(e))$, then $c(pre(e) \rightarrow post(e)(p)) = 3 + c(post(e)(p))$. By $c(U) \geq c(post(e)(p))$ and $|U| \geq 1$, we obtain $c([U,e]p) > c(pre(e) \rightarrow post(e)p)$. By induction hypothesis, we have $\vdash_{\mathbb{AKDL}} (pre(e) \rightarrow post(e)(p)) \leftrightarrow t(pre(e) \rightarrow post(e)(p))$. It is equivalent to $\vdash_{\mathbb{AKDL}} [U,e]p \leftrightarrow t([U,e]p)$ by Definition 77 and the inference rule (RE).

  Then we prove $t(\phi) \in \mathcal{L}_{\text{EL}}$. We have $t([U,e]p) = t(\neg(pre(e) \wedge \neg post(e)(p))) = \neg(t(pre(e)) \wedge t(\neg post(e)(p)))$. We know that $c(pre(e)) \leq c(U)$ and $c(\neg post(e)(p)) \leq c(U) + 1$. So we have $c(pre(e)) \leq n$ and $c(\neg post(e)(p)) \leq n$. By induction hypothesis, we have $t(pre(e)) \in \mathcal{L}_{\text{EL}}$ and $t(\neg post(e)(p)) \leq \mathcal{L}_{\text{EL}}$. Thus, $t([U,e]p) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = [U,e]\neg\psi$: $c([U,e]\neg\psi) = (3 + |U| + c(U)) \cdot (1 + c(\psi)) = 3 + |U| + c(U) + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$.

  We first prove $\vdash_{\mathbb{AKDL}} \phi \leftrightarrow t(\phi)$. We know $c(pre(e) \to \neg[U,e]\psi) = 2 + \max(c(pre(e)), c(\neg[U,e]\psi))$. Since $c(\neg[U,e]\psi) > c(pre(e))$, we have $c(pre(e) \to \neg[U,e]\psi) = 3 + c([U,e]\psi) = 3 + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$. So $c([U,e]\neg\psi) > c(pre(e) \to \neg[U,e]\psi)$. By induction hypothesis, we obtain $\vdash_{\mathbb{AKDL}} (pre(e) \to \neg[U,e]\psi) \leftrightarrow t(pre(e) \to \neg[U,e]\psi)$. It follows that $\vdash_{\mathbb{AKDL}} [U,e]\neg\psi \leftrightarrow t([U,e]\neg\psi)$ by the inference rule (RE) and Definition 77.

  Then we prove $t(\phi) \in \mathcal{L}_{\text{EL}}$. We know $t([U,e]\neg\psi) = t(pre(e) \to \neg[U,e]\psi) = t(\neg(t(pre(e) \wedge t(\neg post(e)(p)))))$. Since $c(pre(e)) \leq c(U)$ and $c([U,e]\psi) < c([U,e]\neg\psi)$, by induction hypothesis, we have $t(pre(e)) \in \mathcal{L}_{\text{EL}}$ and $t([U,e]\psi) \in \mathcal{L}_{\text{EL}}$. Thus, we obtain $t([U,e]\neg\psi) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = [U,e](\psi_1 \wedge \psi_2)$: $c([U,e](\psi_1 \wedge \psi_2)) = (3 + |U| + c(U)) \cdot (\max(c(\psi_1), c(\psi_2)) + 1)$.

  We first prove $\vdash_{\mathbb{AKDL}} \phi \leftrightarrow t(\phi)$. Assuming $c(\psi_1) \geq c(\psi_2)$, we have $c([U,e](\psi_1 \wedge \psi_2)) = 3 + |U| + c(U) + 3 \cdot c(\psi_1) + |U| \cdot c(\psi_1) + c(U) \cdot c(\psi_1)$. We also know $c([U,e]\psi_1 \wedge [U,e]\psi_2) = 1 + \max(c([U,e]\psi_1), c([U,e]\psi_2))$. By $c(\psi_1) \geq c(\psi_2)$, we have $c([U,e]\psi_1 \wedge [U,e]\psi_2) = 4 + |U| + c(U) + c(U) \cdot c(\psi_1)$. So $c([U,e](\psi_1 \wedge \psi_2)) > c([U,e]\psi_1 \wedge [U,e]\psi_2)$. By induction hypothesis, we obtain $\vdash_{\mathbb{AKDL}} ([U,e]\psi_1 \wedge [U,e]\psi_2) \leftrightarrow t([U,e]\psi_1 \wedge [U,e]\psi_2)$. It is equivalent to $\vdash_{\mathbb{AKDL}} [U,e](\psi_1 \wedge \psi_2) \leftrightarrow t([U,e](\psi_1 \wedge \psi_2))$ by (RE) and Definition 77.

  Then we prove $t(\phi) \in \mathcal{L}_{\text{EL}}$. Assuming $c(\psi_1) \geq c(\psi_2)$, we have $c([U,e](\psi_1 \wedge \psi_2)) = 3 + |U| + c(U) + 3 \cdot c(\psi_1) + |U| \cdot c(\psi_1) + c(U) \cdot c(\psi_1)$. We know $t([U,e](\psi_1 \wedge \psi_2)) = t([U,e]\psi_1) \wedge t([U,e]\psi_2)$. Since $c([U,e]\psi_1) < c([U,e](\psi_1 \wedge \psi_2))$ and $c([U,e]\psi_2) < c([U,e](\psi_1 \wedge \psi_2))$, by induction hypothesis, we have $t([U,e]\psi_1) \in \mathcal{L}_{\text{EL}}$ and $t([U,e]\psi_2) \in \mathcal{L}_{\text{EL}}$. Thus, we obtain $t([U,e](\psi_1 \wedge \psi_2)) \in \mathcal{L}_{\text{EL}}$.

- When $\phi = [U,e]K_i\psi$, $c([U,e]K_i\psi) = 3 + |U| + c(U) + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$.

  We first prove $\vdash_{\mathbb{AKDL}} \phi \leftrightarrow t(\phi)$. We know $c(pre(e) \to \bigwedge_{e'_i R_i e} K_i[U,e']\psi) = 2 + \max(c(pre(e)), c(\neg \bigwedge_{e'_i R_i e} K_i[U,e']\psi))$. We also know $c(\neg \bigwedge_{e'_i R_i e} K_i[U,e']\psi) = 1 + c(\bigwedge_{e'_i R_i e} K_i[U,e']\psi) = 1 + |U| - 1 + \max(c(K_i[U,e_1]\psi), \cdots, c(K_i[U,e_{|U|}]\psi)) = 1 + |U| + \max(c([U,e_1]\psi), \cdots, c([U,e_{|U|}]\psi))$. Let $m \in \mathbb{N}$ such that $1 \leq m \leq |U|$ and $c([U,e_m]\psi) = \max(c([U,e_1]\psi), \cdots, c([U,e_{|U|}]\psi))$. Then $c(pre(e) \to \bigwedge_{e'_i R_i e} K_i[U,e']\psi)) = 3 + |U| + c([U,e_m]\psi) = 3 + |U| + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$. So it is easy to see that

$c([U,e]K_i\psi) > c(pre(e) \to \bigwedge_{e'_i R_i e} K_i[U,e']\psi))$. By induction hypothesis, we obtain $\vdash_{\mathbb{AKDL}} (pre(e) \to \bigwedge_{e'_i R_i e} K_i[U,e']\psi)) \leftrightarrow t(pre(e) \to \bigwedge_{e'_i R_i e} K_i[U,e']\psi))$. It is equivalent to $\vdash_{\mathbb{AKDL}} [U,e]K_i\psi \leftrightarrow t([U,e]K_i\psi)$ by (RE) and our translation.

Then we prove $t(\phi) \in \mathcal{L}_{\mathrm{EL}}$. We know $t([U,e]K_i\psi) = \neg t(pre(e) \to \bigwedge_{e' R_i e} K_i[U,e']\psi) = \neg(t(pre(e)) \wedge t(\neg \bigwedge_{e' R_i e} K_i[U,e']\psi))$. We know $c(pre(e)) \leq c(U)$. And $c(\neg \bigwedge_{e' R_i e} K_i[U,e']\psi) = 1 + |U| + \max(c(K_i[U,e_1]\psi), \cdots, c([U,e_m]\psi))$ where $\{e_1, \cdots, e_m\} = [e]^{R_i}$. Since for each $e_k, e_l \in U$, we have $c([U,e_k]\psi) = c([U,e_l]\psi)$. So $c(\neg \bigwedge_{e' R_i e} K_i[U,e']\psi) = 1 + |U| + c(K_i[U,e]\psi) = 2 + |U| + c([U,e]\psi) = 2 + |U| + c([U,e]\psi) = 2 + |U| + 3 \cdot c(\psi) + |U| \cdot c(\psi) + c(U) \cdot c(\psi)$ which is strictly smaller than $c([U,e]K_i\psi)$. By induction hypothesis, $t(\neg \bigwedge_{e' R_i e} K_i[U,e']\psi) \in \mathcal{L}_{\mathrm{EL}}$. Thus, we obtain $t([U,e]K_i\psi) \in \mathcal{L}_{\mathrm{EL}}$.

- When $\phi = [U_1,e_1] \odot_i (U_2,e_2)$, $c([U_1,e_1] \odot_i (U_2,e_2)) = (3 + |U_1| + c(U_1)) \cdot (5 + 2|\Phi| + 2|U_2| + c([U_2,e_2]\theta)) = 15 + 6|\Phi| + 6|U_2| + 3c([U_2,e_2]\theta) + 5|U_1| + 2|\Phi||U_1| + 2|U_1||U_2| + |U_1|c([U_2,e_2]\theta) + 5c(U_1) + 2|\Phi|c(U_1) + 2|U_2|c(U_1) + c(U_1)c([U_2,e_2]\theta)$. First of all, by the axiom (U-N), (U-C), (U-K), (RD$\odot_i$) and the inference rule (RE), we have $\vdash_{\mathbb{AKDL}} [U_1,e_1] \odot_i (U_2,e_2) \leftrightarrow (\bigwedge_{\phi \in \Phi \cup \{\top\}} (pre(e_1) \to \bigwedge_{e'_1 R_i e_1} K_i[U_1,e'_1](\phi \to \wedge_{e'_2 R_i e_2} [U_2,e'_2]\phi)) \wedge (pre(e_1) \to \neg \bigwedge_{e'_1 R_i e_1} K_i[U_1,e'_1] \bigwedge_{\psi \in \Phi} \bigwedge_{e'_2 R_i e_2} ([U_2,e'_2]\psi \vee \neg\psi)))$.

  We first prove $\vdash_{\mathbb{AKDL}} \phi \leftrightarrow t(\phi)$. We have $c(pre(e_1) \to \bigwedge_{e'_1 R_i e_1} K_i[U_1,e'_1](\theta \to \bigwedge_{e'_2 R_i e_2} [U_2,e'_2]\theta)) = 3 + |U_1| + (3 + |U_1| + c(U_1)) \cdot (1 + |U_2| + c([U_2,e'_2]\theta))$. We also have $c(pre(e_1) \to \bigwedge_{e'_1 R_i e} K_i[U_1,e'_1] \bigwedge_{\psi \in \Phi} \bigwedge_{e'_2 R_i e_2} ([U_2,e'_2]\psi \vee \neg\psi)) = 3 + |U_2| + (3 + |U_1| + c(U_1)) \cdot (|\Phi| + |U_2| + c([U_2,e'_2]\theta \vee \neg\theta)) = 12 + 3|\Phi| + 4|U_2| + (3 + |\Phi|)|U_1| + |U_1||U_2| + (3 + |\Phi|)c(U_1) + |U_2|c(U_1) + (3 + |U_1| + c(U_1)) \cdot c([U_2,e'_2]\theta)$. Then we obtain that $c([U_1,e_1] \odot (U_2,e_2)) > c(pre(e_1) \to \bigwedge_{e'_1 R_i e_1} K_i[U_1,e'_1](\theta \to \bigwedge_{e'_2 R_i e_2} [U_2,e'_2]\theta))$ and $c([U_1,e_1] \odot (U_2,e_2)) > c(pre(e_1) \to \bigwedge_{e'_1 R_i e} K_i[U_1,e'_1] \bigwedge_{\psi \in \Phi} \bigwedge_{e'_2 R_i e_2} ([U_2,e'_2]\psi \vee \neg\psi))$. By induction hypothesis, we have $\vdash_{\mathbb{AKDL}} (pre(e_1) \to \bigwedge_{e'_1 R_i e_1} K_i[U_1,e'_1](\theta \to \bigwedge_{e'_2 R_i e_2} [U_2,e'_2]\theta)) \leftrightarrow t(pre(e_1) \to \bigwedge_{e'_1 R_i e_1} K_i[U_1,e'_1](\theta \to \bigwedge_{e'_2 R_i e_2} [U_2,e'_2]\theta))$ and $\vdash_{\mathbb{AKDL}} (pre(e_1) \to \bigwedge_{e'_1 R_i e} K_i[U_1,e'_1] \bigwedge_{\psi \in \Phi} \bigwedge_{e'_2 R_i e_2} ([U_2,e'_2]\psi \vee \neg\psi)) \leftrightarrow t(pre(e_1) \to \bigwedge_{e'_1 R_i e} K_i[U_1,e'_1] \bigwedge_{\psi \in \Phi} \bigwedge_{e'_2 R_i e_2} ([U_2,e'_2]\psi \vee \neg\psi))$. Thus, by the inference rule (RE), it is obtained that $\vdash_{\mathbb{AKDL}} [U_1,e_1] \odot_i (U_2,e_2) \leftrightarrow \bigwedge_{\phi \in \Phi \cup \{\top\}} t(pre(e_1) \to \bigwedge_{e'_1 R_i e_1} K_i[U_1,e'_1](\phi \to \wedge_{e'_2 R_i e_2} [U_2,e'_2]\phi)) \wedge t(pre(e_1) \to \neg \bigwedge_{e'_1 R_i e_1} K_i[U_1,e'_1] \bigwedge_{\psi \in \Phi} \bigwedge_{e'_2 R_i e_2} ([U_2,e'_2]\psi \vee \neg\psi))$. This implies $\vdash_{\mathbb{AKDL}} [U_1,e_1] \odot_i (U_2,e_2) \leftrightarrow t([U_1,e_1] \odot_i (U_2,e_2))$.

  Then we prove $t([U_1,e_1] \odot_i (U_2,e_2)) \in \mathcal{L}_{\mathrm{EL}}$. Since we have proved that $c([U_1, e_1] \bigodot (U_2, e_2)) > c(pre(e_1) \rightarrow \bigwedge_{e'_1 R_i e_1} K_i[U_1, e'_1](\theta \rightarrow \bigwedge_{e'_2 R_i e_2}[U_2, e'_2]\theta))$ and $c([U_1, e_1] \bigodot (U_2, e_2)) > c(pre(e_1) \rightarrow \bigwedge_{e'_1 R_i e} K_i[U_1, e'_1] \bigwedge_{\psi \in \Phi} \bigwedge_{e'_2 R_i e_2}([U_2, e'_2]\psi \vee \neg\psi))$, by induction hypothesis, we know that $t(pre(e_1) \rightarrow \bigwedge_{e'_1 R_i e_1} K_i[U_1, e'_1](\theta \rightarrow \bigwedge_{e'_2 R_i e_2}[U_2, e'_2]\theta)) \in \mathcal{L}_{\mathrm{EL}}$ and $t(pre(e_1) \rightarrow \bigwedge_{e'_1 R_i e} K_i[U_1, e'_1] \bigwedge_{\psi \in \Phi} \bigwedge_{e'_2 R_i e_2}([U_2, e'_2]\psi \vee \neg\psi)) \in \mathcal{L}_{\mathrm{EL}}$. Therefore, it is followed that $t([U_1, e_1] \bigodot_i (U_2, e_2)) \in \mathcal{L}_{\mathrm{EL}}$.

- When $\phi = [U, e] \bigodot_i(\alpha_1 \cup \alpha_2)$, we know $t([U, e] \bigodot_i(\alpha_1 \cup \alpha_2)) = t([U, e] \bigodot_i \alpha_1 \wedge [U, e] \bigodot_i \alpha_2)$. We have $c([U, e] \bigodot_i(\alpha_1 \cup \alpha_2)) = (3 + |U| + c(U)) \cdot (1 + \max(c(\bigodot_i \alpha_1), c(\bigodot_i \alpha_2)))$. Assume $c(\bigodot_i \alpha_1) \geq c(\bigodot_i \alpha_2)$. So $c([U, e] \bigodot_i(\alpha_1 \cup \alpha_2)) = 3 + |U| + c(U) + 3c(\bigodot_i \alpha_1) + |U|c(\bigodot_i \alpha_1) + c(U) \cdot c(\bigodot_i \alpha_1)$. On the other hand, we have $c([U, e] \bigodot_i \alpha_1 \wedge [U, e] \bigodot_i \alpha_2) = 1 + (3 + |U| + c(U)) \cdot c(\bigodot_i \alpha_1)$. This is strictly smaller than $c([U, e] \bigodot_i(\alpha_1 \cup \alpha_2))$. By induction hypothesis, $\vdash_{\mathbb{AKDL}} ([U, e] \bigodot_i \alpha_1 \wedge [U, e] \bigodot_i \alpha_2) \leftrightarrow t([U, e] \bigodot_i \alpha_1 \wedge [U, e] \bigodot_i \alpha_2)$. By the axiom (RNC$\bigodot_i$), the inference rule (RE) and Definition 77, it holds that $\vdash [U, e] \bigodot_i(\alpha_1 \cup \alpha_2) \leftrightarrow t([U, e] \bigodot_i(\alpha_1 \cup \alpha_2))$.

  Now we prove $t([U, e] \bigodot_i(\alpha_1 \cup \alpha_2)) \in \mathcal{L}_{\mathrm{EL}}$. We know $t([U, e] \bigodot_i(\alpha_1 \cup \alpha_2)) = t([U, e] \bigodot_i \alpha_1 \wedge [U, e] \bigodot_i \alpha_2) \in \mathcal{L}_{\mathrm{EL}}$. Since $c([U, e] \bigodot_i \alpha_1) < n$ and $c([U, e] \bigodot_i \alpha_2) < n$, by induction hypothesis, $t([U, e] \bigodot_i \alpha_1) \in \mathcal{L}_{\mathrm{EL}}$ and $t([U, e] \bigodot_i \alpha_2) \in \mathcal{L}_{\mathrm{EL}}$. So we have $t([U, e] \bigodot_i \alpha_1) \wedge t([U, e] \bigodot_i \alpha_2) \in \mathcal{L}_{\mathrm{EL}}$. This implies that $t([U, e] \bigodot_i(\alpha_1 \cup \alpha_2)) \in \mathcal{L}_{\mathrm{EL}}$.

- When $\phi = [U, e]F_i(\alpha_1 \cup \alpha_2)$, the proof is similar to the above case.

□

Next we give the strong completeness proof for $\mathbb{AKDL}$.

Theorem 9 (Strong completeness of $\mathbb{AKDL}$)**.** *For every set of formulas $\Gamma \cup \{\phi\} \subseteq \mathcal{L}_{\mathrm{AKDL}}$, $\Gamma \models \phi$ implies $\Gamma \vdash_{\mathbb{AKDL}} \phi$.*

Proof. Suppose that $\Gamma \models \phi$. By Lemma 11 and soundness of $\mathbb{AKDL}$, we have $t(\Gamma) \models t(\phi)$. Since $t(\Gamma) \subseteq \mathcal{L}_{\mathrm{EL}}$ and $t(\phi) \in \mathcal{L}_{\mathrm{EL}}$, by strong completeness of $\mathbb{EL}$ with respect to the class of epistemic models, we obtain $t(\Gamma) \vdash_{\mathbb{EL}} t(\phi)$. This implies that there is a syntactic proof S where we can derive $t(\phi)$ from a finite set $\Lambda \subseteq t(\Gamma)$ by $\mathbb{EL}$. Let $\Lambda = \langle \psi_1, \psi_2, \cdots, \psi_m \rangle$. For each $\psi_n \in \Lambda$, we have $\vdash_{\mathbb{AKDL}} \psi_n \leftrightarrow t(\psi_n)$. We know that S is a sequence of formulas. Then we can give a $\mathbb{AKDL}$-syntactic proof which can derive ϕ from Λ as follows:

(1)	$\vdash_{\mathbb{AKDL}} \psi_1 \leftrightarrow t(\psi_1)$
(2)	$\vdash_{\mathbb{AKDL}} \psi_2 \leftrightarrow t(\psi_2)$
$\vdots$	
(m)	$\vdash_{\mathbb{AKDL}} \psi_m \leftrightarrow t(\psi_m)$
$\vdots$	S
(m+$\lvert S \rvert$+1)	$\vdash_{\mathbb{AKDL}} \phi$
(m+$\lvert S \rvert$+2)	$\vdash_{\mathbb{AKDL}} \phi \leftrightarrow t(\phi)$
(m+$\lvert S \rvert$+3)	$\vdash_{\mathbb{AKDL}} \phi$

Therefore, we conclude that $\Gamma \vdash_{\mathbb{AKDL}} \phi$. □

Corollary 4. $\mathbb{AKDL}$ *is sound and strongly complete with respect to the class of epistemic models.*

6.4.3 Ought to be vs. Ought to do

The distinction between the notion of 'ought-to-be' and the notion of 'ought-to-do' is significant. We mainly discuss two differences, focused around the problem of 'ought implies can' and the problem around agency.

The problem around 'ought implies can' The approaches where deontic operators are applied to propositions conceptualize the notion of 'ought-to-be'. They characterize what states of affairs ought to be achieved. In English, a sentence like 'it ought to be the case that there were no Second World War' indicates that 'ought-to-be' can be applied to unreal or impossible states of affairs. In this sense, 'ought-to-be' is not in line with the Kantian principle 'ought implies can'. In contrast, 'ought-to-do' requires that the obligations can be done by agents. For example, 'you ought to not start the Second World War' suggests that the Second World War has not happened yet and you have an obligation to refrain from starting it. Hence, we can say that the Kantian principle fits to 'ought-to-do', instead of 'ought-to-be'.

Chapter 3, 4 and 5 all discuss 'ought-to-be'. The notion of knowledge-based conditional obligations is defined over an epistemic betterness structure based on a priority structure. In this section, we only focus on unconditional obligations. Let us review two definitions studied in Chapter 3:

Definition 79. *Given a priority structure $\mathcal{G}$ and an epistemic betterness structure $M = \langle S, \sim_1, \sim_2, \cdots, \sim_n, \leqslant_{\mathcal{G}}, V\rangle$ based on $\mathcal{G}$,*

$$M, s \models \bigcirc\phi \quad \textit{iff} \quad \max_{\leqslant_{\mathcal{G}}} S \subseteq \|\phi\|_M.$$
$$M, s \models \odot_i \phi \quad \textit{iff} \quad \max_{\leqslant_{\mathcal{G}}} [s]^{\sim_i} \subseteq \|\phi\|_M.$$

As for objective unconditional obligation, its definition focuses on all states. The set of states in an epistemic betterness structure can be chosen according to any criteria, e.g., ontic possibility. So there is a state where there is no Second World War and a state where there is the Second World War. The priority structure shows that 'there is no war' is better than 'there is the war' and hence, the first state is better than the second state. We can also easily obtain that it ought to be that there is no war, even though the Second World War has already happened.

However, the logic $\mathbb{AKDL}$ introduced in this Chapter applies the deontic operator to actions. The action of not starting the Second World War can never be executable at the current moment. Therefore, this action will never become an obligation of some agent. According to this comparison to Hansson's objective unconditional obligation, $\mathbb{AKDL}$ is able to conceptualize the notion of 'ought-to-do' properly.

The problem around agency In terms of epistemic unconditional obligation, the semantics of the $\mathcal{L}_{\text{KCDL}}$-formula $\odot_i(\phi|\top)$ makes formula $K_i\phi \to \odot_i(\phi|\top)$ valid. The formula can be read as 'if an agent knows something, it ought to be the case', which is extremely counter-intuitive. For example, if I know President Kennedy was assassinated, then it ought to be the case that President Kennedy was assassinated. If we review Hansson's objective unconditional obligation, i.e., $\bigcirc(\phi|\top)$, a similar formula $\Box\phi \to \bigcirc(\phi|\top)$ is also valid in betterness structures. Moreover, in standard deontic logic, the formula $\Box\phi \to \bigcirc\phi$ is valid as well. However, deontic stit logic does not have the problem. Therefore, it seems that this is a common drawback of these logics which characterize 'ought-to-be' obligations without considering the notion of agency.

In the tradition of studies on 'ought-to-be' obligations, a set of possible states and an ordering on the set are given *a priori*. 'Ought-to-be' obligations are those states of affairs that hold on all the best states. We can alternatively say that 'ought-to-be' obligations are the best states of affairs chosen from a given situation (possible states and an ordering on them). This underlying idea of defining obligations leaves no space for the notion of agency since the situation is fixed in general.

But for these logics of 'ought-to-do' obligations, e.g., Meyer's logic, and these deontic logics involving agency, e.g., deontic stit logic, the problem no longer exists. Also, the knowledge-based 'ought-to-do' obligation defined in this chapter provides an approach to assessing whether an action is ideal. Beginning with an initial epistemic model containing the states that are epistemically possible for some agent, those actions that lead to better models are obligations. If they result in strictly worse situations, they are forbidden. So only executable but not performed actions can be candidates of some agent's obligations in the current situation. These actions *can* change something. The obligations are to make the situations better.

6.5 Resolving paradoxes of classical deontic logic

In this section, we investigate some paradoxes that have plagued deontic logic for a long time. Meyer et al. (1994) and McNamara (2006) provided a list of well-known paradoxes from the deontic logic literature. We mainly focus on the issues about obligations or prohibition. Since the negation of actions is not treated in this chapter, we do not discuss the paradoxes involving 'not doing ...'.

As mentioned in Section 6.3.2, the main similarity between our approach and Meyer's approach is treating obligations as actions which can make transitions between states or situations. The paradoxes mentioned in the following part can also be resolved under Meyer's framework. So our approach works in dealing with these paradoxes since we define obligations with respect to actions as well. Now we will show how our knowledge-based 'ought-to-do' obligation introduced in this chapter resolves these paradoxes properly.

6.5.1 Ross's Paradox and non-deterministic choice

Ross's Paradox (see Ross (1944)) is also known as 'disjunction paradox' which is generally illustrated by the following informal argument:

1. It is obligatory that the letter is mailed.
2. If the letter is mailed, then the letter is mailed or burnt.

---

3. It is obligatory that the letter is mailed or burnt.

The above argument can be formalized in standard deontic logic as $\bigcirc\phi \to \bigcirc(\phi \vee \psi)$, which is obviously valid since $\phi \to (\phi \vee \psi)$ is always true by propositional logic. What is worse, if a deontic logic satisfies the following two conditions, it would suffer from Ross's paradox:

1. using Kripke-style frames and defining obligation operator $\bigcirc$ as a kind of necessity;

2. obligations are propositions and inferences between obligations are therefore based on classical logical inference.

Unfortunately, standard deontic logic and the deontic stit logic both satisfy the above conditions. So it is inevitable that they are challenged by Ross's paradox. However, Meyer (1988)'s deontic logic based on propositional dynamic logic provides a resolution. We can reformulate the paradox by $\mathcal{L}_{\text{AKDL}}$ as follows:

(1) $M, s \models \odot_i(U_m, e_m)$;
(2) $M, s \models m \to (m \vee l)$;

---

(3) $M, s \models \odot_i((U_m, e_m) \cup (U_l, e_l))$.

Formulas (1) and (2) are two premises which represent that the agent i ought to mail the letter ($\bigodot_i(U_m, e_m)$) and in the factual world s, if the letter is mailed, then it is mailed or burnt ($m \to (l \vee m)$). But we cannot derive that the agent i ought to mail the letter or burn it since the action of burning the letter should be captured by an independent action model (U_l, e_l) and the formula $\bigodot_i((U_m, e_m) \cup (U_l, e_l))$ is not true in s. As the axiom (RNC$\bigodot_i$) implies, both actions should be obligations themselves.

The above phenomenon is also in line with the free choice problem in the context of deontic sentences. Sentence (3) should be understood as that mailing the letter is obligatory and burning the letter is also obligatory, but since they cannot be fulfilled at the same moment, the agent can choose freely to fulfill one of them. Therefore, our conceptualization prevents the argument from validity and resolves Ross's paradox.

6.5.2 Weinberger's paradox and sequential composition

Tom's mom commands Tom to close the window and play the piano because if they do not close the window, playing the piano would cause noise and disturb their neighbors. Then Tom can give two arguments as follows:

(i)	Tom ought to close the window and play the piano.
(ii)	Tom ought to close the window.

and

(i)	Tom ought to close the window and play the piano.
(ii*)	Tom ought to play the piano.

From his mother's command, Tom ought to close the window and play the piano. Then by standard deontic logic, this obligation implies that Tom ought to close the window even if Tom is not playing the piano, which makes no sense since closing the window causes poor ventilation or it implies that Tom ought to play the piano while not closing the window, which is not good as it disturbs their neighbors. Standard deontic logic characterizes the two arguments as $\bigcirc(p \wedge q) \vdash \bigcirc p$ or $\bigcirc(p \wedge q) \vdash \bigcirc q$ and makes them valid.

However, based on our framework, the action of closing the window and playing the piano can be formalized by a sequential composition of two actions: $((U_w, e_w); (U_p, e_p))$ or $((U_p, e_p); (U_w, e_w))$. The command from Tom's mother claims that Tom ought to close the window *and* play the piano, which can be captured by the $\mathcal{L}_{\text{AKDL}}$-formula $\bigodot_{\text{Tom}}((U_w, e_w); (U_p, e_p))$ or the formula $\bigodot_{\text{Tom}}((U_p, e_p); (U_w, e_w))$. By the semantics, it means that after updating the original epistemic model by the sequential composition $(U_w \otimes U_p, (e_w, e_p))$ or $(U_p \otimes U_w, (e_p, e_w))$, a better epistemic model comes forth. But it does not imply that the action of closing the window can bring about a better epistemic model itself while not playing the piano, nor does it imply that the action of playing the piano is obligatory while not closing the window.

6.5.3 Penitent paradox

The penitent paradox is generally formalized as a valid formula in standard deontic logic: $F\phi \to F(\phi \wedge \psi)$ which means that if ϕ is forbidden, then ϕ and ψ is forbidden as well. Look at the following argument:

1)	It is forbidden to break your promise to go to your friend's party.
2)	It is forbidden to call an ambulance for a car accident victim and break your promise to go to the party.

It is totally unethical to go to the party while ignoring the victims of the car accident. So the action of calling an ambulance and breaking the promise to the party should not be forbidden. Therefore, the above argument is counter-intuitive.

Standard deontic logic defines prohibition as $F\phi =_{def} \bigcirc\neg\phi$, which means that if $\neg\phi$ is obligatory, ϕ is forbidden. So the penitent paradox can also be formalized as $\bigcirc\neg\phi \vdash \bigcirc(\neg\phi \vee \neg\psi)$ which is equivalent to Ross's paradox.

However, $\mathbb{AKDL}$ assigns independent status for the notion of prohibition which is not defined based on obligation. Hence, we can reformulate the penitent paradox over our framework and resolve it. See the following inference:

$$\frac{1)' \quad F_i(U_b, e_b)}{2)' \quad F_i((U_b, e_b); (U_c, e_c))}$$

Because the sequential composition $((U_b, e_b); (U_c, e_c))$ is a new action model which can be different from (U_b, e_b), it can make some original state better rather than making it worse. The inference above is not valid over our framework. Therefore, we have provided a way to accommodate the penitent paradox.

6.5.4 Forrester's paradox of gentle murder

Forrester's paradox of gentle murder involves both prohibition and obligation. Let us see the following argument:

i)	It is forbidden to break your neighbor's window.
ii)	If you break your neighbor's window, you ought to apologize for it.
iii)	If you apologize for breaking the window, you break your neighbor's window.
iv)	You break your neighbor's window.

The sentences i) - iv) intuitively should be consistent and we can imagine that there is a scenario making all four sentences true. However, they give rise to a conflict if we interpret it in standard deontic logic. See the formal inference below:

$$\begin{array}{ll} \text{i}') & Fb \\ \text{ii}') & b \rightarrow \bigcirc a \\ \text{iii}') & a \rightarrow b \\ \text{iv}') & b \\ \hline \text{v}') & Fb \wedge \bigcirc b \end{array}$$

Formula v′) is derived from formula i′)-iv′) by standard deontic logic. The conflict arises due to $\bigcirc a$ and $\bigcirc a \rightarrow \bigcirc b$. However, in our framework, all these sentences should be understood in the light of 'ought-to-do' rather than 'ought-to-be'. For example, the conditional ii) should be interpreted by dynamic operators rather than material implication since it indicates the obligation $\bigcirc(U_a, e_a)$ after performing the action (U_b, e_b). Similarly, sentence iii) should also be understood as the proposition b is the case after performing an action (U_a, e_a). Therefore, we can reformulate the argument as follows:

$$\begin{array}{ll} \text{i}'') & F_i(U_b, e_b) \\ \text{ii}'') & [U_b, e_b] \bigcirc (U_a, e_a) \\ \text{iii}'') & [U_a, e_a]b \\ \text{iv}'') & \langle U_b, e_b \rangle \top \end{array}$$

Now it is easy to find that formula i′′) - iv′′) are consistent in $\mathbb{AKDL}$. Forrester's paradox of gentle murder has been accommodated properly.

6.5.5 Paradox of derived obligation

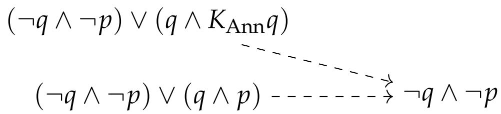
The formula $\bigcirc\phi \rightarrow \bigcirc(\psi \rightarrow \phi)$ is valid in standard deontic logic. This suggests that standard deontic logic can derive any conditional obligations whose consequence itself is an obligation, which gives rises to a lot of counter-intuitive instances. For example,

a.	I ought to keep my promise to go to my friend's party.
b.	It is obligatory that if my family members need help now, I still keep my promise to go to the party.

The argument above is nonsense since the state of affairs that my family members need help is more urgent and it defeats my obligation to keep my promise to the party. If we replace the formula with an equivalent form $\bigcirc\phi \rightarrow \bigcirc(\neg\psi \vee \phi)$, we find it is the same as the formulation shown in Ross's paradox. So we can resolve it as we did in Section 6.5.1

6.6 Formalization of Scenarios

In this section, we formalize the scenarios mentioned in Section 6.1. We use the following atomic propositions:



The best: Sam does not have diabetes and he is not injected with insulin. The second best: 1. Sam has diabetes and he is injected with insulin. 2. Sam has diabetes and Ann knows the fact.

Figure 6.4: $\mathcal{G}_1$ (Scenario 10 and 11)

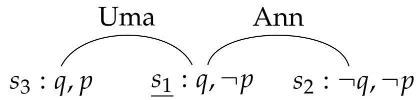


Figure 6.5: M_1 (Scenario 10)

q	: Sam has diabetes.
p	: Sam is injected with insulin.
r	: Sam is in the hospital.
h	: Sam is having a heart attack.
H	: The coin is heads up.
T	: The coin is tails up.
BH	: Chiyo bets heads.
BT	: Chiyo bets tails.
NG	: Chiyo does not gamble.

6.6.1 Scenario 10 and 11

We first give a priority structure $\mathcal{G}_1$ as the criterion for comparing models when we discuss Scenarios 10 and 11.

Priority structure $\mathcal{G}_1$ in Figure 6.4 tells us what the best state of affairs is and what the second best are. We give the initial epistemic model M_1 for Scenario 10. The reflexive relations are omitted.

We also give the initial epistemic model M_1 for Scenario 10. The reflexive relations are omitted. The factual state is s_1. The model M_1 shows the initial situation in Scenario 10 where Uma knows that Sam has diabetes but does not know whether Sam is injected with insulin, and Ann knows that Sam is not injected with insulin but she does not know whether Sam has diabetes. Then we will see what can we get after performing two different actions.

Let us first check the action that Ann injects with insulin. We give the preconditions and postconditions of the action model $(E_{\text{inject}}, e_{\text{inject}})$: $Pre(e_{\text{inject}})) = \top$ and $Post(e_{\text{inject}})(p) = \top$. The truth value of q remains the same in each state. Thus, if we want to check whether Ann ought to inject Sam with

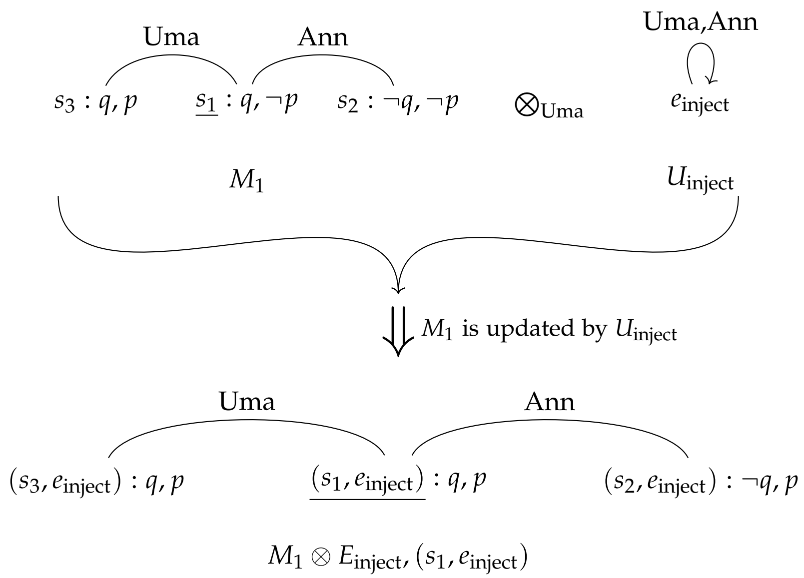


Figure 6.6: Scenario 10 (Ann injects insulin)

insulin, we need to compare $(M, s_1)|_{\sim_{\text{Ann}}}$ (consisting of s_1 and s_2) with $(M_1 \otimes E_{\text{inject}}, (s_1, e_{\text{inject}}))|_{\sim'_{\text{Ann}}}$ (consisting of (s_1, e_{inject}) and (s_2, e_{inject})) (see Figure 6.6). By Definition 73 and the priority structure $\mathcal{G}_1$, we have $s_1 <_{\mathcal{G}_1} (s_1, e_{\text{inject}})$ but $(s_2, e_{\text{inject}}) <_{\mathcal{G}_1} s_2$, which means that $(M, s_1) \not<^{\text{Ann}}_{\mathcal{G}} (M \otimes E_{\text{inject}}, (s_1, e_{\text{inject}}))$. In other words, Ann injecting Sam with insulin could cause a worse state. So we have $M_1, s_1 \not\models \odot_{\text{Ann}}(E_{\text{inject}}, e_{\text{inject}})$. Therefore, Ann does not have the knowledge-based obligation to inject insulin.

Now we check Uma's action of telling Ann about Sam. We give the preconditions and postconditions of action model $(E_{\text{tell}}, e_{\text{tell}}$: $pre(e_{\text{tell}}) = q$ and $Post(e_{\text{tell}})(p) = id$. After Uma tells Ann that Sam has diabetes, we get another updated model shown in the bottom row in Figure 6.7. Since $M_1 \otimes E_{\text{tell}}, (s_1, e_{\text{tell}}) \models K_{\text{Ann}} q$, we have $s_1 <_{\mathcal{G}_1} (s_1, e_{\text{tell}})$ and $s_3 <_{\mathcal{G}_1} (s_3, e_{\text{tell}})$. So if Uma tells Ann that Sam has diabetes, both possible states for Uma are updated to better states. Thus, by Definition 73, it follows that $(M, s_1) \not<^{\text{Uma}}_{\mathcal{G}} (M \otimes E_{\text{tell}}, (s_1, e_{\text{tell}}))$. By semantics, we conclude that $M_1, s_1 \models \odot_{\text{Uma}}(E_{\text{tell}}, e_{\text{tell}})$ which means that Uma ought to tell Ann the fact. This is also in line with our intuition.

In Scenario 11, Uma has told Ann that Sam has diabetes. We continue the analysis in Scenario 10 and Figure 6.7 where Uma tells Ann the fact. As mentioned above, Uma ought to tell Ann that Sam has diabetes. After being told the information, Ann ought to inject Sam with insulin. The updated model $(M_1 \otimes E_{\text{tell}}, (s_1, e_{\text{tell}}))$ shown in Figure 6.7 simulates the situation after Uma tells

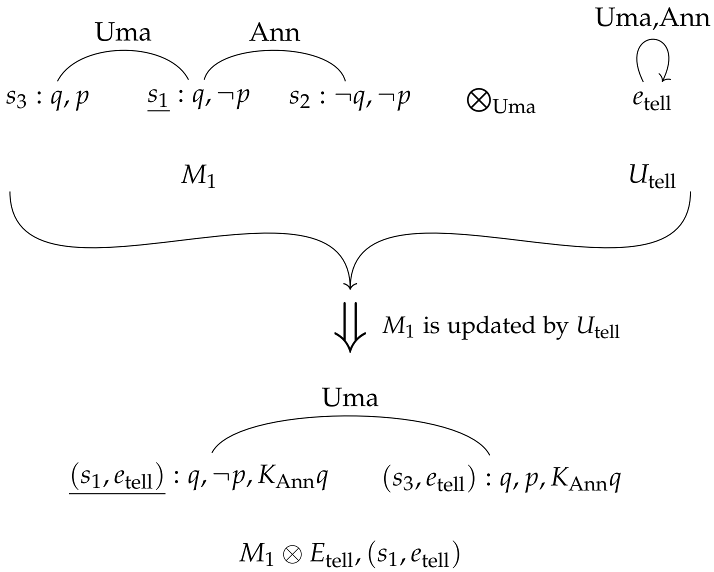


Figure 6.7: Scenario 10 (Uma tells Ann that Sam has diabetes)

Ann the fact.

We need to check the model updated by the action of injecting in $(M_1 \otimes E_{\text{tell}}, (s_1, e_{\text{tell}}))$. Let us see Figure 6.8. Since we only need to consider the states that Ann cannot distinguish from the current state, i.e., (s_1, e_{tell}), we get a new updated model $(M_1 \otimes E_{\text{tell}}) \otimes E_{\text{inject}}$ which only contains one state $((s_1, e_{\text{tell}}), e_{\text{inject}})$. According to the priority structure $\mathcal{G}_1$, we have $(s_1, e_{\text{tell}}) <_{\mathcal{G}_1} ((s_1, e_{\text{tell}}), e_{\text{inject}})$. Thus, we have $M_1 \otimes E_{\text{tell}}, (s_1, e_{\text{tell}}) \models \odot_{\text{Ann}}(E_{\text{inject}}, e_{\text{inject}})$.

6.6.2 Scenario 12

Scenario 12 is given by Horty (2019) to introduce the notion of epistemic ought, which is actually identical to the notion of knowledge-based obligation. We will use our knowledge-based 'ought-to-do' obligations to simulate Scenario 12.

The priority structure $\mathcal{G}_2$ used in Scenario 12 is shown in Figure 6.9.

The initial epistemic model M_3 of Scenario 12 shows the situation where the coin is tossed but Chiyo does not know whether the coin lands heads up or tails up ($\neg K_{\text{Chiyo}} T \wedge \neg K_{\text{Chiyo}} H$). At the moment, Chiyo has not decided whether to bet.

Now we give the process of betting on heads in Figure 6.11. The case of betting on tails is similar. The action model of betting on heads is $(U_{\text{BetH}}, e_{\text{BetH}})$. The preconditions and postconditions are: $pre(e_{\text{BetH}}) = \top$, $post(e_{\text{BetH}})(NG) = \bot$, $post(e_{\text{BetH}})(BT) = \bot$, and $post(e_{\text{BetH}})(BH) = \top$.

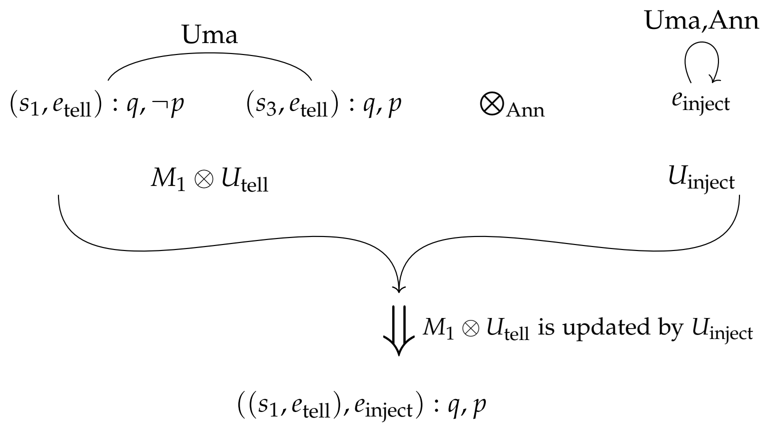


Figure 6.8: Scenario 11 (After Uma tells Ann the fact, Ann injects Sam with insulin)

$$(H \wedge BH) \vee (T \wedge BT)$$

$$\uparrow$$

$$(H \wedge BH) \vee (T \wedge BT) \vee (H \wedge NG) \vee (T \wedge NG)$$

The best: Chiyo bets correctly.
The second best: Chiyo does not bet.
The worst: Chiyo bets incorrectly.

Figure 6.9: $\mathcal{G}_2$ (Scenario 12)

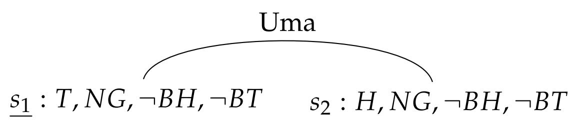


Figure 6.10: M_3 (Scenario 12)

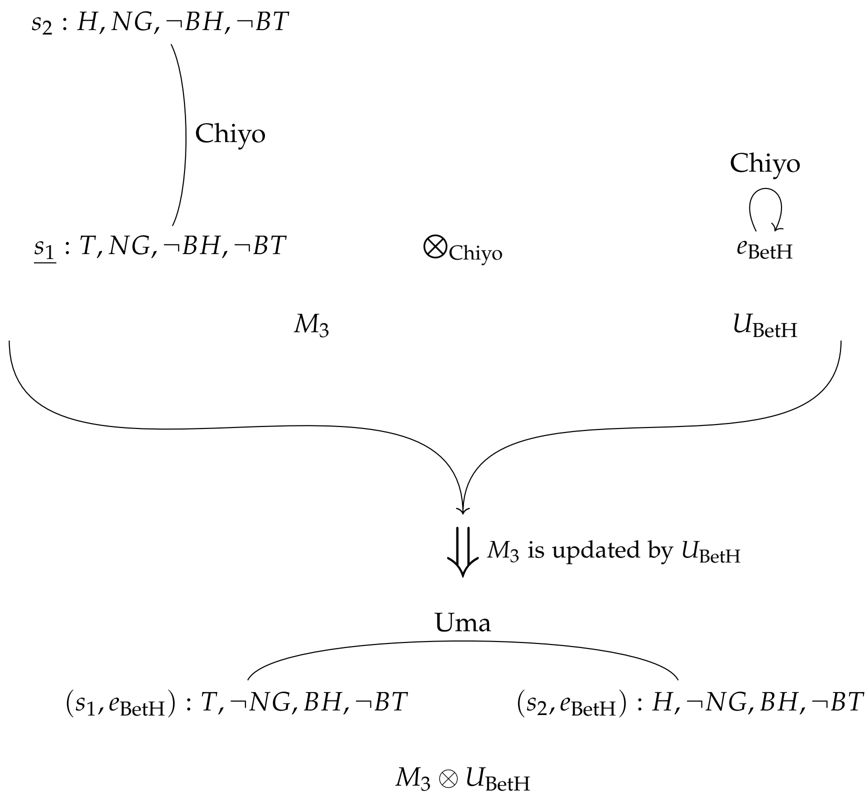


Figure 6.11: Scenario 13 (Uma bets heads)

$$\neg h \wedge (K_{\text{Uma}} h \vee K_{\text{Uma}} \neg h)$$
$$\uparrow$$
$$\neg h \wedge \neg(K_{\text{Uma}} h \vee K_{\text{Uma}} \neg h)$$
$$\uparrow$$
$$h \wedge (K_{\text{Uma}} h \vee K_{\text{Uma}} \neg h)$$
$$\uparrow$$
$$h \wedge \neg(K_{\text{Uma}} h \vee K_{\text{Uma}} \neg h)$$

The best: Sam is not having a heart attack and Uma knows Sam's health state.
The second best: Sam is not having a heart attack and
Uma does not know Sam's health state.
The third best: Sam is having a heart attack and Uma knows Sam's health state.
The worst: Sam is having a heart attack and Uma does not know Sam's health state.

Figure 6.12: $\mathcal{G}_3$ (Scenario 13)

Uma

$\underline{s_1} : h, \neg(K_{\text{Uma}} h \vee K_{\text{Uma}} \neg h)$ $s_2 : \neg h, \neg(K_{\text{Uma}} h \vee K_{\text{Uma}} \neg h)$

Figure 6.13: M_4 (Scenario 13)

After betting on heads, there are two possible states for Chiyo: (s_1, e_{BetH}) and (s_2, e_{BetH}). According to the priority structure $\mathcal{G}_3$, $(s_1, e_{\text{tell}}) <_{\mathcal{G}_2} s_1$ and $s_2 <_{\mathcal{G}_2} (s_2, e_{\text{BetH}})$. This means that betting on heads could lead to a worse state. Then, we have $M_3, s_1 \not\models \odot_{\text{Chiyo}}(U_{\text{BetH}}, e_{\text{BetH}})$.

6.6.3 Scenario 13

'Ought to know something' is also called epistemic obligation in philosophy. We assume that, in Scenario 13, Uma knows that Sam is a patient at her practice and Sam is living in the sickroom of Uma's hospital. 'Uma knows Sam's health state' is expressed by 'Uma knows whether Sam is having a heart attack'. It should be Uma's obligation. The priority structure $\mathcal{G}_3$ in Figure 6.12 is used for Scenario 13.

The initial situation is modelled by M_4 in Figure 6.13, where Uma does not know whether Sam is having a heart attack.

Then we give the dynamic process of Uma's asking about Sam's health state

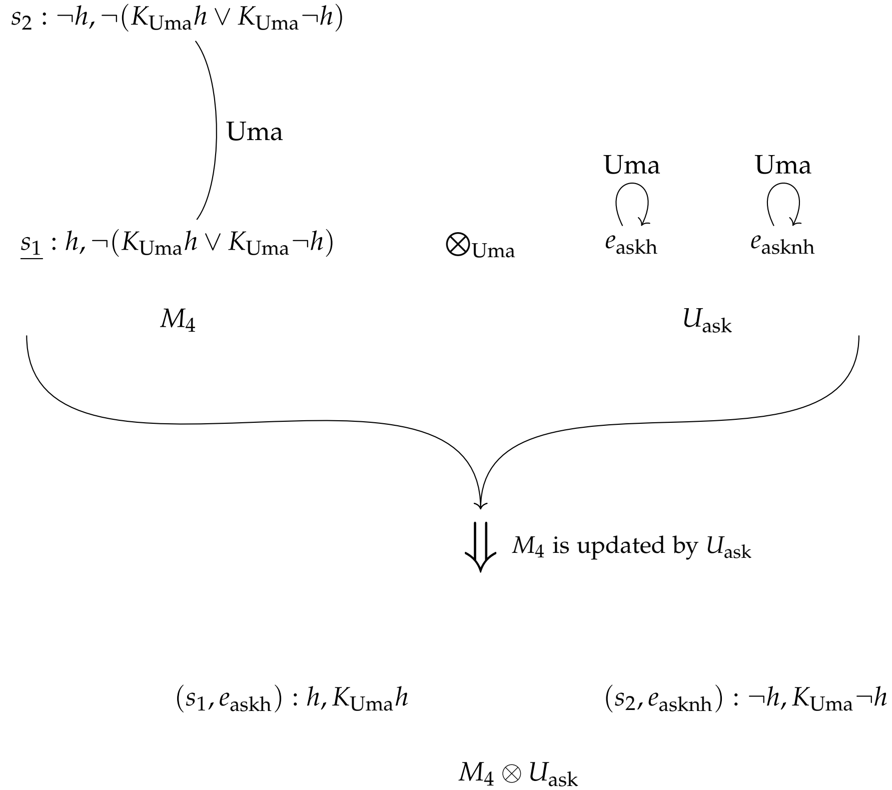


Figure 6.14: Scenario 13 (Uma asks whether Sam is having a heart attack)

in Figure 6.14.

The action model U_{ask} contains two events which represent the action of asking that h is the case and the action of asking that h is not the case. The preconditions are $pre(e_{\text{askh}}) = h$ and $pre(e_{\text{asknh}}) = \neg h$. The postconditions for both events and any formulas are *id*.

According to the priority structure $\mathcal{G}_3$ in Figure 6.12, we have $s_1 <_{\mathcal{G}_3} (s_1, e_{\text{askh}})$. That also implies that $(M_4, s_1) <^i_{\mathcal{G}_3} (M_4 \otimes U_{\text{ask}}, (s_1, e_{\text{askh}}))$. Similarly, it holds that $s_2 <_{\mathcal{G}_3} (s_2, e_{\text{asknh}})$ which also implies that $(M_4, s_2) <^i_{\mathcal{G}_3} (M_4 \otimes U_{\text{ask}}, (s_2, e_{\text{asknh}}))$. Thus, the action of asking about Sam's health state always leads to some better states. So we have $M_4, s_1 \models \odot_{\text{Uma}}(U_{\text{ask}}, e_{\text{askh}})$.

6.7 Discussion and conclusion

In this chapter, we conceptualized the notion of knowledge-based 'ought-to-do' obligation based on dynamic epistemic logic. The basic idea of defining the notion is that an action ought to be done if and only if the action always improves the initial situation to better situations. As illustrated in Section 6.3, this definition differs from the classical definition of obligations which requires that obligations must lead to the optimal or the most ideal situations. On the other hand, in the previous research on the notion of knowledge-based obligations (see Section 1.4), each agent knows their own knowledge-based obligations. However, the knowledge-based 'ought-to-do' obligations defined in this chapter might not be known by the agents. These are two main differences between our approach and the classical one. Even so, our approach and the classical approach are actually not that different. We can transfer our approach to the classical one by adding some constraints.

Restrictions to the number of actions However, by restricting the number of actions available for each agent, there are only finitely many different consequence situations and therefore we can compare them to find the most ideal consequence situations. Following this restriction, we can define obligations in Meyer's approach. For example, given an epistemic model (M, s) as the initial situation, an agent i has only three actions available for him/her: α_1, α_2, and α_3 (See Figure 6.15). The action α_1 can lead to a better situation, an incomparable situation and a worse situation. So according to our definition of knowledge-based 'ought-to-do' obligations in this chapter, α_1 is not an obligation. The actions α_2 and α_3 always lead to better situations. Thus, they are obligations according to our definition.

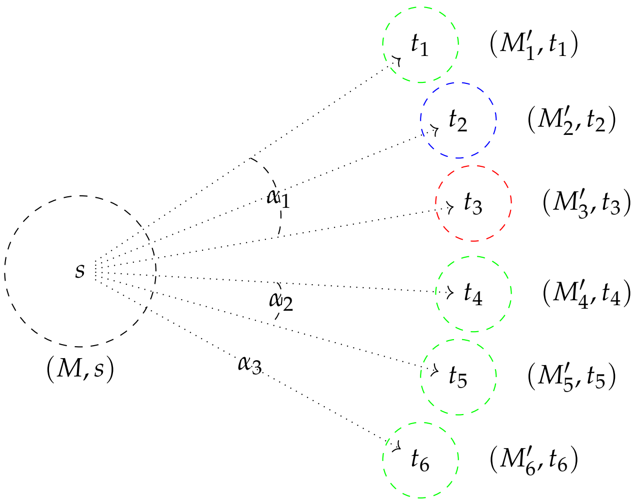
However, if we can compare these green updated models which are all better (than (M, s)) updated models, we can find the most ideal situations. Then according to Meyer's definition: an action ought to be done if and only if it always leads to the most ideal situations, we can tell which actions of α_1, α_2 and α_3 are obligatory. Consider the following cases:

1. If (M_1', t_1) is the best and it is better than other green models, then no action is the obligation since no action can always lead to the most ideal situation.

2. If (M_4', t_4) and (M_5', t_5) are the most ideal, then α_2 is the obligation.

3. If (M_6', t_6) are the most ideal, then α_3 is the obligation.

It is easy to understand the case 2 and 3. Case 1 deserves further discussions. If (M_1', t_1) is the unique optimal updated model, it implies that the action α_1 can lead to the best situation and can also lead to the worst situation (M_3', t_3). It is not a safe choice for the agent to perform α_1. In contrast, performing α_2



Green: the updated model is better than the initial model. blue: the updated model is neither better nor worse than the initial model. red: the updated model is worse than the initial model.

Figure 6.15

or α_3 always brings about some better situations, which means that it is safe to perform these two actions. In this sense, we can also say it is obligatory to perform α_2 or α_3 since they are the safest choices when the situation is (M, s). It again justifies the definition of knowledge-based 'ought-do-do' obligations provided in this chapter which is based on 'better consequences' rather than 'the optimal consequences'.

Knowing obligations in the single-agent case As mentioned in Section 6.4, one of the differences between knowledge-based 'ought-to-do' obligations studied in this chapter and other characterizations of knowledge-based obligations is that the formula $\odot_i \alpha \rightarrow K_i \odot_i \alpha$ is not valid. This means that an agent might not know his/her own knowledge-based 'ought-to-do' obligations. However, this difference can be removed by adding the following two constraints on the action models:

1. There is only one agent in the group, i.e., $G = \{i\}$;
2. The epistemic relation R_i in the action model is connected.

The two constraints imply that for each action model, the agent cannot epistemically distinguish any events from the actual event in the action model.

In other words, given an action model U, if $e_1 \in U$ and $e_2 \in U$, the agent considers the pointed model (U, e_1) and the pointed model (U, e_2) to be the same action. Thus, in the following part, action models, instead of pointed action models will be the object of the deontic operator. Let (M, s) be an arbitrary pointed epistemic model and let $dom(M)$ represent the domain of states of the epistemic model M. Given a priority structure $\mathcal{G}$,

$M, s \models \odot_i U$ iff (1) for each t with $s \sim_i t$, there is an $e \in U$ such that $(t, e) \in dom((M, s)|_{\sim_i} \otimes U)$;
(2) $(M, s)|_{\sim_i} < (M, s)|_{\sim_i} \otimes U$.

where

$(M, s)|_{\sim_i} < (M, s)|_{\sim_i} \otimes U$ iff (1′) for each $t \in dom((M, s)|_{\sim_i})$, if there is an $e \in U$ such that $(t, e) \in dom((M, s)|_{\sim_i} \otimes U)$, then $t \leqslant_{\mathcal{G}} (t, e)$;
(2′) there is a $t \in dom((M, s)|_{\sim_i})$ and there is an $e \in U$ such that $(t, e) \in dom((M, s)|_{\sim_i} \otimes U)$ and $t <_{\mathcal{G}} (t, e)$.

The condition (1) means that the action U is executable on all the states that the agent i cannot epistemically distinguish from the actual state s. In terms of the condition (2), it must satisfy the conditions (1′) and (2′). The condition (1′) represents that for all updated states, they are not worse than their original states. The condition (2′) means that there at least exists one original state such that it is updated to a strictly better state.

It is worth noting that if $M, s \models \odot_i U$, then for each t such that $t \sim_i s$, it also holds that $M, t \models \odot_i U$ according to the above semantics of $\odot_i U$. Let me explain this. Firstly, the action model U is executable on all the states that the agent i cannot epistemically distinguish from the actual state. Secondly, we have $(M, s)|_{\sim_i} = (M, t)|_{\sim_i}$ and $(M, s)|_{\sim_i} \otimes U = (M, t)|_{\sim_i} \otimes U$. Therefore, we can conclude that $M, s \models K_i \odot_i U$, which means that the knowledge-based 'ought-to-do' obligations must be known by the agent.

Conclusion This chapter conceptualized the notion of knowledge-based 'ought-to-do' obligation based on dynamic epistemic logic. We also extended the action models introduced in Section 2.5 by non-deterministic choice and sequential composition between actions in order to investigate obligations involving compound actions. Then a method for comparing states and different epistemic models was provided based on priority structures. In Section 6.4, we established the logic of knowledge-based 'ought-to-do' and the axiomatization $\mathbb{AKDL}$ was proved to be sound and strongly complete via new reduction axioms for obligation and prohibition. Section 6.5 accommodated several influential paradoxes of deontic logic in $\mathbb{AKDL}$. In the end, we formalized the four scenarios mentioned in Section 6.1.

Chapter 7

Conclusion

This thesis provides two novel definitions of the notion of knowledge-based obligation and also studies the logic of obligation change due to three types of change. The first is epistemic change, which means that the updates on the information of agents would also update their obligations. The second is factual change. It represents that those actions that change facts can also bring about new obligations. The last is norm change, which represents that if the norms that agents are following are updated, then their obligations would be changed accordingly. Different logic systems are established for all the logics mentioned above. We do these work by combining epistemic logic, dynamic epistemic logic, preference logic and deontic logic altogether. Therefore, a comprehensive and coherent framework is provided. This thesis contributes to the systematic research on the problem how knowledge affects agents' obligations in the field of deontic logic. In the remaining part of this chapter, we first look back at two main approaches of characterizing knowledge-based obligations in this thesis. The second section shows how we capture dynamic obligations with respect to different updates. The third section reviews two main techniques used in the strong completeness proofs. The last section shows how the research of this thesis can be developed in the future.

7.1 Knowledge-based obligations

The core notion studied by this thesis is knowledge-based obligations. Although there have been several formalism established for the notion, such as Pacuit et al. (2006), Broersen (2008), Horty (2019), this thesis provides two new frameworks for capturing knowledge-based obligations: Hansson's conditional style of knowledge-based 'ought-to-be' obligations and knowledge-based 'ought-to-do' obligations.

7.1.1 Knowledge-based 'ought-to-be' obligations

The first characterization of knowledge-based obligations follows the tradition of 'ought-to-be' obligations. Moreover, the formalism is based on conditional settings and, as a result, the relation between agents' knowledge and their obligations is clearly addressed by the principle of epistemic detachment: $\models_{\mathbb{KCDL}} (\bigodot_i(\phi|\psi) \wedge K_i\psi) \rightarrow \bigodot_i(\phi|\top)$. According to our logic of knowledge-based conditional obligations $\mathbb{KCDL}$, an agent's knowledge triggers his/her own knowledge-based conditional obligations. An important characteristic of our formalism is that every agent knows all their own knowledge-based conditional obligations. So their unconditional obligations depend on what they know.

In terms of the semantic apparatus, this thesis extends Hansson's betterness structures with epistemic relations. Rather than considering all possible worlds, we only focus on those states that the agent cannot distinguish from the real state. The agent i has a knowledge-based conditional obligation $\bigodot_i(\phi|\psi)$ if and only if all the best ψ-worlds that i cannot epistemically distinguish from the real state satisfy ϕ (see Definition 28). With the constraint on the set of epistemically indistinguishable states, the notion of knowledge-based conditional obligations differs from Hansson's objective conditional obligations in the sense that it takes agents' epistemic information into account.

7.1.2 Knowledge-based 'ought-to-do' obligations

The second characterization of knowledge-based obligations applies deontic operators to actions rather than propositions and therefore a new definition of knowledge-based 'ought-to-do' obligations is given. There are two factors motivating us to develop an 'ought-to-do' style of knowledge-based obligations: (1) in previous research on knowledge-based obligation, actions play fundamental roles and their basic ideas are similar: the obligatory action is good, which is also known by the agent; (2) there is one common counter-intuitive result in many 'ought-to-be' style deontic logics, i.e., if one situation is the case, it ought to be the case. An 'ought-to-do' tuning resolves several problems left by the 'ought-to-be' approach.

Dynamic epistemic logic provides the basic formal framework for our knowledge-based 'ought-to-do' obligations. Each action is represented by an action model or a non-deterministic choice between several action models, which gives us a lot of control on expressing how an action updates a situation. An action model transfers an epistemic model to a new epistemic model. An action model is obligated with respect to some agent's knowledge if and only if the action model updates the initial epistemic model to a better one (see Definition 75). It is worth noting that our 'ought-to-do' obligations are defined based on 'better' rather than 'best'. It characterizes a type of safe obligations that agents can perform given some certain situation.

Priority structures (see Definition 15) are criterion on assessing whether a state in the initial epistemic model is updated to a better one. Given an $\mathbb{AKDL}$ axiom system, the priority structure is fixed and all updates on epistemic models are judged by the priority structure.

Knowledge-based 'ought-to-do' obligations are 'knowledge-based' in the sense that we only need to consider whether each state that one agent cannot epistemically distinguish from the actual state is improved to a better state. The obligations are determined by agents' initial knowledge and also the knowledge after doing the action.

7.2 Making static deontic logic dynamic

In our first characterization of knowledge-based obligations in 'ought-to-be' style, all obligations are static since they never change once an epistemic betterness structure is given. However, various factors can change obligations in our real life. This thesis mainly investigates three factors: information change, factual change and norm change.

7.2.1 Dynamic obligations due to information/factual change

The logic of knowledge-based conditional obligations studied in Chapter 3 shows that knowledge triggers agents' knowledge-based conditional obligations. A follow-up question is: how do an agent's obligations change if his/her knowledge changes? As mentioned above, we combine Hansson's betterness structures with epistemic relations and knowledge can be defined as the standard epistemic logic did. Dynamic epistemic logic has been widely used for capturing information and factual change for decades, where action models build the bridge between the initial models and the updated models. Therefore, action models can lead from an initial epistemic betterness structure to an updated epistemic betterness structure. During this process, facts (truth values of propositions on some states) and information (epistemic relations) are changed by an action model. As a consequence, the betterness relation is changed as well since it is determined by a given priority structure and truth values of epistemic logic formulas. Priority structures play the role of assessing which states of affairs are better or worse. Given a priority structure, it remains the same throughout.

Therefore, alongside the information change, agents would obtain some new obligations and some obligations might be defeated. Based on the dynamic version of epistemic betterness structures, several philosophical notions of obligations can be formalized over the novel framework. A prima facie obligation (see Definition 45) is an unconditional obligations when the agent knows something at a particular moment. All-things-considered obligations (see Definition 46) of an agent are those unconditional obligations when the agent knows everything. Safe knowledge-based obligations (see Definition 47) are the unconditional obligations whatever the agent knows.

In terms of semantics, we use a combination of dynamic epistemic logic and Hansson's betterness structures. Dynamic epistemic logic is used for characterizing information and factual change. Obligations change due to the updates on betterness relations which is controlled by priority structures.

7.2.2 Dynamic obligations due to norm change

In addition to epistemic and factual changes, norm change can also update an agent's obligations. The distinction between norms and obligations is known as prescription and description in deontic logic. A normative system is used for prescribing new norms, which could bring about new obligations to agents. So updates on normative systems generally result in obligation change.

Chapter 5 mainly discusses the problem of how a normative system, namely an ideality sequence (see Definition 54) in our terminology, affects one agent's obligations. The core notion is 'relativized conditional obligations based on ideality sequences', i.e., $\bigcirc_{\mathcal{I}}(\_|\_)$ (see Definition 64). It is formalized very similarly to Hansson's dyadic obligation operator except that it is relativized to ideality sequences. Every ideality sequence $\mathcal{I}$ is a normative system, which represents the prescriptive aspect of our framework. And betterness structures based on some ideality sequence describe the conditional obligations.

Two elementary updates on ideality sequences are introduced: deletion and postfixing. Deletion corresponds to abolishing a norm and postfixing is relevant to the notion of derogation in law. We showed how these updates on ideality sequences affect obligations.

Our logic of relativized conditional obligations $\mathbb{PCDL}$ resolves a famous philosophical problem in inferences between norms: Jørgensen's dilemma. It provides a novel conceptualization on the dilemma. A 'valid' inference between norms should preserve the success of the premises and the set of obligations brought about by the concluded norm is a subset of the set of obligations brought about by the premise norms.

7.3 Technical contributions

This thesis establishes several axiom systems for the notion of knowledge-based obligations: $\mathbb{KCDL}$, $\mathbb{DKCDL}$, $\mathbb{PCDL}$, and $\mathbb{AKDL}$. These systems are proved to be sound and strongly complete with respect to different classes of models. The strong completeness proofs involve different techniques, such as canonical models, the 'step-by-step' method, and the reduction axiom method. Next we will review two important methods used in these proofs.

7.3.1 The 'step-by-step' method

Deontic logics in Hansson's style investigate dyadic obligation operators. Their original method for completeness proof is using canonical models where each possible world is a witness of all obligations under a certain condition. However, this method does not work in the multi-agent case. In Chapter 3, the method of 'step-by-step' constructs a model which satisfies each consistent set of $\mathcal{L}_{\text{KCDL}}$-formulas. This model does not include all maximal consistent sets of formulas. Rather, it starts with an initial set and builds up a bigger model stepwise with those sets needed to make it perfect.

In the light of the 'step-by-step' method, we proved that the axiomatization $\mathbb{KCDL}$ is strongly complete with respect to the class of epistemic betterness structures. Since an epistemic betterness structure is a generalization of Hansson's betterness structure, the completeness result for $\mathbb{KCDL}$ (see Theorem 2) gives an answer to the open question left by Parent (2014) for Hansson's system **F**+(CM) (see Corollary 1).

7.3.2 The Kangerian-Andersonian reduction

The Kangerian-Andersonian reduction was originally put forward to show that the classical deontic operators can be reduced to classical modal operators, i.e., $\Box$ and proposition constants. However, Chapter 4 and 5 not only find the Kangerian-Andersonian reductions to show knowledge-based conditional obligations (see Proposition 3) and relativized conditional obligations (see Proposition 9) can be reduced to the universal operators, but also use them as the main axioms to prove the strong completeness of $\mathbb{DKCDL}$ (see Theorem 4) and $\mathbb{PCDL}$ (see Theorem 7). By Kangerian-Andersonian reduction axioms, we can reduce the strong completeness proofs to the classical strong completeness proof for epistemic logic $\mathcal{S}5$.

7.4 Future research

This thesis has investigated some problems of knowledge-based obligations. We list several possible directions and considerations that can extend our research in the future.

7.4.1 Collective obligations and group knowledge

In many other branches of philosophical logic, it is common to lift single-agent notions to multi-agent notions. For example, the operator K_i in epistemic logic represents one agent's knowledge, whereas there are also D_G, E_G, and C_G to represent distributed knowledge, mutual knowledge and common knowledge of a group of agents. Similarly, in deontic logic, there is also a notion of obligation for a group of agents so that the group, as a whole, is obliged to accomplish a given task. It is called collective obligations. Although this thesis does not involve collective obligations, this notion has been studied for decades in both deontic logic and moral philosophy. When a group task is assigned, each member should bear their own parts and hence gets an individual obligation. *How a collective obligation distributes over a group* is a question that moral philosophers, political philosophers and deontic logicians must confront.

In the field of deontic logic, there have been many frameworks for conceptualizing the notion of collective obligations. Carmo and Pacheco (2000) investigated the notion of collective agency by a new deontic and action logic where a group of agents is treated as an institutionalized agent. Grossi et al. (2004) also conceptualized the collective obligations by the notion of plan and task allocation in computer science and addressed the problem which member should be responsible for the failure of their group task. Tamminga and Duijf (2017) studied the conditions under which the collective obligations coordinate individual obligations based on game-theoretical frameworks. Based on the bringing-it-about modality, Porello (2018) provided new modalities to capture three types of group norms which in turn induce three types of collective obligations. Duijf and Van De Putte (2021) studied the cases where a group makes a collective decision, but no individual member of the group can be held responsible for this decision. The most relevant work to our thesis is from Cholvy and Garion (2002). They extended Boutilier's conditional obligations to a group notion. A new formula $I(\phi|IU(KB))$ was provided where $IU(KB)$ represents the knowledge base of a given set of agents and the formula can be read as 'the group has an obligation of ϕ based on their knowledge base'.

In philosophy, we can see more discussions suggesting the notion of group knowledge to play important roles in distributing collective obligations to individual obligations. We list several claims about this issue from philosophical literature:

1. (simplest) Members are *only* obligated to do their own part, provided others are doing theirs.

2. (Lawford-Smith (2012) 1) A member of the collective is obligated to take a capacity-relative share in joint work when given a *belief* that the other members would do the same.

3. (Lawford-Smith (2012) 2) When a collective has an obligation to see to it that ϕ, every individual member of the collective has an obligation to take a capacity-relative share in fulfilling the obligation, unless she has the reasonable belief that at least one other member of the collective will not take a capacity-relative share in fulfilling the obligation.

4. (Aas (2015)) Every member not only does their parts under the condition that they believe others will do their parts as well, but also before they fulfill this collective obligation, they should contribute another collective

action – changing the way they are organized – to make themselves form a real organization.

As far as we found, we do not see deontic logic which aims to investigate the notion of knowledge-based collective obligations. However, as the above philosophers argued, group knowledge indeed affects the collective obligations and individual obligations. Inspired by epistemic logic where the group notions of knowledge are developed based on the single-agent knowledge, we can define three different knowledge-based collective conditional obligations based on our logic of knowledge-based conditional obligations in Chapter 3.

Given a set of agents $G = \{1, 2, \cdots, n\}$, let $M = \langle S, \leqslant, \sim_1, \sim_2, \cdots, \sim_n, V\rangle$ be an epistemic betterness structure. Let $\sim_\cap = \bigcap_{i \in G} \sim_i$, $\sim_\cup = \bigcup_{i \in G} \sim_i$, and let $\twoheadrightarrow$ be the reflexive and transitive closure of $\sim_\cup$.

1. D-Epistemic Collective Obligation $\odot_D(\phi|\psi)$:

$$M, s \models \odot_D(\phi|\psi) \text{ iff } max_{\leqslant|[s]^{\sim_\cap}} \|\psi\|_M \subseteq \|\phi\|_M.$$

   The truth of $\odot_D(\phi|\psi)$ depends on the best ψ-worlds over these possible worlds that can decide the distributed knowledge of G.

2. M-Epistemic Collective Obligation $\odot_E(\phi|\psi)$:

$$M, s \models \odot_E(\phi|\psi) \text{ iff } max_{\leqslant|[s]^{\sim_\cup}} \|\psi\|_M \subseteq \|\phi\|_M.$$

   The truth of $\odot_E(\phi|\psi)$ depends on the best ψ-worlds over these possible worlds that can decide the mutual knowledge of G.

3. C-Epistemic Collective Obligation $\odot_C(\phi|\psi)$:

$$M, s \models \odot_C(\phi|\psi) \text{ iff } max_{\leqslant|[s]^{\twoheadrightarrow}} \|\psi\|_M \subseteq \|\phi\|_M.$$

   The truth of $\odot_C(\phi|\psi)$ depends on the best ψ-worlds over these possible worlds that can decide the common knowledge of G.

Do the above definitions characterize the notion of knowledge-based collective obligations properly? How do they relate to the knowledge-based individual obligations provided in this thesis? What is the axiomatization for these collective obligations? They must be interesting questions for future research.

7.4.2 From Individual Obligations to Social Obligations

Although $\mathbb{KCDL}$ defines epistemic conditional obligations in the multi-agent case, it does not formalize any obligations that are essentially related to multiple agents. For example, Matteo is a soldier in Cesar's troop. When war breaks out, Matteo has an obligation to Cesar to obey the commands given by Cesar. The obligation of Matteo is directed to his commander Cesar. We can call this type of obligation a 'social obligation' which is mostly inspired by the notion of social commitments given by Castelfranchi (1995). It could be captured by introducing a new operator $\odot_{(i,j)}(\phi|\psi)$ which intuitively means that i has an obligation to j to see to it that ϕ under the condition that ψ.

This type of obligations are relevant to social commitment. As Gilbert (1999) argued, obligations between two agents are generated from their joint commitments. Dignum et al. (1996), earlier than Gilbert, had provided a formal characterization on actions of intention and commitment which changes agent's obligations. Cholvy and Garion (2002) also used a notion of individual commitment to define individual obligations to a collective obligations. Dunin-Kęplicz and Verbrugge (2004), inspired by Castelfranchi (1995), defined social commitment as: $COMM(i,j,\alpha) \leftrightarrow INT(i,\alpha) \wedge GOAL(j, done(i,\alpha)) \wedge CBEL_{\{i,j\}}(INT(i,\alpha) \wedge GOAL(j, done(i,\alpha)))$, which intuitively means that i commits to j to do plan α if and only if i has the intention to do α and j should be interested in i fulfilling its intention and they commonly believe the above conditions. In terms of speech acts, commanding, requesting, committing, etc. would affect the agent's obligations. Yamada (2006) established a multi-agent epistemic deontic Logic $\mathbb{MEDL}$ where the formula $[Com_{(i,j)}\phi]K_j \bigcirc_{(j,i,i)} \phi$ means that after i commands j that ϕ, j will know that j has an obligation to i that ϕ in the name of i.

Considering the strong relevance between social obligations and social commitments mentioned above (also witness Dignum and Royakkers (1998), Royakkers and Dignum (1999)), we may give a characterization of $\odot_{(i,j)}(\phi|\psi)$ based on Dunin-Kęplicz and Verbrugge (2004)'s idea:

$$\odot_{(i,j)}(\phi|\psi) \leftrightarrow \odot_i(\phi|\psi) \wedge K_j \odot_i(\phi|\psi) \wedge \odot_j(\odot_i(\phi|\top)|\psi)$$

Moreover, one agent as an authority can even change other agents' obligations. In the context of normative systems, Herzig et al. (2011) formalized *authorization* with a formula $P_i(+P_j(+\phi))$ which intuitively means that agent i has the permission to authorize agent j to achieve ϕ.

7.4.3 Beyond consequentialism

In this thesis, we introduced several deontic logics which define obligations based on actions. In Meyer (1988)'s approach, an action ought to be done if and only if all the consequence worlds that it leads to are ideal. Deontic stit logic tries to find optimal actions by comparing which action brings about the most ideal histories. In Chapter 6, we follow the similar idea that if it is obligatory to fulfill an action, then this action must lead to more ideal situations. It is worth noting that all these approaches follow the principle of utilitarianism or consequentialism, because they all define obligations as actions which lead to good consequences.

As mentioned in Section 2.1, there are other ethical theories, such as deontological theory (see Kant (2002, 2005)), social contractarianism (see Hobbes (1914), Rousseau (1795)), and contractualism (see Scanlon et al. (1998)) besides consequentialism. Deontology is the main foil of consequentialism. It posits that whether an action is our obligation cannot be justified by its effects. On the contrary, it depends on whether the action complies with a moral norm.

If we plan to construct a deontic logic based on deontology, our criteria on assessing actions should no longer be comparing their consequences. Instead, we need to judge whether an action complies with given moral norms. But there are some foreseeable problems. For example, how to define the inference between two obligations in this sense? If an action complies with a moral norm, which also means that the action is obligatory, how to derive another action which also follows the moral norm? It seems that this problem would lead to reducing the inference between obligations to the inference between moral norms, which is not what we really expect. So how to establish a deontic logic without resorting to consequentialism is a question to be resolved.

There are still many other interesting topics involving both deontic logic and knowledge. For example, the interactions between permission and knowledge deserve special attention as well. You are permitted to operate a machine in a factory only if you know how to use it. Lots of possible directions and questions need our further research.

Summary

This thesis investigates the problem of knowledge-based obligations and how knowledge affects obligations. Chapter 3 provides our first static logic of knowledge-based conditional obligations. Chapter 4 and 5 study how obligations change due to information, factual and norm change. Finally, we characterize a new notion of knowledge-based 'ought-to-do' obligations and its axiomatization is established.

We first introduce the notion of knowledge-based conditional obligations which is defined over epistemic betterness structures. This definition addresses the relation between agents' knowledge and obligations, which is indicated by a very important theorem in our research: the epistemic detachment, i.e., $\models_{\mathbb{KCDL}} (\odot_i(\phi|\psi) \wedge K_i\psi) \rightarrow \odot_i(\phi|\top)$. Therefore, we can say a knowledge-based conditional obligation pre-encodes what obligation an agent has if the agent knows some information.

In the semantic aspect, epistemic relations for each agent have been added to betterness structures to construct epistemic betterness structures. Rather than considering all possible worlds, we only focus on those states that the agent cannot epistemically distinguish from the real state. The agent i has knowledge-based conditional obligation $\odot_i(\phi|\psi)$ if and only if all the best ψ-worlds that i cannot distinguish from the real state satisfy ϕ. Accordingly, we establish the logic of knowledge-based conditional obligations and provide a sound and strongly complete axiomatization $\mathbb{KCDL}$ following the 'step-by-step' method rather than the classical approach of canonical models. Since an epistemic betterness structure is a generalization of Hansson's betterness structure, the completeness result for $\mathbb{KCDL}$ gives an answer to the open question left by Parent (2014) for Hansson's system $\mathbf{F}$+(CM).

Then the static logic of knowledge-based conditional obligations is extended by dynamic operators on epistemic and factual change. We therefore are able to characterize obligation change due to epistemic or factual change. Specifically, dynamic epistemic logic was introduced for characterizing these updates. Epistemic relations in an epistemic betterness structure are changed by action models and therefore knowledge-based conditional obligations are changed. Moreover, the action models used in this thesis also include the function of factual change which can update the truth values of propositions on each state.

Factual change would also change the betterness relation based on one given priority structure.

According to the characterization of dynamic information, several philosophical notions of obligations can be formalized over the novel framework. A *prima facie* obligation is an unconditional obligations when the agent knows something at a particular moment. All-things-considered obligations of an agent are those unconditional obligations when the agent knows everything. Safe knowledge-based obligations are the unconditional obligations whatever the agent knows.

The axiom system $\mathbb{DKCDL}$ is established based on Kangerian-Andersonian reduction for the deontic operator and we can prove it to be sound. By KA-reduction axioms and reduction axioms for the dynamic operator, we can translate each $\mathcal{L}_{\text{DKCDL}}$-formula to a syntactically equivalent formula of the classical epistemic logic. Therefore we can derive the strong completeness of $\mathbb{DKCDL}$ from the strong completeness of the classical epistemic logic.

In addition to epistemic and factual changes, norm change can also update agents' obligations. The distinction between norms and obligations is known as prescription and description in deontic logic. A normative system is used for prescribing new norms, which could bring about new obligations to agents. So updates on normative systems generally cause obligation change.

Chapter 5 mainly discusses the problem of how a normative system, namely an ideality sequence in our terminology, affects one agent's obligations. The core notion is 'relativized conditional obligations based on ideality sequences', i.e., $\bigcirc_{\mathcal{I}}(\_|\_)$. It is formalized very similarly to Hansson's dyadic obligation operator except that it is relativized to ideality sequences. Every ideality sequence $\mathcal{I}$ is a normative system, which represents the prescriptive aspect of our framework. And betterness structures based on some ideality sequence describe the conditional obligations.

We introduce four updates on ideality sequences: deletion, postfixing, prefixing and insertion, among which deletion and postfixing are elementary updates. We show how these updates on ideality sequences affect obligations by a running example. We do not involve many technical difficulties in the system $\mathbb{PCDL}$ since the Kangerian-Andersonian reduction axiom for the relativized conditional obligation facilitates the completeness proof. We spent comparatively more words on explaining how this new logic resolves a famous philosophical problem in inferences between norms: Jørgensen's dilemma. Norms are not propositions and hence they cannot have truth values. However, classical inferences consisting of propositions are valid if they preserve the truth of premises. The logic $\mathbb{PCDL}$ provides a novel conceptualization on the dilemma. A 'valid' inference between norms should preserve the success of the premises and the set of obligations brought about by the concluded norm is a subset of the set of obligations brought about by the premise norms.

Finally, we put forward a new approach to characterize the notion of knowledge-based obligations. We call it Knowledge-based 'ought-to-do' obligations. Beforehand, all obligations discussed in this thesis are defined with respect to propositions, which means that the deontic operators are applied to propositions. However, how to formalize obligations based on actions has long been investigated by deontic logicians. This question is also known as the question how to provide a proper formalism of 'ought-to-do' instead of the classical 'ought-to-be'. Chapter 6 jumps out of the tradition of 'ought-to-be' and shows a way of capturing the notion of knowledge-based 'ought-do-do' obligations. It is not only knowledge-based, which is analogous to knowledge-based conditional obligations defined in Chapter 3, but also action-based.

Dynamic epistemic logic plays an important role again in this chapter. Each action is represented by an action model or a non-deterministic choice between several action models, which gives us a lot of control on expressing how an action updates a situation. The deontic operator is applied directly to these actions. An action model is obligatory with respect to some agent's knowledge if and only if the action model updates the initial epistemic model to a better one. Priority structures remain important for comparing the initial model and the updated model. In other words, an obligated action improves the current situations. Similarly, the notion of prohibition is given as: an action is prohibited if and only if it always makes the current situation worse. A logic of knowledge-based 'ought-to-do' obligations, i.e., $\mathbb{AKDL}$ is provided and we can prove it to be sound and strongly complete by reduction axioms (different from Kangerian-Andersonian reduction). The logic $\mathbb{AKDL}$ is also able to conceptualize several influential dilemmas of deontic logic.

In summary, this thesis establishes a comprehensive and coherent framework for studying the problem of knowledge-based obligations. It combines epistemic logic, dynamic epistemic logic, preference logic and Hansson's deontic logic altogether to characterize how knowledge affects an agent's obligations.

Samenvatting

Dit proefschrift onderzoekt het probleem van kennis-gebaseerde verplichtingen en hoe kennis verplichtingen beïnvloedt. In Hoofdstuk 3 staat onze eerste statische logica van kennis-gebaseerde voorwaardelijke verplichtingen centraal. In hoofdstuk 4 en 5 bespreken we hoe verplichtingen veranderen als gevolg van informatie, en als gevolg van veranderende feiten en normen. Ten slotte karakteriseren we een nieuwe notie van op kennis gebaseerde '*ought-to-do*'-verplichtingen en wordt de axiomatisering ervan vastgesteld in hoofdstuk 6.

We introduceren eerst het begrip van kennis-gebaseerde voorwaardelijke verplichtingen, dat wordt gedefinieerd over *epistemic betterness structures*. Deze definitie behandelt de relatie tussen kennis en verplichtingen van actoren, die wordt aangegeven door een zeer belangrijke stelling in ons onderzoek: de epistemische onthechting, d.w.z. $\models_{\mathbb{KCDL}} (\odot_i(\phi|\psi) \wedge K_i\psi) \rightarrow \odot_i(\phi|\top)$. Daarom kunnen we zeggen dat een op kennis gebaseerde voorwaardelijke verplichting vooraf codeert welke verplichting een actor heeft als de actor bepaalde informatie kent.

Vanuit het semantische aspect zijn epistemische relaties voor elke actor toegevoegd aan *betterness structures* om zo *epistemic betterness structures* te maken. In plaats van alle mogelijke werelden te beschouwen, richten we ons alleen op die toestanden die de agent epistemisch gezien niet kan onderscheiden van de echte toestand. De actor i heeft een op kennis gebaseerde voorwaardelijke verplichting $\odot_i(\phi|\psi)$ dan en slechts dan als alle beste ψ-werelden die i niet kan onderscheiden van de echte toestand voldoen aan ϕ. Dienovereenkomstig stellen we de logica vast van op kennis gebaseerde voorwaardelijke verplichtingen en bieden we een correcte en sterk volledige axiomatisering $\mathbb{KCDL}$ volgens de 'stap-voor-stap'-methode in plaats van de klassieke benadering van canonieke modellen. Aangezien een epistemische *betterness structure* een generalisatie is van Hansson's *betterness structure*, geeft het volledigheidsresultaat voor $\mathbb{KCDL}$ een antwoord op de open vraag die Parent (2014) heeft achtergelaten voor het systeem van Hansson **F**+(CM).

Vervolgens wordt de statische logica van kennis-gebaseerde voorwaardelijke verplichtingen uitgebreid door dynamische operatoren op epistemische en feitelijke verandering. We zijn daarom in staat om verandering van verplichtingen te karakteriseren als gevolg van epistemische of feitelijke verandering. In het bijzonder werd dynamische epistemische logica geïntroduceerd om deze updates te karakteriseren. Epistemische relaties in een epistemische *betterness structure* worden veranderd door actiemodellen en daarom worden op kennis-gebaseerde voorwaardelijke verplichtingen veranderd. Bovendien bevatten de actiemodellen die in dit proefschrift worden gebruikt ook de functie van feitelijke verandering die de waarheidswaarden van proposities voor elke toestand kan bijwerken. Feitelijke verandering zou ook de *betterness relation* veranderen op basis van een gegeven prioriteitsstructuur.

Volgens de karakterisering van dynamische informatie kunnen verschillende filosofische noties van verplichtingen worden geformaliseerd binnen het nieuwe kader. Een *prima facie* verplichting is een onvoorwaardelijke verplichting wanneer de actor iets weet op een bepaald moment. Alles-overwogen verplichtingen van een actor zijn onvoorwaardelijke verplichtingen waarin de actor alles weet. Veilige kennis-gebaseerde verplichtingen zijn de onvoorwaardelijke verplichtingen, wat de actor ook maar weet.

Het axioma-systeem $\mathbb{DKCDL}$ is opgesteld op basis van een Kangeriaans-Andersoniaanse reductie voor de deontische operator en we kunnen bewijzen dat deze correct is. Door KA-reductie-axioma's en reductie-axioma's voor de dynamische operator, kunnen we elke $\mathcal{L}_{\mathrm{DKCDL}}$-formule vertalen naar een syntactisch equivalente formule van de klassieke epistemische logica. Daarom kunnen we de sterke volledigheid van $\mathbb{DKCDL}$ afleiden uit de sterke volledigheid van de klassieke epistemische logica.

Naast epistemische verandering en feitelijke verandering, kan normverandering ook de verplichtingen van actoren actualiseren. Het onderscheid tussen normen en verplichtingen staat bekend als voorschrift en beschrijving in deontische logica. Er wordt gebruik gemaakt van een normatief systeem voor het voorschrijven van nieuwe normen, die nieuwe verplichtingen voor actoren zouden kunnen meebrengen. Updates op normatieve systemen veroorzaken dus over het algemeen verandering van verplichtingen.

Hoofdstuk 5 behandelt voornamelijk het probleem van hoe een normatief systeem, namelijk een *ideality sequence* in onze terminologie, de verplichtingen van een actor beïnvloedt. Het kernbegrip is 'gerelativeerde voorwaardelijke verplichtingen op basis van *ideality sequences*', d.w.z. $\bigcirc_{\mathcal{I}}(\_|\_)$. Het is op dezelfde manier geformaliseerd als Hansson's dyadische verplichtingsoperator, behalve dat het wordt gerelativeerd tot *ideality sequences*. Elke *ideality sequence* $\mathcal{I}$ is een normatief systeem, dat het voorschriftelijke aspect van ons raamwerk vertegenwoordigt. *Betterness structures* gebaseerd op een of andere *ideality sequence* beschrijven de voorwaardelijke verplichtingen.

We introduceren vier updates over *ideality sequences*: verwijdering, postfixing, prefixing en insertie, waarvan verwijdering en postfixing elementaire updates zijn. We laten aan de hand van een doorlopend voorbeeld zien hoe deze updates van *ideality sequences* de verplichtingen beïnvloeden. Het systeem $\mathbb{PCDL}$ brengt niet veel technische problemen met zich mee, aangezien het Kangeriaans-Andersoniaanse reductieaxioma voor de gerelativeerde voorwaardelijke verplichting het bewijs van volledigheid faciliteert. We hebben relatief meer woorden besteed aan het uitleggen hoe deze nieuwe logica een beroemd filosofisch probleem oplost in gevolgtrekkingen tussen normen: Jørgensen's dilemma. Normen zijn geen proposities en kunnen daarom geen waarheidswaarden hebben. Klassieke gevolgtrekkingen bestaande uit proposities zijn echter geldig als ze de waarheid van premissen behouden. De logica $\mathbb{PCDL}$ biedt een nieuwe conceptualisering van het dilemma. Een 'geldige' gevolgtrekking tussen normen moet het succes van de premissen behouden en de reeks verplichtingen die door de gesloten norm worden veroorzaakt, is een subreeks van de reeks verplichtingen die door de premissennormen wordt veroorzaakt.

Ten slotte stellen we een nieuwe benadering voor om het begrip kennisgebaseerde verplichtingen te karakteriseren. We noemen het kennis-gebaseerde 'ought-to-do'-verplichtingen. Vooraf zijn alle verplichtingen die in dit proefschrift worden besproken gedefinieerd met betrekking tot proposities, wat betekent dat de deontische operatoren worden toegepast op proposities. Hoe verplichtingen op basis van acties kunnen worden geformaliseerd, is echter al lang onderzocht door deontische logici. Deze vraag staat ook wel bekend als de vraag over hoe je een juist formalisme van '*ought-to-do*' in plaats van het klassieke '*ought-to-be*' kunt geven. Hoofdstuk 6 springt uit de traditie van '*ought-to-be*' en toont een manier om de notie van op kennis gebaseerde '*ought-to-do*'-verplichtingen te vatten. Het is niet alleen op kennis gebaseerd, wat analoog is aan kennisgebaseerde voorwaardelijke verplichtingen zoals gedefinieerd in Hoofdstuk 3, maar het is ook gebaseerd op actie.

Dynamisch-epistemische logica speelt in dit hoofdstuk opnieuw een belangrijke rol. Elke actie wordt weergegeven door een actiemodel of een nietdeterministische keuze tussen verschillende actiemodellen, wat ons veel controle geeft over hoe een actie een situatie bijwerkt. De deontische operator wordt direct op deze acties toegepast. Een actiemodel is verplicht met betrekking tot de kennis van een actor dan en slechts dan als het actiemodel het aanvankelijke epistemische model *update* naar een beter model. *Priority structures* blijven belangrijk voor het vergelijken van het initiële model en het bijgewerkte model. Met andere woorden, een verplichte handeling verbetert de huidige situatie. Evenzo wordt het begrip verbod gegeven als: een handeling is verboden dan en slechts dan als het de huidige situatie altijd erger maakt. Een logica van kennis-gebaseerde '*ought-to-do*'-verplichtingen, d.w.z. $\mathbb{AKDL}$ wordt gegeven en we kunnen bewijzen dat deze correct en sterk volledig is door middel van reductie-axioma's (anders dan Kangeriaanse-Andersonische reductie). De logica $\mathbb{AKDL}$ is ook in staat om verschillende invloedrijke dilemma's van deontische logica te conceptualiseren.

Samenvattend stelt dit proefschrift een alomvattend en coherent kader vast voor het bestuderen van het probleem van kennis-gebaseerde verplichtingen. Het combineert epistemische logica, dynamisch-epistemische logica, voorkeurslogica en deontische logica van Hansson om te karakteriseren hoe kennis de verplichtingen van een actor beïnvloedt.

Bibliography

Lennart Åqvist. *Introduction to Deontic Logic and the Theory of Normative Systems*. Bibliopolis, 1987.

Sean Aas. Distributing collective obligation. *J. Ethics & Soc. Phil.*, 9:1, 2015.

Thomas Ågotnes, Wiebe van der Hoek, Juan A Rodríguez-Aguilar, Carles Sierra, and Michael J Wooldridge. On the logic of normative systems. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 1175–1180, 2007.

Carlos E Alchourrón. Detachment and defeasibility in deontic logic. *Studia Logica*, 57(1):5–18, 1996.

Alan Ross Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 67(265):100–103, 1958.

Alan Ross Anderson and Omar Khayyam Moore. The formal analysis of normative concepts. *American Sociological Review*, 22(1):9–17, 1957.

Lennart Åqvist. Good Samaritans, contrary-to-duty imperatives, and epistemic obligations. *Noûs*, 1(4):361–379, 1967.

Lennart Åqvist. Systematic frame constants in defeasible deontic logic. In *Defeasible Deontic Logic*, pages 59–77. Springer, 1997.

Guillaume Aucher, Guido Boella, and Leendert van der Torre. Prescriptive and descriptive obligations in dynamic epistemic deontic logic. In *International Workshop on AI Approaches to the Complexity of Legal Systems*, pages 150–161. Springer, 2009.

Alexandru Baltag and Sonja Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science*, 165:5–21, 2006a.

Alexandru Baltag and Sonja Smets. Dynamic belief revision over multi-agent plausibility models. In *Logic and the Foundations of Game and Decision Theory (LOFT)*, pages 11–24. University of Liverpool, 2006b.

Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. *Logic and the foundations of game and decision theory (LOFT)*, 3: 9–58, 2008.

Can Başkent, Loes Olde Loohuis, and Rohit Parikh. On knowledge and obligation. *Episteme*, 9(2):171–188, 2012.

Mathieu Beirlaen and Christian Straßer. Non-monotonic reasoning with normative conflicts in multi-agent deontic logic. *Journal of Logic and Computation*, 24 (6):1179–1207, 2014.

Nuel Belnap. Backwards and forwards in the modal logic of agency. *Philosophy and phenomenological research*, 51(4):777–807, 1991.

Nuel Belnap and Michael Perloff. Seeing to it that: a canonical form for agentives. *Theoria*, 54(3):175–199, 1988.

Nuel Belnap, Michael Perloff, and Ming Xu. *Facing the future: agents and choices in our indeterminist world*. Oxford University Press on Demand, 2001.

Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal logic*. Cambridge University Press, 2002.

Genevieve Bossu and Pierre Siegel. Saturation, nonmonotonic reasoning and the closed-world assumption. *Artificial Intelligence*, 25(1):13–63, 1985.

Craig Boutilier. Conditional logics of normality: a modal approach. *Artificial Intelligence*, 68(1):87–154, 1994.

Jan Broersen. Action negation and alternative reductions for dynamic deontic logics. *Journal of Applied Logic*, 2(1):153–168, 2004.

Jan Broersen. A logical analysis of the interaction between 'obligation-to-do'and 'knowingly doing'. In *International Conference on Deontic Logic in Computer Science*, pages 140–154. Springer, 2008.

John P Burgess. Basic tense logic. In F. Guenthner and D. Gabbay, editors, *Handbook of Philosophical Logic*, pages 89–133. Springer, 1984.

José Carmo and Olga Pacheco. Deontic and action logics for collective agency and roles. In *International Workshop on Deontic Logic in Computer Science*, pages 93–124, 2000.

Cristiano Castelfranchi. Commitments: From individual intentions to groups and organizations. In V. Lesser, editor, *Proceedings of the First International Conference on Multi-agent Systems*, pages 41–48. MIT Press, 1995.

Laurence Cholvy and Christophe Garion. Collective obligations, commitments and individual obligations: a preliminary study. In *International Workshop on Deontic Logic in Computer Science*, pages 55–71, 2002.

Tiago De Lima, Lambèr M. M. Royakkers, and Frank Dignum. A logic for reasoning about responsibility. *Logic Journal of the IGPL*, 18(1):99–117, 2010.

Frank Dignum and Lambèr M. M. Royakkers. Collective obligation and commitment. In *Proceedings of 5th International conference on Law in the Information Society*, 1998.

Frank Dignum, John-Jules Ch. Meyer, Roel J Wieringa, and Ruurd Kuiper. A modal approach to intentions, commitments and obligations: Intention plus commitment yields obligation. In *Deontic Logic, Agency and Normative Systems*, pages 80–97. Springer, 1996.

Hein Duijf and Frederik Van De Putte. The problem of no hands: responsibility voids in collective decisions. *Social Choice and Welfare*, pages 1–38, 2021.

Barbara Dunin-Kęplicz and Rineke Verbrugge. A tuning machine for cooperative problem solving. *Fundamenta Informaticae*, 63(2-3):283–307, 2004.

Fred Feldman. *Doing the best we can: An essay in informal deontic logic*, volume 35. Springer Science & Business Media, 1986.

Fred Feldman. A simpler solution to the paradoxes of deontic logic. *Philosophical Perspectives*, 4:309–341, 1990.

Dov Gabbay, John F Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre. *Handbook of deontic logic and normative systems*, volume 2. College Publication, forthcoming.

Jelle Gerbrandy and Willem Groeneveld. Reasoning about information change. *Journal of logic, language and information*, 6(2):147–169, 1997.

Margaret Gilbert. Obligation and joint commitment. *Utilitas*, 11(2):143–163, 1999.

Guido Governatori and Antonino Rotolo. Defeasible logic: Agency, intention and obligation. In *International Workshop on Deontic Logic in Computer Science*, pages 114–128. Springer, 2004.

Guido Governatori and Antonino Rotolo. Changing legal systems: Legal abrogations and annulments in defeasible logic. *Logic Journal of the IGPL*, 18(1): 157–194, 2010.

Guido Governatori, Antonino Rotolo, and Giovanni Sartor. Temporalised normative positions in defeasible logic. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pages 25–34, 2005.

Guido Governatori, Antonino Rotolo, Régis Riveret, Monica Palmirani, and Giovanni Sartor. Variants of temporal defeasible logics for modelling norm modifications. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 155–159, 2007.

Patricia S Greenspan. Conditional oughts and hypothetical imperatives. *The Journal of Philosophy*, 72(10):259–276, 1975.

Davide Grossi, Frank Dignum, Lambèr M. M. Royakkers, and John-Jules Ch. Meyer. Collective obligations and agents: Who gets the blame? In Alessio Lomuscio and Donald Nute, editors, *Deontic Logic in Computer Science*, pages 129–145, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

Davide Grossi, Barteld Kooi, Xingchi Su, and Rineke Verbrugge. How knowledge triggers obligation. In *International Workshop on Logic, Rationality and Interaction*, pages 201–215. Springer, 2021.

Susan Haack. *Philosophy of logics*. Cambridge University Press, 1978.

Jörg Hansen. Deontic logics for prioritized imperatives. *Artificial Intelligence and Law*, 14(1-2):1–34, 2006.

Jörg Hansen. *Imperatives and Deontic Logic: On the Semantic Foundations of Deontic Logic*. PhD Thesis, Leipzig University, 2008.

Jörg Hansen, Gabriella Pigozzi, and Leendert van der Torre. Ten philosophical problems in deontic logic. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2007.

Bengt Hansson. An analysis of some deontic logics. *Nous*, pages 373–398, 1969.

David Harel, Jerzy Tiuryn, and Dexter Kozen. *Dynamic Logic*. MIT Press, Cambridge, MA, USA, 2000. ISBN 0262082896.

Andreas Herzig, Emiliano Lorini, Frédéric Moisan, and Nicolas Troquard. A dynamic logic of normative systems. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, volume 2011, pages 228–233. Citeseer, 2011.

Jaakko Hintikka. Knowledge and belief: An introduction to the logic of the two notions. *Studia Logica*, 16, 1962.

Kaarlo Jaakko Juhani Hintikka. *Quantifiers in deontic logic*. Societas Scientiarum Fennica, 1957.

Thomas Hobbes. *Leviathan*. JM Dent, 1914.

John F Horty. Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, 9(1-2):69–91, 1993.

John F Horty. Agency and obligation. *Synthese*, 108(2):269–307, 1996.

John F Horty. *Agency and deontic logic*. Oxford University Press, 2001.

John F Horty. Epistemic oughts in stit semantics. *Ergo, an Open Access Journal of Philosophy*, 6, 2019.

John F Horty and Eric Pacuit. Action types in stit semantics. *The Review of Symbolic Logic*, 10(4):617–637, 2017.

Joris Hulstijn. Need to know: Questions and the paradox of epistemic obligation. In *International Conference on Deontic Logic in Computer Science*, pages 125–139. Springer, 2008.

Jonathan Jenkins Ichikawa and Matthias Steup. The analysis of knowledge. `https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/`, 2018.

Andrew JI Jones and Ingmar Pörn. Ideality, sub-ideality and deontic logic. *Synthese*, 65(2):275–290, 1985.

Jørgen Jørgensen. Imperatives and logic. *Erkenntnis*, 7:288–296, 1937.

Stig Kanger. New foundations for ethical theory. In *Deontic logic: Introductory and systematic readings*, pages 36–58. Springer, 1970.

Immanuel Kant. *Critique of practical reason*. Hackett Publishing, 2002.

Immanuel Kant. *The moral law: Groundwork of the metaphysic of morals*. Psychology Press, 2005.

Barteld Kooi. Expressivity and completeness for public update logics via reduction axioms. *Journal of Applied Non-Classical Logics*, 17(2):231–253, 2007.

Barteld Kooi and Allard Tamminga. Conditional obligations in strategic situations. In *Proceedings of the 3rd International Workshop on Normative Multiagent Systems*, 2008.

Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44: 167–207, 1990.

Piotr Kulicki and Robert Trypuz. Connecting actions and states in deontic logic. *Studia Logica*, 105(5):915–942, 2017.

Holly Lawford-Smith. The feasibility of collectives' actions. *Australasian Journal of Philosophy*, 90(3):453–467, 2012.

Daniel Lehmann. Knowledge, common knowledge and related puzzles (extended summary). In *Proceedings of the 3rd Annual ACM Symposium on Principles of Distributed Computing*, pages 62–67, 1984.

David Lewis. *Counterfactuals*. Oxford: Blackwell, 1973.

David Lewis. Semantic analyses for dyadic deontic logic. In *Logical Theory and Semantic Analysis*, pages 1–14. Springer, 1974.

Kevin Leyton-Brown and Yoav Shoham. Essentials of game theory: A concise multidisciplinary introduction. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2(1):1–88, 2008.

Fenrong Liu. *Changing for the better: Preference dynamics and agent diversity*. PhD Thesis, University of Amsterdam, 2008.

Alessio Lomuscio and Marek Sergot. On multi-agent systems specification via deontic logic. In *International Workshop on Agent Theories, Architectures, and Languages*, pages 86–99. Springer, 2001.

Alessio Lomuscio and Marek Sergot. Deontic interpreted systems. *Studia Logica*, 75(1):63–92, 2003.

Emiliano Lorini. On the logical foundations of moral agency. In *International Conference on Deontic Logic in Computer Science*, pages 108–122. Springer, 2012.

David Makinson. On a fundamental problem of deontic logic. *Norms, Logics and Information Systems: New Studies in Deontic Logic and Computer Science*, 49:29, 1999.

David Makinson and Leendert Van Der Torre. Input/output logics. *Journal of Philosophical Logic*, 29(4):383–408, 2000.

David Makinson and Leendert van der Torre. What is input/output logic? input/output logic, constraints, permissions. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2007a.

David Makinson and Leendert van der Torre. What is input/output logic? input/output logic, constraints, permissions. In *Normative Multi-agent Systems*, number 07122 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2007b. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany. URL `http://drops.dagstuhl.de/opus/volltexte/2007/928`.

Rosja Mastop. Norm performatives and deontic logic. *European Journal of Analytic Philosophy*, 7(2):83–105, 2011.

HJ McCloskey. Ross and the concept of a prima facie duty. *Australasian Journal of Philosophy*, 41(3):336–345, 1963.

Paul McNamara. Deontic logic. In *Handbook of the History of Logic*, volume 7, pages 197–288. Elsevier, 2006.

John-Jules Ch. Meyer. A different approach to deontic logic: deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29(1):109–136, 1988.

John-Jules Ch. Meyer, Frank Dignum, and Roelf Johannes. The paradoxes of deontic logic revisited: A computer science perspective. *Technical Report UU-CS-1994- 38*, 1994.

John Stuart Mill. *On liberty*. London: Longman, Roberts & Green, 1859.

Donald Nute. *Defeasible deontic logic*, volume 263. Springer Science & Business Media, 2012.

Eric Pacuit, Rohit Parikh, and Eva Cogan. The logic of knowledge based obligation. *Synthese*, 149(2):311–341, 2006.

Xavier Parent. On the strong completeness of Åqvist's dyadic deontic logic G. In *International Conference on Deontic Logic in Computer Science*, pages 189–202. Springer, 2008.

Xavier Parent. Maximality vs. optimality in dyadic deontic logic. *Journal of Philosophical Logic*, 43(6):1101–1128, 2014.

Xavier Parent. Completeness of ÅQVIST'S systems E and F. *Review of Symbolic Logic*, 8(1), 2015.

Jan Plaza. Logics of public communications. *Synthese*, 158(2):165–179, 2007.

Henri Poincaré. *Dernières pensées*. E. Flammarion, 1913.

Daniele Porello. A logic for reasoning about group norms. In Jan M. Broersen, Gabriella Pigozzi, Cleo Condoravdi, and Shyam Nair, editors, *Deontic Logic and Normative Systems - the 14th International Conference on Deontic Logic in Computer Science*, pages 301–315, 2018.

Henry Prakken and Marek Sergot. Dyadic deontic logic and contrary-to-duty obligations. In *Defeasible Deontic Logic*, pages 223–262. Springer, 1997.

Arthur N. Prior. *Formal Logic*. Oxford University Press, 1955.

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

Jean-Jacques Rousseau. *A Treatise on the Social Contract, Or, The Principles of Political Law*. DI Eaton, 1795.

Lambèr M. M. Royakkers and Frank Dignum. From collective to individual commitments. In *Proceedings of the 7th international conference on Artificial intelligence and law*, pages 192–193, 1999.

Leonard J Savage. *The foundations of statistics*. Courier Corporation, 1972.

Thomas Scanlon et al. *What we owe to each other*. Harvard University Press, 1998.

Krister Segerberg. A deontic logic of action. *Studia logica*, 41(2-3):269–282, 1982.

Yoav Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*. MIT Press, Cambridge, MA, USA, 1988a. ISBN 0262192691.

Yoav Shoham. *Reasoning about Change: Time and Causation from the Standpoint of AI*. MIT Press, Cambridge, MA, 1988b.

Xingchi Su. Knowledge-based conditional obligation. *Short Papers, Advances in Modal Logic AiML 2020*, pages 112–116, 2020.

Allard Tamminga. Deontic logic for strategic games. *Erkenntnis*, 78(1):183–200, 2013.

Allard Tamminga and Hein Duijf. Collective obligations, group plans and individual actions. *Economics & Philosophy*, 33(2):187–214, 2017.

Johan Van Benthem. *Exploring logical dynamics*. Center for the Study of Language and Information, 1997.

Johan Van Benthem. *Modal logic for open minds*. Stanford: Center for the Study of Language and Information, 2010.

Johan van Benthem and Fenrong Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17(2):157–182, 2007.

Johan van Benthem, Sieuwert van Otterloo, and Olivier Roy. Preference logic, conditionals, and solution concepts in games. *Modality Matters*, pages 61–76, 2006.

Johan van Benthem, Davide Grossi, and Fenrong Liu. Priority structures in deontic logic. *Theoria*, 80(2):116–152, 2014.

Ron van der Meyden. The dynamic logic of permission. *Journal of Logic and Computation*, 6(3):465–479, 1996.

Leendert van der Torre. *Reasoning about obligations: defeasibility in preference-based deontic logic*. Tinbergen Institute Research Series 140. Amsterdam: Thesis, 1997.

Leendert van der Torre and Yao-Hua Tan. An update semantics for deontic reasoning. In Paul Mcnamara and Henry Prakken, editors, *Norms, Logics and Information Systems*, pages 73–90. IOS Press, 1998.

Leendert van der Torre and Yao-Hua Tan. Contrary-to-duty reasoning with preference-based dyadic obligations. *Annals of Mathematics and Artificial Intelligence*, 27(1-4):49–78, 1999.

Hans van Ditmarsch and Barteld Kooi. Semantic results for ontic and epistemic change. In *Logic and the Foundations of Game and Decision Theory (LOFT)*, pages 87–117. Amsterdam UP, 2008.

Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic epistemic logic*. Springer Science & Business Media, 2007.

Bas C Van Fraassen. The logic of conditional obligation. In *Exact Philosophy*, pages 151–172. Springer, 1973.

Georg Henrik Von Wright. Deontic logic. *Mind*, 60(237):1–15, 1951.

Georg Henrik Von Wright. A note on deontic logic and derived obligation. *Mind*, 65(260):507–509, 1956.

Georg Henrik Von Wright. A new system of deontic logic. In *Deontic Logic: Introductory and Systematic Readings*, pages 105–120. Springer, 1970.

Yanjing Wang. A logic of knowing how. In *International Workshop on Logic, Rationality and Interaction*, pages 392–405, 2015.

Yanjing Wang. Beyond knowing that: A new generation of epistemic logics. In *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*, pages 499–533. Springer, Cham, 2018.

Yanjing Wang and Qinxiang Cao. On axiomatizations of public announcement logic. *Synthese*, 190(1):103–134, 2013.

Malte Willer. Dynamic foundations for deontic logic. *Deontic Modality*, pages 324–354, 2016.

Georg Henrik von Wright. Is there a logic of norms? *Ratio Juris*, 4(3):265–283, 1991.

Chao Xu, Yanjing Wang, and Thomas Studer. A logic of knowing why. *Synthese*, 198(2):1259–1285, 2021.

Tomoyuki Yamada. Acts of commanding and changing obligations. In *International Workshop on Computational Logic in Multi-Agent Systems*, pages 1–19. Springer, 2006.

Tomoyuki Yamada. Logical dynamics of some speech acts that affect obligations and preferences. *Synthese*, 165(2):295–315, 2008.

Tomoyuki Yamada. Acts of requesting in dynamic logic of knowledge and obligation. *European Journal of Analytic Philosophy*, 7(2):59–82, 2011.

Appendix A

Strong Completeness of $\mathbb{KCDL}$ in the single-agent case

The basic strategy of proving completeness is attributed to Parent (2014). Parent (2014) still uses the canonical models to prove the strong completeness of **F**+(CM). It is different from the 'step-by-step' approach used in Chapter 3 for the multi-agent case. In terms of $\mathbb{KCDL}$ in the single-agent case, we can also use canonical models to give a proof. But we provide a new definition on $\leqslant$ in the canonical models which keeps $\leqslant$ transitive. Let Γ be a consistent set of $\mathcal{L}_{KCDL}$-formulas. We need to establish a canonical model which satisfies Γ. Let Γ_0 be some maximal consistent extension of Γ. Γ_0^{ψ} denotes $\{\phi \mid \odot(\phi|\psi) \in \Gamma_0\}$ and $K^{-1}\Delta$ denotes $\{\phi \mid K\phi \in \Delta\}$. We will distinguish two cases: (1) Principal case: there is a formula ω such that $\Gamma_0^{\omega} \subseteq \Gamma_0$; (2) Limiting case: there is no formula ω such that $\Gamma_0^{\omega} \subseteq \Gamma_0$.

Before we proceed, the following lemma is necessary.

Lemma 12. *The following formulas are derivable in* $\mathbb{KCDL}$*:*

1. $\odot(\psi_1|\phi) \wedge \odot(\psi_2|\phi) \wedge \cdots \odot(\psi_n|\phi) \to \odot(\psi_1 \wedge \psi_2 \cdots \wedge \psi_n|\phi)$ $(n \geqslant 2)$
2. *If* $\vdash \psi \to \gamma$, *then* $\vdash \odot(\psi|\phi) \to \odot(\gamma|\phi)$.
3. $\odot(\phi|\phi \vee \psi) \wedge \odot(\psi|\psi \vee \gamma) \to \odot(\gamma \to \psi|\phi)$
4. $\neg K\neg\phi \to \neg \odot(\bot|\phi)$
5. $\odot(\gamma|\phi \vee \gamma) \wedge \odot(\psi|\phi) \to \odot(\phi \to \psi|\gamma)$
6. $(\odot(\phi|\phi \vee \psi) \wedge \odot(\psi|\psi \vee \gamma)) \to \odot(\phi|\phi \vee \gamma)$

Principal Case

Definition 80. *(The Canonical Model Generated by* Γ_0*, Principal Case) A canonical model generated by* Γ_0 *is a tuple* $M^{\Gamma_0} = \langle W, \sim, \leqslant, V\rangle$ *where*

1. $W = \{(\Delta, \psi) \mid \Delta$ *is a MCS and* $\Gamma_0^{\psi} \subseteq \Delta\}$[1];

2. $(\Delta, \psi) \sim (\Sigma, \chi)$ *iff* $K^{-1}\Delta \subseteq \Sigma$;

3. $(\Delta, \psi) \leqslant (\Sigma, \chi)$ *iff* $(\odot(\chi|\chi \vee \psi) \in \Gamma_0$ *and* $\psi \notin \Sigma)$ *or* $(\Delta = \Sigma$ *and* $\psi = \chi)$.

4. $V(p) = \{(\Delta, \psi) \mid p \in \Delta\}$ *for any* $p \in \boldsymbol{P}$.

Lemma 13. *(1)* $\sim$ *is an equivalence relation and total; (2) Let* Δ *be a MCS. If* $\odot(\phi|\phi \vee \psi) \notin \Delta$*, then* $\Delta^{\phi \vee \psi} \cup \{\neg\phi\}$ *is consistent; (3) Let* Δ *and* Δ_1 *be two MCSs. If* $\odot(\phi|\psi) \notin \Delta_1$ *and* $K^{-1}\Delta \subseteq \Delta_1$*, then* $\Delta^{\psi} \cup \{\neg\phi\}$ *is consistent.*

Now we can prove the Truth Lemma based on M^{Γ_0}.

Lemma 14. *(Truth Lemma) Let* $M^{\Gamma_0} = \langle W, \sim, \leqslant, V\rangle$ *be a canonical model generated by* Γ_0*. For all* $(\Delta, \psi) \in W$ *and all* ϕ*,* $M^{\Gamma_0}, (\Delta, \psi) \models \phi$ *iff* $\phi \in \Delta$.

Proof. We prove it by induction on the structure of ϕ. When ϕ is a Boolean formula, the proof is standard. When $\phi = K\beta$, it is almost the same as Parent (2014).

When $\phi = \odot(\alpha|\beta)$:

- $(\Rightarrow)$ Suppose that $\odot(\alpha|\beta) \notin \Delta$. By Lemma 13(3), $\Gamma_0^{\beta} \cup \{\neg\alpha\}$ is consistent. So $\Gamma_0^{\beta} \cup \{\neg\alpha\}$ can be extended into a MCS Δ_1. Since $\Gamma_0^{\beta} \subseteq \Delta_1$, $(\Delta_1, \beta) \in W$. Let (Δ_2, γ) be an arbitrary state in W such that $\beta \in \Delta_2$. By Definition 80, $(\Delta_2, \gamma) \not> (\Delta_1, \beta)$. By Lemma 13(1), $(\Delta_1, \beta) \sim (\Delta, \psi)$. Thus, $(\Delta_1, \beta) \in \max_{\leqslant|[(\Delta,\psi)]^{\sim}} \|\beta\|_{M^{\Gamma_0}}$. By the inductive hypothesis, $M^{\Gamma_0}, (\Delta_1, \beta) \models \neg\alpha$. So $M^{\Gamma_0}, (\Delta, \psi) \not\models \odot(\alpha|\beta)$.

- $(\Leftarrow)$ Suppose that $\odot(\alpha|\beta) \in \Delta$. Let $(\Delta_1, \theta) \in \max_{\leqslant|[(\Delta,\psi)]^{\sim}} \|\beta\|_{M^{\Gamma_0}}$. We want to show that $\odot(\theta|\beta \vee \theta) \in \Gamma_0$. Assume, to reach a contradiction, that $\odot(\theta|\beta \vee \theta) \notin \Gamma_0$. By Lemma 13(2), $\Gamma_0^{\beta\vee\theta} \cup \{\neg\theta\}$ is consistent. So it can be extended into a MCS Δ_2 such that $\Gamma_0^{\beta\vee\theta} \cup \{\neg\theta\} \subseteq \Delta_2$. So $(\Delta_2, \beta \vee \theta) \in W$. By the axiom ($\odot$Id), $\beta \vee \theta \in \Delta_2$. So $\beta \in \Delta_2$. By ($\odot$Id) again, we have $\odot(\beta \vee \theta|\beta \vee \theta \vee \theta) \in \Gamma_0$. Since $\theta \notin \Delta_2$, $(\Delta_1, \theta) \leqslant (\Delta_2, \beta \vee \theta)$. And we know $\odot(\theta|\beta \vee \theta \vee \theta) \notin \Gamma_0$. So $(\Delta_2, \beta \vee \theta) \nleqslant (\Delta_1, \theta)$. Thus, $(\Delta_1, \theta) < (\Delta_2, \beta \vee \theta)$. By Lemma 13(1), $(\Delta_1, \theta) \sim (\Delta_2, \beta \vee \theta)$. By the inductive hypothesis, $M^{\Gamma_0}, (\Delta_2, \beta \vee \theta) \models \beta$, which contradicts $(\Delta_1, \theta) \in \max_{\leqslant|[(\Delta,\psi)]^{\sim}} \|\beta\|_{M^{\Gamma_0}}$. Thus, $\odot(\theta|\beta \vee \theta) \in \Gamma_0$. Let γ be an arbitrary formula such that $\gamma \in \Gamma_o^{\beta}$. So $\odot(\gamma|\beta) \in \Gamma_0$. We also have $\odot(\theta|\beta \vee \theta) \in \Gamma_0$. Thus, by Lemma 12(v), $\odot(\beta \to \gamma|\theta) \in \Gamma_0$. Thus, $\beta \to \gamma \in \Gamma_0^{\theta}$. So $\beta \to \gamma \in \Delta_1$. Thus, $\gamma \in \Delta_1$. So $\alpha \in \Delta_1$ as well. Therefore, $M^{\Gamma_0}, (\Delta, \psi) \models \odot(\alpha|\beta)$.

□

[1]MCS represents the maximal $\mathcal{L}_{KCDL}$-consistent set.

Lemma 15. *(Verification Lemma) M^{Γ_0} is reflexive, transitive and $\sim$-smooth.*

Proof. (Reflexivity and Transitivity) Reflexivity is easily verified by Definition 80. Transitivity can be obtained by Lemma 12(iii) and Lemma 12(vi).

($\sim$-smoothness) Let $(\Delta, \theta) \in M^{\Gamma_0}$ such that $M^{\Gamma_0}, (\Delta, \theta) \models \beta$:

- When $\odot(\theta|\theta \vee \beta) \in \Gamma_0$: Assume that $(\Delta, \theta) \notin \max_{\leqslant|[(\Delta,\theta)]^{\sim}} \|\beta\|_{M^{\Gamma_0}}$. This means that there exists $(\Sigma, \lambda) \in M^{\Gamma_0}$ such that $(\Sigma, \lambda) > (\Delta, \theta)$ and $\Sigma \in \|\beta\|_{M_0^{\Gamma}}$. By Definition 80(iii), $\odot(\lambda|\lambda \vee \theta) \in \Gamma_0$ and $\theta \notin \Sigma$. By Lemma 12(v), $\odot(\lambda|\lambda \vee \theta) \wedge \odot(\theta|\theta \vee \beta) \rightarrow \odot(\beta \rightarrow \theta|\lambda) \in \Gamma_0$. So $\odot(\beta \rightarrow \theta|\lambda) \in \Gamma_0$. So $\beta \rightarrow \theta \in \Sigma$, which implies that $\theta \in \Sigma$. Contradiction.
- When $\odot(\theta|\theta \vee \beta) \notin \Gamma_0$, we will show that there is $(\Sigma, \beta \vee \theta) \in M^{\Gamma_0}$ such that $(\Sigma, \beta \vee \theta) > (\Delta, \beta)$ and $(\Sigma, \beta \vee \theta) \in \max_{\leqslant|[(\Delta,\theta)]^{\sim}} \|\beta\|_{M^{\Gamma_0}}$. Since $\odot(\theta|\theta \vee \beta) \notin \Gamma_0$, by Lemma 13(2), $\Gamma_0^{\beta\vee\theta} \cup \{\neg\theta\}$ is consistent. So it can be extended into a MCS Σ such that $\Gamma_0^{\beta\vee\theta} \cup \{\neg\theta\} \subseteq \Sigma$. By Definition 80, $(\Sigma, \beta \vee \theta) \in M^{\Gamma_0}$. Since $\neg\theta \in \Sigma$, we have $\beta \in \Sigma$. Since for any $(\Lambda, \lambda) \geqslant (\Sigma, \beta \vee \theta)$, $\neg(\beta \vee \theta) \in \Lambda$. So $\neg\beta \in \Lambda$. Thus, $(\Sigma, \beta \vee \theta) \in \max_{\leqslant|[(\Delta,\theta)]^{\sim}} \|\beta\|_{M^{\Gamma_0}}$. By the axiom ($\odot$Id), we have $(\Sigma, \beta \vee \theta) > (\Delta, \beta)$.

□

Limiting Case

Definition 81. *(The Canonical Model Generated by (Γ_0, ω), Limiting Case) Take an arbitrary formula ω, the canonical model generated by (Γ_0, ω) is a tuple $M^{(\Gamma_0,\omega)} = \langle W', \sim', \leqslant', V' \rangle$ where $\sim'$ and V' are defined as in Definition 80(ii) and (iii), W' and $\leqslant'$ are defined as follows:*

1. *$W' = W \cup \{(\Gamma_0, \omega)\}$, where $W = \{(\Delta, \psi) \mid \Delta$ is a MCS and $\Gamma_0^{\psi} \subseteq \Delta\}$;*
2. *$\leqslant' = \leqslant \cup \{\langle(\Gamma_0, \omega), (\Gamma_0, \omega)\rangle\} \cup \{\langle(\Gamma_0, \omega), (\Delta, \psi)\rangle \mid (\Delta, \psi) \in W\}$, where $\leqslant$ is defined as in Definition 80(iii).*

The truth lemma and verification lemma for Limiting case can be proved easily based on Lemma 14.

Theorem 10. $\mathbb{KCDL}$ *is strongly complete with respect to the class of epistemic betterness structures that are reflexive, transitive and $\sim$-smooth.*

Acknowledgements

At the moment, outside the window is a blue sky intertwined with dark clouds, which is typical Dutch weather. Amounts of papers are piled on my desk, as well as the final manuscript of my doctoral dissertation spread out on the computer screen. Such a scene reminds me of the first time when I wrote a diary on a computer in the fourth grade of elementary school. At that time I was proud of learning to type words, but now I am proud of completing the last thesis of my student life. This is the beginning and the end of a journey.

The journey started in a small town in Qingyang City, Gansu Province, an industrial town in the Chinese Loess Plateau of the Northwest China, where people came from various surrounding provinces. Since there were few locals, no fixed dialect there. People needed to communicate with people from different places, and Mandarin became the only lingua franca. Similar standardization exists in all aspects of modern society. They are the cornerstones underpinning the modern technology and life. In that small town, I began to study Chinese, mathematics, English, science and other subjects. Chinese is for communication; mathematics is the foundation of rationality; English is for the international vision; and science is for technology. Without exception, these studies are standardized summaries of human discoveries over thousands of years. Therefore, I would like to thank Minglan Ma, Fenglan Xu, Yuhua Qi, Junying Yang and other teachers for their earnest teaching and great encouragement in the elementary school, which helped me to lay a solid foundation for future study and maintain confidence.

The first turning point of this journey occurred after our family relocated to Xi'an. How to adapt to the famous middle school in the big city became the first non-academic challenge of mine. Students there come from different family backgrounds and have broader social perspectives. I have since learned that knowledge outside the book is just as important. Because it is directly related to the social network where we live. A social network is built on the relationship between people. More precisely, it is built on the public values committed by people, that is, universal values, such as fairness, justice, law, kindness, integrity, etc. In terms of study, Euclidean plane geometry has aroused my great interest. Its idea of axiom system that deduces many conclusions from axioms is the most refreshing point. Looking back now, me now is just a 'strict expansion' to me then who was interested in doing geometry proof exercises. In the three years of junior high school, the accumulation of knowledge and the improved understandings on society are the most valuable assets I have learned from many classmates and teachers. I would like to thank my teachers, Shanshan Wang, Feng Zhang, and Tao Yang for their great help. In particular, Feng Zhang gave me great encouragement when I was in difficulty in mathematics, which helped me regain my confidence. I would also like to thank Xi Zhang, Jinwen Li, Rui Zhang, Yifang Li, Qizhe Zhang, Yutong Hu and other friends for their help in all aspects. Thank you for all the wonderful things you gave me in the junior high school.

My senior high school life is entangled with exams in three years. It is also the beginning of self-recognition. I gradually realized that in aspect of the acquisition and understanding of knowledge, there are inherent differences from person to person. Everyone has their own areas of expertise and weakness. ‘Don’t be hard on yourself, don’t despise others, do your best within your ability’ may be the best mentality that a person who is in a fiercely competitive environment can hold. I would like to thank Xuri Zhang, Hongjie Li, Zhaoliang Chen and other teachers for their teaching, especially Hongjie’s encouragement which is a rare compliment from my teachers in the senior high school.

The most comprehensive stage of shaping a person’s values must be the undergraduate period. Out of curiosity and the desire to study both arts and science, I applied to the Department of Philosophy, Sun Yat-sen University, majoring in logic. During the four years, two questions have always been raised repeatedly in my ears: first, what is the study of logic? Second, what career can you do after graduating as a logic student?

The first question is essential. It is also a question that every logician explores in their life. I have heard many answers, and at the end of my doctoral thesis, I also give a provisional answer: logics uses symbolic language and mathematical models to establish rigorous description systems and characterization models for realistic concepts or rules. One of the core ideas of logic and even of analytic philosophy is the rigorous definition of everyday or philosophical concepts, which not only has its own set of standardized terminology like other mature disciplines, but its purpose itself is to standardize other concepts.

However, the second question is a bit tacky. It concerns whether learning logic can bring economic benefits. The fact is that learning logic does not bring direct economic returns to logicians. After graduation, most students in my major choose to work. They choose different paths for the better future that they believe in. There are also a small number of students who are still willing to stay in the logic-related academia. And they are also sticking to the road for the better future that they believe in. As the theme of Chapter 6 in this thesis shows: an action that will lead us to a better future is something that we are obliged to do. Therefore, it is not important if learning logic can make money or not. What matters is whether you can take the path that you think leads you to a better future. I would like to thank Jianying Cui, Yuping Shen, Hu Liu, Wei Wang, Xuefeng Wen, and other professors from the Institute of Logic and Cognition of Sun Yat-sen University for leading me into the gate of logic, so that I can truly appreciate the magnificent scenery on the road to the logic study and approach to a better future.

Apart from these two questions about logic, the university campus itself is a miniature model of society. There are people from all over the world and different social strata. They manage official organizations, student organizations, autonomous associations, etc. in the campus. Different organizations influence and restrain each other. Everyone plays multiple roles and needs to balance study and activities. If I had a superficial perception of the social network since junior high school, then when I was in university, I got a profound experience of social network. Among my two dormitory mates, one may be a student wearing Armani from a developed big city, and the other is a youth from a rural region who has only two clothes for four years. This is the unevenness of society and the great tension between different classes. I can only watch but can't do anything about it. I also began to understand that universal values are not only the foundation of our society, but also the goal of it. I would like to thank Xiang Lu, Huaiyu Zhang, Huawei Zhou, Tong Wu, Xiaoyun Qin, Yinya Huang, Xuan Yang and other classmates who majored in logic for the wonderful friendship.

The life of a master's student was purely about academic training. I started to learn the most obscure things in logic and tried to apply them to research. The three years at Peking University were not happy time. The academic pressure made me feel self-doubting. Fortunately, it was these stressful three years that made me clearly realize that my passion for logic is not groundless, and the difficulties and pressures are not heavy enough to kill my love for this field. I would like to express my great appreciate to my master supervisor Yanjing Wang. It was his strict academic training that set the start point of my logic research. It was also because of his recommendation that I was able to get the opportunity to study in the Netherlands. Many thanks for the careful teaching from my teachers, Beihai Zhou, Zhuanghu Liu, Taotao Xing, and Bo Chen. Studying with you have also laid a solid foundation on logic research for me. I would also like to thank Guangze Yang, Yueqing Zhang, Zhenze Zhao, Fang Lin, Yiran Chen, Guanjun Wang, Shuai Zhang, Zijian Wang and other friends for taking care of me in my study and life.

Reflecting on the four years of my PhD in Groningen, it was the four years when I really became mature. When I started to engage in academic activities as an independent researcher, the first thing that I felt was the pain of research. Doctoral research is basically standing on the boundary of human knowledge, trying to push the extension of human knowledge a little bit further. With the refinement of research topics, there is no longer colleagues whom I can often discuss problems with. Although I can listen to advice from my supervisors, I have to solve every problem by myself. Every night facing the whiteboard to write mathematical proofs is just like dancing with logical symbols. It is indeed a bit like an ascetic, cultivating his own way, without speaking to outsiders.

However, the research itself is very sweet. It is because those problems have not been touched or overcome by the predecessors. The challenges will satisfy the ambition and the desire of conquest of a young person. I especially remember the situation where my supervisor, Barteld, heard an unsolvable problem. His response was always: "Interesting!" instead of "Difficult". Replacing the anxiety of being unable to solve the problem with the fun of solving the problem is the spirit of optimism. Moreover, it is the confidence that a scholar should have.

Apart from the loneliness in academia, life loneliness is a bigger challenge. In a remote town, I stay alone in my office or residence place for night by night. Shrouded in the warm light that Dutch people like most, I have read books, watched movies, and played video games. For the first one or two years, I thought that loneliness was not a problem. But, when the unpredictable outbreak of COVID-19 pandemic disrupted everyone's lives around the world, the long-term lockdown has deprived me of the few opportunities to contact people. The isolation policy has made people feel helpless and lonely. In the growing sense of loneliness, I can feel the torment in my heart coming from nothingness. Fortunately, I have friends by my side. Those getting-together dinners and chats delighted our lives. And my kitty Cloud, you are the elf in my room.

The four years of my PhD life are the four years in which my academic ability has improved the most. It is also the four years in which I hone my mind. The pain can only be understood by myself. But only through these experiences can we be able to do better in the future.

In the end, I would like to express my special thanks to many people who have given me the most sincere help in my growth and taught me the most profound truth.

First of all, I would like to thank Shuya Feng, Xiaolong Wei, Hao Su, and Xiangyun Hu, for the friendship they forged since our childhood, which is so precious that it usually reminds me of fond childhood memories. I would like to thank Yixu Mu for showing your friendliness in my first day in the junior high school. Now our steel-like friendship has been around for almost 20 years. I'm sorry that I could not be present in your wedding. I will congratulate you in person when I get back to Beijing. Thank you Xin Xin for giving me the opportunity to be the groomsman of your wedding. Our friendship lasts forever, I believe. Then I would like to thank Jiaping Sun and Zhiqiao Lin for their great help during the middle school time when I was your desk mates. I wish you a happy marriage and a long-term relationship. Thank you Huaidong Yang for your kindness and thoughtfulness. I am looking forward to seeing you again. Thank you Qiaoya Fu and Zhendong Lu for your help during my senior high school and university time. Your wonderful fate is God's praise. Thank you Wenyi Jing for teaching me to get along with friends with a relaxed attitude. I will send you my belated wishes to your marriage when I go to Chengdu. I would like to thank Su Wu for discussing academic problems with me. Thank you Tianhong Ren for your help in my daily life. You taught me that people must pursue their own hearts in order to live a happy and free life. Thanks to Qi

Yan for leading our undergraduate class into a big loving family. Thanks to Wei Wu, Shibo Fu, Xiaoman Wang, Zesi Zhang, Jiawei Huang, Mingyao Sun, and Xinyi Hu from the Team Yuexiao. It is because of you that my university life is so splendid and beautiful. I would like to thank Jiqi Liu for playing badminton with me during the master's stage. We progress together and I believe that life will always go smoothly. Thank you Dan Li for your willingness to become the main female player of our team, and also for your willingness to let me be your listener. Ryan is the biggest gift from God. Thanks to Guanjun Wang for your valuable information and advice on my job search. See you in Hangzhou! Thank you Shuai Liu for sharing your expertise on philosophy. We have a peace and comfortable friendship. Thank you Kaicheng Fang for your support during my master stage. Your nuttiness is what I think young people should have. Thank you Hengyu Zhang for teaching me how to live independently. I would like to thank Yuzhen Feng and Yanping Gong for inviting me to your happy and cozy dinners for several times. Thanks Mo Liu for recommending me some textbooks on decidability theory. I would like to thank Hang Yang taught me how to do fitness. I wish you an early birth of your child. Thanks Yuzhe Zhang for many academic discussions. I would also like to taste the meals cooked by you more in the future. Thanks Chen Yifei for giving Zhang Yuzhe a warm embrace. Thanks Sha Luo for sharing so much information about graduation and looking for jobs. Thanks Zhenxing Zhang for your invitation to the dinners, which made life in lockdown period a lot more fun. Thanks Zheng Heng for organizing the yard barbecue. Thank you Xu Teke for accompanying me out to taste many delicious food. Thank you Stipe Pandzic, Yuri Santos, Abby Toth, Nicole Orzan, Edoardo Baccini as office colleagues for having lunch and chatting with us every day. I would like to thank Chenglong Deng for bringing me into the Groningen badminton team, and for taking care of me in every aspect of my life. Thank you Yangyang Guo, Shuai Fang, Taiting Chen, Mengzhi Wu, Yang Li, Xiaoping Zheng, Zhiheng Cheng, Xipeng Liu, Di Wen, and Xiaoyin Fang for playing badminton with me. Our weekly sports time is a rare group activity during my PhD. Thanks to Kritika and Crystal for all the help I received when I first arrived in the faculty of philosophy. I would like to thank Cesar Reigosa for organizing the wonderful academic activities as the chairman of the PhD Association of the faculty of philosophy. Thanks to Xiaojun Zhang and Ling Li for chatting with me when I was lonely, and giving me warm 'sunshine' on my lonely days. I would also like to thank Mingpu Ren for designing this thesis cover which gives these obscure words a colorful and beautiful appearance.

Then I would like to solemnly thank my three supervisors, Barteld Kooi, Rineke Verbrugge, and Davide Grossi. In the past four years, I have received the most comprehensive and considerate guidance. The supervisors not only provided me a new direction on the academic road, but also, in our regular meetings, constantly shared new knowledge, methods and serious research attitude with me. Every time when I see the detailed revision comments given by them on the manuscript of my papers, I have been reminded over again and again that academics should always be rigorous and serious. They were gentle enough to point out when I was wrong and gave me the biggest compliments when I was doing well. When my submission was rejected and I was disheartened, their words of encouragement always reignited my fighting spirit. They are the biggest contributors to the completion of this thesis. I would like to thank Bart Verhij, Jan Broersen and Fenrong Liu for reviewing this thesis. Their valuable comments made it better and brought me a lot of new inspiration. Thanks to Cor Steging for helping me translate and revise the Dutch summary part. Thanks to Allard Tamminga for his life and academic advice when I just arrived in Groningen. Thanks to Hans van Ditmarsch for his encouragement and valuable advice on me and my work during several academic conferences. Thanks to Pauline Kleingeld for being my annual assessment examiner. Special thanks to Edoardo Baccini and Francisco Trucco as my paranymphs for their tremendous help with my PhD defense.

Finally, I would like to thank all of my family members. I have lived in a large and harmonious family since I was a child. For this reason, I had a very happy childhood. As the saying goes, lucky people use childhood to heal their whole life, and that’s what I am. Even though I encountered many hardships after leaving home to study, whenever I think of my family, the warmth that flows from the bottom of my heart can immediately heal the wound. I would like to especially thank my father Sheng Su and my mother Ping Hu. They taught me to be kind and humility. They taught me to be upright and how to get along with others. The parenting kindness can never be repaid. At this moment, let me give this doctoral dissertation to you as a gift. Please also let me take good care of you in the future. And wish you health and happiness.

After all these thanks, I want to go back to logic itself. The reason why the complex human society can continuously emerge new orders and break old customs in the chaos is precisely because people need standardized regulations to realize the desire of communication. And the changes of human desires in turn create new standardization. In the past, logic tried to provide a standardized norm for human cognition. But the demand for understanding human beings led logic to the direction of describing human cognition. Logic is such an ambitious discipline that, like sociology, can describe and explain people and society, and also like mathematics, can regulate people and society. I cannot say how much I can contribute to logic in the future. But logic has taught me to fly to a higher dimension to see the world clearly. However, in the society, the growing life experience has taught me that only by diving into the world and experiencing the various practical situations can I keep calm and peace mentally.

Logic and life are my gains.

致谢

此刻，窗外是乌云和蓝天交织的天空，这是典型的荷兰天气。桌前堆满了看过的和没看过的论文，还有铺在电脑屏幕上的博士学位论文终稿。这样的场景让我想起我小学四年级第一次用电脑写下一篇完整日记时的心情，只不过那时的我是因为学会了打字而骄傲，现在的我是因为完成了学生生涯最后一篇大论文而骄傲。这是一段旅程的起点和终点。

这段旅程开始于中国甘肃省庆阳市的一个小镇，那是大西北黄土高原中的工业小镇，那里的人们来自周围各个省份，却少有当地人，因此没有固定的方言。人们需要与来自不同地方的人交流，普通话就成了唯一的通用语言。类似的标准化存在于现代社会的方方面面，它们是支撑着现代科技和便利生活的基石。在那个小镇，我开始学习语文，数学，英语，科学等科目，语文为交流之便，数学为形式之根本，英语为国际化视野，科学为技术之利。无一例外，这些学问都是对人类经岁累月的发现的标准化总结。因此我要感谢马明兰老师、徐凤兰老师，齐瑜华老师，杨军英老师等诸位老师的谆谆教诲和莫大鼓励，让我为将来的学习打下坚实的基础并且保持自信。

这段旅程的第一个转折点发生在我们举家搬迁到西安市之后，初来乍到的小镇学生如何适应大城市的知名中学成了我要面对的第一个非学术难题。同学们的家庭背景各不相同，拥有更加宽广的社会视野。我也自此才明白书本外的见识同样重要，因为这些知识才直接关乎我们所处的社会网络。社会网络是构建在人与人的关系之上的，更准确的说，它建立在人与人共同承诺的普世价值之上，比如：公平，正义，法治，良善，诚信等。在学习方面，欧氏平面几何引起了我极大的兴趣，其从公理出发推出诸多结论的理论系统最让我耳目一新，如今回头看，现在做逻辑学研究的我不过是那时最爱做几何证明题的我的“严格扩张”罢了。初中三年，知识的累积和对社会认识的提升，是我从诸多同学及老师身上所学到的最宝贵的财富。在此感谢王珊珊老师，张峰老师，杨涛老师给予我的特别的关照，尤其是张峰老师在我数学学习陷入困难之际给予了我极大的肯定和鼓励，让我重振信心。还要感谢张玺、李锦文、张蕊、李怡芳、张启哲、胡雨桐等同学在生活和学习中的无处不在的帮助，谢谢你们在初中阶段给予我的美好的一切。

高中是与考试纠缠的三年，也是学会自我认知的开始。我逐渐明白在知识的获取和理解上，人与人有着天生的差异，每个人都有其擅长与不擅长的领域。“不苛求自己，不轻视他人，能力之内，竭尽全力”也许是一个身处激烈竞争的人所能持有的最佳心态。感谢张旭日老师，李宏杰老师，陈昭亮老师等诸位师长的教导，尤其是李宏杰老师的鼓励，是我高中阶段不可多得的来自于老师的夸奖了。

如果说塑造一个人价值观最全面的阶段，那一定是大学本科时期。出于好奇以及文理兼修的愿望，我报考了中山大学哲学系逻辑学专业。四年大学时光，耳边总有两个问题反复被提起：一是逻辑学到底研究什么？二是逻辑学专业毕业到底能从事什么工作?

第一个问题充满了学究气，亦是每个逻辑学家终其一生探索的问题，我听到过很多答案，在博士论文的最后，我也给出一个我自己的暂时性的答案：逻辑学用符号语言和数学模型为现实概念或规则建立严格的描述系统和刻画模

型。逻辑学乃至分析哲学的核心理念之一就是严格地定义日常或哲学概念，它不仅像其它成熟的学科一样有一套自己的标准化术语，其目的本身就是标准化其它概念。

然而，第二个问题则俗气了许多，它关心的是学习逻辑学是否能够带来经济回报，而赚钱是所有社会个体都必须要面对的问题。事实是学习逻辑并不能给逻辑学家带来直接的经济回报，本科毕业后大部分同学选择就业或者转读其它专业研究生，他们为了他们相信的更好的未来选择了不同的道路；也有少部分同学依然留在逻辑学学术圈寒窗苦读，他们也是为了他们相信的更好的未来坚守了道路。正如这篇博士论文第六章的主题一样：一个能够把我们带领到更好的未来的行为是我们有义务去做的。所以学习逻辑赚不赚钱不重要，重要的是能不能走自己认为通往更好的未来的道路。感谢崔建英老师，沈瑜平老师，刘虎老师，王玮老师，文学锋老师等诸位中山大学逻辑与认知研究所的老师把我领进逻辑学的大门，让我真正能够有幸领略逻辑学中的无限风光，让我能够走上通往更好未来的道路。

除去这两个关于逻辑学的问题，大学校园本身就是一个社会的微缩模型。在一方不大的园子里，有来自五湖四海的人们，来自不同社会阶层的人们，管理着学校的领导机构、官方组织、学生组织、自治社团等，不同组织之间互相影响，相互牵制。每一个人都扮演着多种角色，需要平衡好日常学习和校园活动。如果说初中开始我对于社会网络的概念有了感知，那么大学时期就是对社会网络有了深刻的体验。我的两个邻节点中，一个可能是穿着阿玛尼的城市子弟，另一个则是四年只有两套单衣月生活费300元都不到的乡下青年。社会的参差以及身处他们之间那种巨大的张力，我只能看着却无能为力。我也开始明白，社会网络的基础是普世价值，但同时，社会网络的目标也是普世价值，我们的社会依然需要向这一目标前进。在此我要感谢2011级逻辑学专业的陆翔、张怀宇、周华威、吴桐、卢振东、秦晓昀、黄殷雅、杨璇等同学和我建立的美好的大学友谊。

硕士研究生的生活是纯粹的学术训练，我真正开始了学习逻辑学中有难度的知识并且经历了将所学转换为新的学术成果的过程。在北京大学的这三年并不快乐，学术的压力让我对自己的能力和未来产生了极大的不自信。但是值得庆幸的是，正是这充满压力的三年，让我清楚认识到我对于逻辑学的热情并非空穴来风，困难和压力不足以消磨我对这一学科的喜爱，那么我就值得在这条道路上坚持下去。我要特别感谢我的硕士导师王彦晶，是他对于学术的严格要求和给我们的全面训练让我能够有资格开始逻辑学的研究，也是因为他的推荐，我才能够获得去荷兰深造的机会。我也要感谢北京大学逻辑学教研室的周北海老师、刘壮虎老师、邢滔滔老师、陈波老师的悉心教导，跟随你们学习也为我打下了牢固的逻辑学基础。在此还要感谢杨广泽、张月青、赵真泽、林芳、陈逸然、王冠军、张帅、王子剑等同学在学习与生活上对我的照顾。

而回望在格罗宁根读博士的这四年，是我真正成熟起来的四年。当我以一个独立研究者的姿态开始从事学术活动时，我首先感受到的是研究的苦，颇有拔剑四顾心茫然的无力感，更何况，我的剑还并不锋利。博士研究本就是站在人类知识的边缘上，试图将人类知识的外延再推开一点点，而随着研究课题的细化，我的身边不再有能时常讨论问题的同学，虽然可以从导师处听取建议，但是终究要自己独立解决。每一个面对白板写证明的夜晚，都是和逻辑符号的孤独起舞。这确乎有点像苦行僧，修着自己的道，外人不可语。但是，研究本

身又很有甜头，正是因为那些问题是前人未能触及或者攻克的，它们才会满足年轻人的野心和征服欲。尤记得导师Barteld每每听到我阐述一个无法解决的问题时，他的回应总是："Interesting！"而非"Difficult"。用解决问题的有趣去替代无法解决问题的焦虑，这是乐观精神，更是学者该有的自信。

除去学术上的孤独感，生活上的孤独感是更大的挑战。在异国他乡的偏远小城，绝大部分的深夜，一个人待在办公室或者住所里，在荷兰人最喜欢的暖光灯下，我看过书，看过电影，打过游戏。开始的一两年，我自认为孤独并不是问题，但是世事难料，突如其来的百年一遇的新冠疫情打乱了世界上每一个人的生活。长期的封锁和居家办公剥夺了我本就不多的与人接触的机会，隔离政策更是让人感到脱离社会的无助与孤单。在逐渐增长的孤独感之中，我能感受到自己内心的煎熬，那是一种来自于虚无的包围和折磨。好在身边有朋友，是和他们的一次次聚餐和闲聊中，我感受到了人作为社会的动物的本性，也汲取了最纯粹的快乐。还有我的小猫咪Cloud，你是陪伴我的小精灵。

博士四年是我学术能力进步最快的四年，也是我磨练心智的四年。个中苦楚，只有自己明白。但是只有经历过这些，在未来才能游刃有余。

在最后，我要特别感谢一些人，是他们在我的成长中给予了我最诚挚的帮助，教会了我最深刻的道理。

首先，我要感谢冯舒雅、魏小龙、苏浩、胡翔云，和他们从小缔结的友谊，是时至今日依然能带给我美好童年回忆的珍贵感情。感谢牟一旭在我刚上初中时伸出的橄榄枝，如今我们钢铁般的友谊已经快二十年了，抱歉你的婚礼我无法到场，待我回北京再行祝贺。感谢辛鑫给了我平生第一次当伴郎的机会，和你的友谊是那种不论多久不见都不会生疏的情感。然后我要感谢孙嘉平、林郅乔在和我坐同桌的中学时光里给予我的莫大帮助，祝你们婚姻幸福，百年好合。感谢杨怀东的善良与体贴，期待重逢时。感谢付乔雅、卢振东夫妇在我高中和大学阶段给予的帮助，你们奇妙的缘分是老天的褒奖。感谢荆文奕教会我用轻松愉快的心态和同学相处，等我去成都再给你送上迟到的新婚祝福。感谢伍素在本科阶段对我学习上的帮助，和你愉快的学术讨论开启了我走向逻辑学研究的道路。感谢任天鸿对我生活上的帮助，在你身上我学到了人要追求自己的本心才能活得快活洒脱。感谢闫琦作为班长带领本科班级成为了一个有爱的大家庭。感谢约宵队的吴玮、伏仕波、王晓曼，张泽思、黄嘉炜、孙明耀、胡昕宜，正是因为你们，我的大学生活才是那么的灿烂美好。感谢刘纪琪在硕士阶段陪我打球，我们一起进步，无论是球技还是学术，人生总会顺利起来。感谢李丹愿意成为球队的主力女将，也感谢你愿意让我做你的倾听对象，如今的如意郎君是你用赤诚之心换来的幸福。感谢王冠军在我找工作期间提供的宝贵信息和建议，有机会我们杭州再见。感谢柳帅关于学术的分享，和你的关系有种平淡而舒服的感觉。感谢方凯成在我当班长期间的鼎力支持，你的洒脱和专注是我认为的年轻人该有的气质。感谢张恒禹教会我如何独立生活。感谢冯玉振和宫艳萍夫妇对我生活上的照顾，每次聚餐都是非常快乐的回忆。感谢刘默向我推荐的关于可判定性理论的书籍。感谢杨航作为办公室同事介绍给我健身的理念，还有好吃的各种烤箱制品，祝你早生贵子。感谢张语哲和我进行的诸多学术讨论，还有做的一手好菜，日后还想多多品尝。感谢陈一菲给了张语哲一个温暖的怀抱。感谢罗莎分享的诸多关于毕业和工作的信息。感谢张振兴的聚餐邀请让隔离生活多了很多乐趣。感谢郑珩多次贡献家中大院让我们体会了露天烧烤的快乐。感谢徐特科陪我出去品尝美食。感谢Stipe

Pandzic、Yuri Santos、Abby Toth、Nicole Orzan作为办公室同事和我们一起吃饭聊天的快乐日常。感谢邓成龙把我领进格罗宁根羽球大家庭，更感谢他在生活中方方面面对我的照顾。感谢郭洋洋、方帅、陈泰廷、吴梦之、李扬、郑潇平、程志恒、刘曦鹏、温迪、方效寅陪我一起打球，我们每周的运动时光是我读博期间难得的团体活动。感谢Kritika在我初到哲学系时给予我的诸多帮助。感谢Cesar Reigosa担任哲学系博士生协会主席负责人期间组织的精彩学术活动。感谢张潇俊和李灵在我孤单的时候陪我闲聊，给了我寂寞日子里温暖的阳光。更要感谢任明普设计的论文封面让我晦涩的文字有了一个色彩纷呈又美观大方的面子。

然后我要郑重地感谢我的三位导师Barteld Kooi, Rineke Verbrugge, Davide Grossi，四年里我接受到了最全面和最体贴的博士指导，各位老师不仅在学术道路上给我指引了全新的方向，更是在一次次的日常会议中，不断和我分享新的知识、思维方法以及认真研究的态度。每每看到各位老师给我的论文草稿提出的详尽的修改意见，就像是一次次的提醒，做学术要永远保持严谨认真，宁少勿滥。在我犯错的时候他们能够温和地指出问题，在我做得还不错的时候给予了我最大的表扬。当我的论文被一次次拒稿心灰意冷之际，你们的充满鼓励的话语总是能让我重燃斗志。这一篇论文的成型，你们是最大的功臣。感谢Bart Verhij、Jan Broersen和刘奋荣老师担任我的论文评审，他们的宝贵意见让这篇论文变得好也带给了我很多新的启发。感谢Cor Steging帮助我翻译和修改该论文中的荷兰语总结。感谢Allard Tamminga在我初到格罗宁根时给予我的生活和学术上的建议。感谢Hans van Ditmarsch在学术会议中给予我的肯定和宝贵建议。感谢Pauline Kleingeld担任我的年度考核考官。还要特别感谢Edoardo Baccini和Francisco Trucco作为我的答辩"小蜜蜂"为我的博士答辩做出的巨大帮助。

最后我要感谢我所有的家人，从小我就生活在一个和睦的大家庭中，也因此我拥有了一个非常幸福的童年。俗话说，幸运的人用童年来治愈一生，我便是如此。即便在离家读书后遇到了诸多困难，但是每每想起家人，那种从心底流淌而出的暖意都能立刻抚平伤口。我要特别感谢我的父亲苏胜和母亲胡萍，他们教会我学习的态度和谦卑的品行，他们教会我做人的风骨和处事的道理。父母的养育之恩无以为报，此时此刻，就让我把这本博士论文赠予你们，回馈你们的谆谆教诲。也请让我在今后的漫漫岁月里，好好照顾你们，祝福你们健康幸福。

在感谢之后，我想回到逻辑学本身。纷繁复杂的人类社会之所以能够在一片混沌中不断涌现新的秩序和破除旧的习俗，正是因为人们需要标准化的规定来实现交流的欲望，而人类欲望的变迁又反过来缔造了新的标准化系统。曾经的逻辑学试图为人类的认识提供一个标准化的规范，但是对于认识人类自己的诉求又引领逻辑学走向了描述人类的方向。逻辑学是如此有野心的一门学科，既要像社会学一样能够描述和解释人与社会，又要像数学一样规范人与社会。我无法说我未来能为逻辑学作出多少贡献，但是逻辑学教会了我要能飞到更高维的视角才能看清楚世界。但是，身处社会之中，不断增长的人生阅历又教会了我只有沉下去体验世间百态才能处变不惊。

逻辑与生活，都是我的收获。