Conditional Beliefs in Action

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Informatiker

Christoph Schwering

aus Euskirchen

Berichter: Universitätsprofessor Gerhard Lakemeyer, Ph.D.
Universitätsprofessorin Dr. rer. nat. Gabriele Kern-Isberner

Tag der mündlichen Prüfung: 29. Juni 2016

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.

Abstract

Humans rarely have sound or even complete knowledge about their environment. Instead, we usually picture different contingencies what the world could be like. For example, we might believe that a specific box is presumably empty, and that otherwise it most plausibly contains a gift. On the grounds of (conditional) beliefs like this we act. Sometimes we perceive new information that refutes some of these contingencies; then we revise our beliefs appropriately. To a human, all this is natural and mundane.

For a machine to do the same, it needs a formal representation and understanding of *conditional belief* and *actions* and *perception*. In this thesis we develop a formalization and semantics of these concepts in one coherent logical language and investigate their interaction. The main contributions are as follows.

First we introduce a method to express that a set of conditional beliefs is all that is believed. This captures the idea that a (conditional) knowledge base covers the agent's beliefs to their full extent. We refer to this concept as *only-believing*, as it generalizes Levesque's only-knowing to conditional beliefs. It can also be considered a semantic version of Pearl's meta-logical System Z.

Then we investigate the belief *projection problem*, which refers to determining what is believed after a number of actions have occurred. Solving the belief projection problem is essential to reason about beliefs in dynamic systems, like a robot for example. We propose two solutions in the framework of Reiter's situation calculus. Namely, we extend the well-known concepts of query regression and knowledge base progression to conditional beliefs.

Finally, as a step towards practical reasoning about beliefs and contingencies, we develop a *limited-reasoning system* for conditional beliefs. We complement Lakemeyer and Levesque's limited first-order inference with a novel sound first-order consistency test. Together, these techniques enable us to approximate the notions of conditional belief and only-believing in a way that is sound and decidable for an important class of problems.

Zusammenfassung

Selten verfügen Menschen über korrektes oder gar vollständiges Wissens über ihre Umwelt. In der Regel haben wir lediglich Vermutungen, wie die Welt unter bestimmten Bedingungen aussehen könnte. Zum Beispiel könnten wir glauben, dass ein Paket vermutlich leer ist, andernfalls aber am ehesten ein Geschenk enthält. Aufgrund solcher (bedingter) Vermutungen handeln wir. Wenn sich eine Vermutung aufgrund neuer Information als falsch erweist, passen wir unsere Vorstellungen entsprechend an.

Um ähnliches zu leisten, muss eine Maschine über eine formale Repräsentation und über ein Verständnis von Konzepten wie *bedingte Vermutung*, *Aktion* und *Wahrnehmung* verfügen. Die vorliegende Arbeit entwickelt eine Formalisierung und Semantik dieser Begriffe in einer logischen Sprache. Die wesentlichen Beiträge sind wie folgt.

Zunächst führen wir ein Konzept ein um auszudrücken, dass eine Wissensbasis bestehend aus bedingten Vermutungen die Vorstellungen eines künstlichen Akteurs vollständig erfasst. Wir bezeichnen dieses Konzept als *Only-Believing*, da es Levesques Only-Knowing auf den Fall der bedingter Vermutungen verallgemeinert. Es kann auch als semantische Version von Pearls metalogischem System Z aufgefasst werden.

Dann untersuchen wir das *Projektionsproblem*, bei dem es darum geht zu entscheiden, was man nach einer Sequenz von Aktionen glaubt. Dieses Problem zu lösen ist unerlässlich um in dynamischen Systemen – etwa einem Roboter – über Vermutungen zu schließen. Wir entwickeln zwei Lösungen im Rahmen von Reiters Situationskalkül; und zwar verallgemeinern wir die bekannten Ansätze von Regression und Progression für bedingte Vermutungen.

Schließlich entwickeln wir als Beitrag in Richtung praktikablen Schließens über Vermutungen und Eventualitäten ein *System für eingeschränktes Schließen*. Dem abgeschwächten Inferenzmechanismus für Logik erster Stufe von Lakemeyer und Levesque stellen zu diesem Zweck einen korrekten Konsistenztest zur Seite. Diese beiden Techniken erlauben uns bedingte Vermutungen und Only-Believing zu approximieren und dabei Korrektheit und Entscheidbarkeit zu erhalten.

Acknowledgements

The journey to this doctorate was exciting and enriching, and I am very grateful for this experience. Above all, I would like to thank Gerhard Lakemeyer for supervising and mentoring me for the past four years. While he granted me the freedom (and responsibility) to choose a field of research on my own, his door and his mailbox were always open to discuss ideas and problems. His views, his comments, and his rigour have shaped my understanding of science and artificial intelligence, and their influence on this thesis couldn't be bigger.

I would also like to extend my particular thanks to Gabriele Kern-Isberner for taking on the role of the second examiner of this thesis and for her comments on my work from the belief-revision perspective. Furthermore I want to thank Martin Grohe and Bastian Leibe for acting as additional examiners at my defence.

Two visits to The University of New South Wales were turning points for my thesis: the first brought me into contact with my niche of research, and the second led to a good deal of the results. I am thankful to everybody involved in making these visits possible, especially to Gerhard Lakemeyer for initiating the first one and supporting the second, to Gabriele Kern-Isberner for acting as evaluator, and of course to Maurice Pagnucco for hosting and supervising me in Sydney.

I was very warmly welcomed to Gerhard Lakemeyer's Knowledge-Based Systems Group and the Chair for Computer Science 5 at RWTH Aachen as well as to the Artificial Intelligence group at the UNSW School of Computer Science and Engineering. I want to thank my coworkers and friends from these groups for making me feel at home in both Aachen and Sydney and for making the past four years such an enjoyable time. A special thanks goes to Jens Claßen and to my office mate Stefan Schiffer; I benefited hugely from their support and their wisdom especially in teaching-related matters. I would also like to thank the secretarial staff, Daniele Glöckner, Gabriele Hoeppermanns, and Claudia Puhl, for always helping me with any administrative matters.

Lastly, I want to acknowledge the financial support from the B-IT Research School, the German Academic Exchange Service (DAAD), the German Science Foundation (DFG), and the German taxpayer, who most likely eventually payed the bill.

Contents

1 Introduction

With the advent of computing hardware, a group of scientists founded a new field of research called *artificial intelligence* (McCarthy, Minsky, et al. 1955), which McCarthy (2007) later described as "the science and engineering of making intelligent machines." While there are many different interpretations of the term "intelligence" itself (Neisser et al. 1996), most research on artificial intelligence, including this thesis, is concerned with *common sense*.

Common sense is fundamental for humans to get by in daily life. As we rarely have sound or even complete knowledge about our environment, we usually employ common sense to reach assumptions about what the world could be like. On the grounds of these beliefs we act. We acquire new information when needed, and when such information is contrary to our previous beliefs, instead of questioning everything we revise our beliefs appropriately. We do so "with common sense," and with such a routine in our daily lives that, unlike, say, a mathematical exercise, it puts no burden on us.

A computer, on the other hand, solves arithmetical problems with ease, but is usually not endowed with common sense. The machine not only has no commonsense knowledge, such as "when I drop an object and hear a clink, it is presumably broken." It does not even know any such concepts as *belief*, *action*, or *perception*. To "understand" these concepts and to reason about them, a computer needs a formal representation thereof and computable procedures to draw inferences from that representation. In this thesis, we develop a logical theory for that purpose.

This theory will allow us to express beliefs and to reason about their consequences: how do our beliefs change after we perform actions or after we obtain new information? Of course, there is no singular answer to that. In fact, the jury is still out on each individual subfield involved: reasoning about knowledge and belief, conditional logic, action theories, and belief revision. Considering the great many intricate philosophical issues still under debate, we take a rather pragmatic stand here. We aim for a *workable* theory that unifies conditional belief, actions, and belief change. After a thorough investigation a fully fledged logic without much attention to computational complexity, we also devise a decidable fragment, which is foundational for practical use.

1.1 Motivation

As a motivating example, consider the following scenario.

Example 1.1.1 It is Christmas, the presents have just been handed out. Now it is time to put away the rubbish. You are holding a gift box. In the belief that somebody took out the present already, you drop the box on the rubbish heap. Just when it hits the ground, a clinking noise rings out. It looks like you were wrong: it seems there *is* a gift in the box, and now it is broken. Anxious to know, you open the box and pull out a shiny, unimpaired object. Apparently you were wrong again: yes, there was a gift in the box – and now you even know what it is – but it is not broken after all.

What we witnessed here is the interplay of beliefs, actions, and perception as it happens constantly in our daily lives. An intelligent agent – like a robot for the cleaning work in our example – needs to deal with belief change, be it because they see (or cause themselves) physical change, or because they find evidence their beliefs were wrong in the first place.

While easy for humans, such commonsense behaviour is hard to realize for robots. The field of *cognitive robotics* as envisaged by Reiter and his colleagues (Levesque and Lakemeyer 2008; Levesque and Reiter 1998; Reiter 2001) aims "to develop an understanding of the relationship between the knowledge, the perception, and the actions of [an autonomous] robot" (Levesque and Reiter 1998).

Here, we are concerned with this problem in a setting where the agent's beliefs are not only incomplete, but may also turn out wrong in the face of new information. We aim for a logical language suitable to represent and reason about problems like Example 1.1.1. While robots only serve motivational purposes in this thesis, our fundamental approach stands in the tradition of cognitive robotics, and we hope one day it will help to control robots at a high abstraction level.

1.2 Contributions

This thesis presents a logical framework for representation of and reasoning about defeasible beliefs in dynamic domains. The following three questions are a motif throughout the thesis.

1. How can we capture the meaning of a conditional knowledge base in a semantically perspicuous way?

2. How do conditional beliefs change in the face of physical actions and new infor-

mation?

3. How can reasoning about conditional beliefs be kept decidable and, sometimes, tractable?

We investigate these questions within the framework of mathematical logic. The key features of the approach are the following.

First-order modal logic The language is a first-order modal dialect. First-order logic allows us to express properties of objects and quantify over them. In Example 1.1.1, quantification is used to say that there is some (unidentified) object x which is broken. Formally, this can be written as $\exists x\,\mathrm{Broken}(x)$. Quantification is a very powerful tool, but also the source of first-order logic's undecidability.

Conditional beliefs The language features a modal operator to express conditional belief: the formula $\mathbf{B}(\mathrm{InBox}(x) \wedge \mathrm{Fragile}(x) \Rightarrow \mathrm{Broken}(x))$ denotes the belief that if x is in the box and fragile, then it is presumably broken. Many other interpretations of conditionals are possible, including the counterfactual one: we believe that if x *was* in the box and fragile, then it *would* be broken.

Only-believing A second belief modality allows to capture *all* that is believed. Such a concept is of great use in knowledge representation, where we are typically interested in what follows from a knowledge base. Only-believing allows to express that a conditional knowledge base is believed and everything else is not believed (except the logical consequences of the knowledge base). It is this implicit representation of non-belief that makes only-believing special.

Actions and informing Another family of modal operators concerns actions. Actions not only can have physical effects, but also carry information. An example of the former is the physical action of dropping the box in Example 1.1.1; an example of the latter is the clink, which conveys the information that something might be broken. Such information will be incorporated into the agent's beliefs by means of belief revision.

Belief projection The belief projection problem refers to determining what is believed after a number of actions have occurred. It can be cast as a logical entailment problem: given a knowledge base and a query in the form of a logical formula with actions and beliefs, does the knowledge base logically imply the query? Solving the problem requires to take the dynamics out of this reasoning task. We offer two approaches: one by regression, where the query is rewritten to "roll back" the effects of the actions; and one by progression, where the knowledge base is updated to account for the effects of actions.

Reduction to non-modal reasoning After applying regression or progression, one is typically left with a logical entailment problem without actions but that still involves conditional belief. We show that these conditional beliefs can also be eliminated by reducing them to first-order reasoning tasks. That way, a large class of reasoning problems in our framework can be reduced to non-modal first-order reasoning.

Limited reasoning First-order reasoning is known to be undecidable. To address this, we devise a new logic of conditional belief with a weaker inference mechanism which still retains the expressivity of a first-order language. We show that this logic is sound with respect to the original language under certain circumstances. While this logic does not accommodate actions, this is no effective limitation thanks to our solutions of the projection problem.

1.3 Outline

The thesis is structured as follows. After reviewing related work in the next chapter, we introduce the logical foundations in Chapter 3. This includes the first-order logical language $\mathcal{L}$, which serves as basis of all the languages to come. For future reference, we also present two other logics known from the literature: the logic of knowledge and only-knowing $\mathcal{OL}$, and a variant thereof called $\mathcal{ES}$ that adds actions and sensing.

Chapter 4 introduces a novel logic of conditional belief and only-believing called $\mathcal{BO}$. Most notably, we prove a unique-model property for only-believing, a very useful result for the rest of the thesis. We also give a representation theorem, which allows to simplify many reasoning tasks in $\mathcal{BO}$. Besides these results, we discuss the relation of $\mathcal{BO}$ to its ancestor $\mathcal{OL}$ and to another framework of conditional beliefs called System Z.

Chapter 5 extends $\mathcal{BO}$ to accommodate actions. The result is called $\mathcal{ESB}$, for epistemic situation calculus with beliefs. We extensively investigate the belief projection problem in $\mathcal{ESB}$ and prove regression, revision, and progression theorems. $\mathcal{ESB}$ outranks the other languages developed in this thesis in terms of their expressivity; in this sense, it is the premier stage of this thesis. We also compare $\mathcal{ESB}$ briefly to its ancestor $\mathcal{ES}$.

The next two chapters address the issue of undecidability that tarnishes the languages up until then. Chapter 6 introduces two non-standard semantics of a non-modal first-order logic. These so-called limited semantics are sound or complete, respectively, with regard to a meaningful fragment of unlimited first-order logic. They set the stage for the

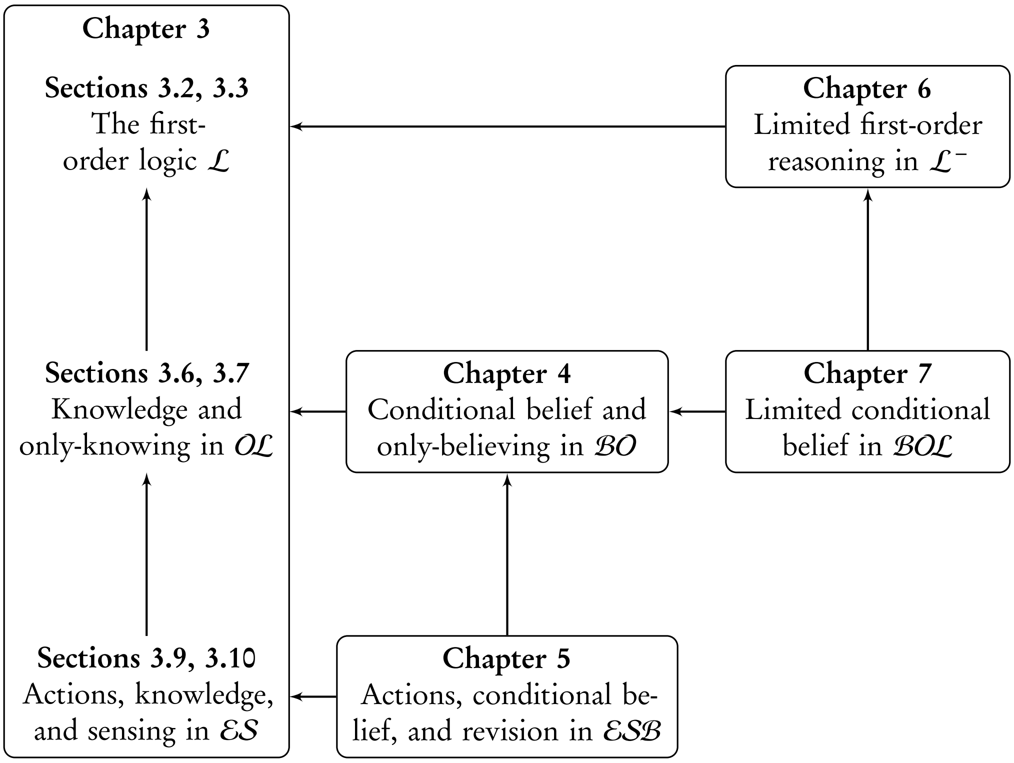


Figure 1.1: The dependency graph of the thesis.

logic of limited conditional belief, $\mathcal{BOL}$, presented in Chapter 7.

Most chapters conclude with a brief discussion and pointers to future work. A comprehensive conclusion of the thesis is finally drawn in Chapter 8. Numerous long proofs from Chapters 4, 5, 6, 7 are presented in Appendices A, B, C, D, respectively.

The survey of related work in Chapter 2 is not essential for the understanding of the rest of the thesis; the chapter can hence be skipped. The subsequent chapters however largely build on each other. Figure 1.1 gives an overview of these interdependencies.

2 Relevant Literature

This chapter surveys relevant literature from the different fields we touch in this thesis. We begin with related work on formalizations of knowledge and belief, which is the broad field this thesis belongs to. Following this, we first present belief revision frameworks and then theories of actions, upon which our epistemic situation calculus with beliefs is based. Finally we discuss ways to make first-order reasoning decidable, which is relevant to our work on limited reasoning.

Occasionally we use some logical notation in this chapter without covering the details. The reader not familiar with logical notation is referred to Sections 3.2 and 3.3 for an introduction to a first-order logic.

2.1 Knowledge and belief

The nature of knowledge has been the subject of debate among philosophers at least since Plato's dialogue *Theaetetus* in the 4th century BC (Chappell 2013). The importance of knowledge for artificial intelligence was recognized from its beginning. In a seminal paper, McCarthy (1959) envisions a program called Advice Taker whose behaviour improves just through new information it is told – reprogramming the system would be no longer required. The information is to be represented as declarative sentences, and the Advice Taker should deduce from this knowledge which action to take next.

Arguably the most popular approach for formal analysis of knowledge is by modal logic. Modern epistemic logic is in large part due to Hintikka (1962), who employed modal logic to represent knowledge and belief. Knowing and believing are called *modalities* (Halpern 1999). In classical modal logic, the main difference between knowledge and belief is that the former is factually true, while the latter is always consistent (Fagin, Halpern, et al. 1995). This understanding reflects one of the candidate definitions discussed by Plato: knowledge is justified true belief.

While originally a field of philosophy, modal logic is today also widely used in knowledge representation. (It is remarkable at this point that McCarthy (1997) opposed modal logic and preferred classical first-order logic instead. Halpern (1999) counters these arguments and argues that in many cases, propositional or first-order modal logic is the right tool to model modalities.)

A more general concept than knowledge or belief is that of *conditional belief*, which is the main theme of this thesis. We will employ it to stratify beliefs so as to reason about more or less likely contingencies. Our use of the concept is related to conditional logics, which add if-then statements closer to human intuition than classical logic's material implication. An overview of modal logic and conditional logic is given below.

Levesque and Lakemeyer (2001) emphasize that a knowledge-based systems needs to make available its knowledge to the rest of the system. To this end, Levesque (1984b) proposed a functional view that resembles a database interface and considers a knowledge representation system as a part of a larger knowledge-based system. As this view is at least implicitly present throughout this thesis, we detail it in this section as well.

2.1.1 Modal logic for knowledge

Modal logic extends propositional logic with a new unary operator that can be given many different interpretations. To illustrate the concept, we call the operator $\mathbf{K}$ and understand it as knowledge. The formulas of this language are of the form $\neg\alpha$, $(\alpha \vee \beta)$, $(\alpha \wedge \beta)$, $(\alpha \supset \beta)$, and $\mathbf{K}\alpha$.

The modern semantics of modal logic is due to Kripke (1959). Truth of a formula is defined with respect to a Kripke structure, which consists of a set of possible worlds e and an accessibility relation $\rightarrow\, \subseteq e \times e$. A possible world is a hypothetical truth assignments of all propositions. The accessibility relation links these worlds to each other.[1]

In modal epistemic logic, the link is interpreted as "considered possible:" if $w \rightarrow w'$, then the agent considers w' a possible world when w is the actual one. A sentence α is known when all accessible possible worlds satisfy α. More formally, $\rightarrow, e, w \models \mathbf{K}\alpha$ iff[2]

[1] We simplify the traditional definition of Kripke structures in order to be closer to the concept of worlds used in the rest of the thesis. The classical understanding of a Kripke structure separates the world and its valuation of propositions. That way, different worlds can have the same valuation, which is not possible in our definition. In languages with only a finite set of propositions, subtle differences between both definitions arise. For example, if there is only a single proposition P, then $P \wedge \mathbf{K}P \wedge \neg\mathbf{K}\neg P \wedge \mathbf{K}\neg\mathbf{K}P$ is satisfiable, namely by the classical Kripke structure where $e = \{w_1, w_2, w_3\}$, $w_1 \rightarrow w_2 \rightarrow w_3$ (and $w_i \nrightarrow w_j$ otherwise), and the valuation is such that P in w_1 and w_2 is true and false in w_3. Clearly, this construction implies $w_1 \neq w_2$. In our definition, there are only two worlds, namely the one that satisfies P and the one that does not, and therefore the sentence is not satisfiable. When there is an infinite supply of propositions, however, there are enough worlds available to replicate the same effect as in classical Kripke structures by using worlds which differ only in propositions that do not occur in the formula in question. Such issues will not be relevant in this thesis.

[2] We use "iff" as an abbreviation for "if and only if" throughout the thesis.

$\rightarrow, e, w' \models \alpha$ for all $w' \in e$ with $w \rightarrow w'$.

Modal logic can also be characterized axiomatically by taking the standard propositional axioms (Kleene 2002) and adding a selection of the following axioms for the modal operator **K**.

- K: $\mathbf{K}\alpha \wedge \mathbf{K}(\alpha \supset \beta) \supset \mathbf{K}\beta$;
- T: $\mathbf{K}\alpha \supset \alpha$;
- D: $\mathbf{K}\alpha \supset \neg\mathbf{K}\neg\alpha$;
- 4: $\mathbf{K}\alpha \supset \mathbf{K}\mathbf{K}\alpha$;
- 5: $\neg\mathbf{K}\alpha \supset \mathbf{K}\neg\mathbf{K}\alpha$.

An established convention is to name the logics after the axioms. A very frequent combination is K45, which is also referred to as S5 and asserts positive and negative introspection. KT45 is the logic of knowledge and KD45 is the logic of belief (Fagin, Halpern, et al. 1995): T specifies that knowledge is true, and D means that belief is consistent. Depending on which conditions the accessibility relation $\rightarrow$ satisfies, different modal logics are obtained.

Except for K, these axioms correspond to certain conditions on the semantic accessibility relation $\rightarrow \subseteq e \times e$. Namely, the condition for the axioms are as follows.

- T: $\rightarrow$ is reflexive, that is, for every w, $w \rightarrow w$;
- D: $\rightarrow$ is serial, that is, for every $w \in e$, there is some $w' \in e$ such that $w \rightarrow w'$;
- 4: $\rightarrow$ is transitive, that is, if $w_1 \rightarrow w_2$ and $w_2 \rightarrow w_3$, then $w_1 \rightarrow w_3$;
- 5: $\rightarrow$ is Euclidean, that is, if $w_1 \rightarrow w_2$ and $w_1 \rightarrow w_3$, then $w_2 \rightarrow w_3$.

A natural dual to the **K** operator is to refer to the *in*accessible worlds of a Kripke structure. We denote such an operator by **N**, and its semantics can be defined as $\rightarrow, e, w \models \mathbf{N}\alpha$ iff $\rightarrow, e, w' \models \alpha$ for all $w' \in e$ with $w \not\rightarrow w'$. $\mathbf{K}\alpha$ can be interpreted as knowing at least α, and $\mathbf{N}\neg\alpha$ as knowing at most α. Together, $\mathbf{K}\alpha \wedge \mathbf{N}\neg\alpha$ capture the intuition that *all* the agent knows is α, also referred to as *only-knowing*. Operators that refer to inaccessible worlds have been studied by Levesque (1990) and independently by Ben-David and Gafni (1989); the original idea is due to Humberstone (1987). Levesque (1990) considers a first-order language with infinitely many extra-logical symbols. Unlike general Kripke structures, models may not limit the set of considered worlds, and the (implicit) accessibility relation needs to adhere to the K45 constraints. Ben-David and Gafni (1989) go beyond that by allowing arbitrary modal logics with general Kripke structures. However, Levesque and Lakemeyer (2001) argue that this generality has counterintuitive consequences. For example, the sentence $\mathbf{K}P \land \mathbf{N}\neg P \supset \neg\mathbf{K}Q$, which stipulates the intuitively true statement that if all we know is P, then we do not know Q, is falsifiable by Kripke structures whose worlds all satisfy P and Q.

It is remarkable that adding such an operator that refers to the inaccessible worlds makes finding a proof theory much harder. In fact, as Halpern and Lakemeyer (1995) show, any such proof theory must be non-recursive in the first-order case. The argument is simple. For any non-modal formula α, validity of $\mathbf{N}\alpha \supset \neg\mathbf{K}\alpha$ coincides with α being falsifiable. Hence, if we had a recursive axiom system, by enumerating all valid formulas we would implicitly also enumerate all falsifiable non-modal formulas, which Church (1936a,b) and Turing (1936) showed to be impossible for first-order logic. Levesque (1990) proposed a non-recursive axiom system for his K45 logic of only-knowing which includes the axiom schema $\mathbf{N}\alpha \supset \neg\mathbf{K}\alpha$ for any non-modal α that is not valid. However, Halpern and Lakemeyer (1995) proved that Levesque's proof theory is actually incomplete for the first-order case, and as of today no complete axiom system is known (apart from trivial ones such as declaring every valid formula an axiom).

Levesque's logic of only-knowing will be present throughout this thesis. We shall introduce a couple of modal logics of knowledge, belief, and action. Following Levesque (1990), the operators for knowledge and belief will be K45 operators whereas truth of knowledge (T) or consistency of belief (D) will not be required. Our use of the terms knowledge and belief is a bit more informal: the term knowledge shall be seen as a hint that the agent considers this information very certain and indefeasible, whereas with belief the agent usually takes very well into account that she might be wrong.

2.1.2 Conditionals

Classical logic provides only one connective to make if-then statement, namely the material implication. The material implication $\alpha \supset \beta$ is equivalent to the disjunction $\neg\alpha \lor \beta$. As a consequence, when the antecedent is false, the implication is vacuously true.

Often, this is contrary to human intuition. Consider the following two conditionals.

- If polar bears lived in Antarctica, they would eat penguins.
- If penguins lived in the Arctic, they would eat polar bears.

The first one is certainly very plausible. But it is rather unthinkable that a group of penguins would bring down a polar bear. With material implications, however, both would come out true, since both antecedents are false. The false antecedent is what makes them special: they are *counterfactual*.

A similar example involves belief. In Example 1.1.1, after dropping the box and hearing a clink, we believe something inside the box broke. But we might also – less plausibly – believe that if there is something metallic in the box, this caused the clinking noise and presumably nothing is broken. This is a *conditional belief*: most plausibly we believe something else, but keep other possibilities and their consequents in mind. Often, conditional beliefs are also understood counterfactually, although the reference point here is not the real world but the most-plausible belief.

Similarly to epistemic logic, conditional logic has its roots in philosophy topic. Three early logical accounts of conditionals exist: Lewis (1973) and Stalnaker (1968) devise possible-worlds semantics for conditionals; Adams (1965) adopts probabilities for conditionals; Gärdenfors (1978) characterizes conditionals with belief revision techniques.

The approach we consider here is the notion of a *system of spheres*, due to Lewis (1973) and Grove (1988). A system of spheres consists of multiple sets of possible worlds. These spheres are totally ordered by a subset relation, with the rationale that the worlds from the narrowest sphere are the most plausible ones. A conditional $\alpha \Rightarrow \beta$ is evaluated in a system of spheres by testing if the most-plausible worlds that satisfy α also satisfy β. In Lewis' semantics of counterfactuals, spheres are relative to individual possible worlds.

It is easy to see that conditionals, other than material implications, are nonmonotonic: it is very well possible that $\alpha \Rightarrow \beta$ comes out true, but $\alpha \land \gamma \Rightarrow \beta$ might not. For example, if polar bears lived in Antarctica and were herbivores, they would not eat penguins. Similarly, counterfactual conditionals are also neither transitive nor does contraposition hold.

Among others, conditionals are helpful to determine causality (Lewis 1979). For example, dropping a box is a cause of things in the box breaking, because if we had not dropped the box, nothing would have broken. However, this counterfactual but-for alone is not sufficient to obtain the intuitively correct causes; more advanced models are needed (Halpern 2015; Halpern and Pearl 2005).

From the more practical perspective of artificial intelligence, conditionals are especially relevant due to their nonmonotonic nature (Kern-Isberner 2001). Pearl (1990) proposed System Z, a reasoning framework where conditionals are ranked in a specific, unambiguous way and nonmonotonic inferences can be drawn from them. Goldszmidt and Pearl (1996) extended System Z to account for qualitative probabilities.

Conditionals are central to this thesis – albeit not from a philosophical perspective but a rather pragmatic view. We will also briefly meet System Z again in Chapter 4, when we extend Levesque's logic of knowing and only-knowing to the case of conditional beliefs.

2.1.3 A functional view

Any knowledge-based system needs an interface to interact with the outside world. Levesque (1984b) characterizes a knowledge-based system as a service with (at least) two operations:

- $\mathbf{TELL}[\alpha, e] = e'$ computes a new state e' that incorporates the new knowledge α into the previous state e;
- $\mathbf{ASK}[\alpha, e] \in \{\text{yes}, \text{no}\}$ determines whether α is known in the current state e.

The importance of a **TELL** operation was stressed before already by McCarthy (1959) for the Advice Taker program.

Levesque (1984b) proposed a first-order $\mathsf{K45}$ logic to perspicuously capture the notion of the state e and new knowledge α. The state e is taken to be a set of worlds considered possible by the agent, which serves as semantic primitive of the logic: truth of knowledge is defined as $e, w \models \mathbf{K}\alpha$ iff $e, w' \models \alpha$ for all $w' \in e$. An accessibility relation as in Kripke structures is not needed; the $\mathsf{K45}$ constraints are satisfied implicitly by using the same set e on the right-hand side of the definition of $\mathbf{K}\alpha$. The obvious definitions of **TELL** and **ASK** are then

- $\mathbf{TELL}[\alpha, e] = e \cap \{w \mid e, w \models \alpha\}$;
- $\mathbf{ASK}[\alpha, e] = \text{yes}$ iff $e \models \mathbf{K}\alpha$.

The question arises if and how these semantic operations can be expressed syntactically. Levesque and Lakemeyer (2001) showed that in practice it is sufficient to restrict oneself to *representable* states. A state e is (finitely) representable when there is some (finite) set Φ of objective sentences, that is, sentences that mention no **K** operator, such that $e = \{w \mid w \models \phi \text{ for all } \phi \in \Phi\}$.

A seemingly simple solution for $\mathbf{ASK}[\alpha, \{w \mid w \models \phi\}]$ where ϕ is an objective sentence is to check if $\mathbf{K}\phi \supset \mathbf{K}\alpha$ is valid. However, this does not handle negative introspection correctly: while $\mathbf{ASK}[\neg\mathbf{K}Q, \{w \mid w \models P\}] = \text{yes}$, the sentence $\mathbf{K}P \supset \mathbf{K}\neg\mathbf{K}Q$ is not valid, because $\{w \mid w \models P \wedge Q\}$ satisfies $\mathbf{K}P$ and $\mathbf{K}Q$.

The problem is that $\mathbf{K}\phi$ alone does not capture the lack of knowledge. Only-knowing (Levesque 1990) addresses this problem: at least for objective sentences, it maximizes the epistemic state, and thus the agent's lack of knowledge. Formally, only-knowing α is defined as $e, w \models \mathbf{O}\alpha$ iff for all w', $w' \in e$ iff $e, w' \models \alpha$. As an equivalent alternative, we could take $\mathbf{O}\alpha$ as abbreviation for $\mathbf{K}\alpha \land \mathbf{N}\neg\alpha$ where the semantics of the new $\mathbf{N}$ operator is $e, w \models \mathbf{N}\alpha$ iff $e, w' \models \alpha$ for all $w' \notin e$; for this thesis, though, we prefer the all-in-one definition of $\mathbf{O}\alpha$.

With only-knowing, **ASK** and **TELL** can be characterized syntactically. For objective ϕ, $\mathbf{ASK}[\alpha, \{w \mid w \models \phi\}] = \text{yes}$ reduces to proving validity of $\mathbf{O}\phi \supset \mathbf{K}\alpha$. Moreover, for objective ϕ and ψ, $\mathbf{TELL}[\alpha, \{w \mid w \models \phi\}] = \{w \mid w \models \phi \land \psi\}$ reduces to proving validity of $\mathbf{O}\phi \supset (\mathbf{K}\alpha \equiv \mathbf{K}\psi)$ (Levesque and Lakemeyer 2001).

In this thesis, we have this view of knowledge representation as a service for other systems in mind. While we will not explicitly use **TELL** and **ASK** procedures, the concept is still present: the idea to represent a knowledge base by way of only-knowing is a core concept in this thesis, and actions will allow to modify these knowledge bases. The logical language sketched in the previous paragraphs are introduced formally in Chapter 3.

2.2 Belief revision

Belief revision theory addresses the question of adequately adjusting beliefs in the face of new information. Theories of belief change have emerged from philosophy, but have also drawn considerable interest in artificial intelligence.

Technically, most work on belief revision assumes a (propositional) logical language with a consequence relation $\models$ (Peppas 2008). Revision then operates on sets of sentences in that language which are closed under logical consequence. Such a set is called *belief set*.

The question of belief revision is then: when the agent's belief set is Φ, and she is now told a new piece of information ϕ, what is the new belief set $\Phi * \phi$? (Note the implicit assumption that the agent fully trusts the new information.)

When ϕ is consistent with Φ, the answer is simple: the new belief set $\Phi * \phi$ is the deductive closure of $\Phi \cup \{\phi\}$. This is called belief *expansion* and denoted by $\Phi + \phi$.

Often, however, ϕ contradicts some information in Φ. These contradicting beliefs need to be given up so that $\Phi * \phi$ can accommodate ϕ without being inconsistent. In general, there is more than one way to do so. For example, when we believe $P \land (P \supset \neg Q)$ and we are then told that Q holds, we could give up P, or $P \supset Q$, or both to make room for Q.

Informally, it has become evident already that belief revision is related to giving up beliefs. In belief revision theory, giving up belief in ϕ is called *contraction* and denoted by $\Phi \div \phi$. So one way to define revision is in terms of contraction followed by expansion:

$$\Phi * \phi = (\Phi \div \neg\phi) + \phi.$$

In English, this equation states that revising Φ by ϕ is the same as first giving up belief in $\neg\phi$ and then adding ϕ to it. This is called the Levi identity (Peppas 2008).

2.2.1 Postulates for belief revision

The Levi identity does not get us very far, unless we had a more precise definition of contraction. In their seminal paper, Alchourrón, Gärdenfors, and Makinson (1985) gave the answers to both, revision and contraction. They characterized the revision $\Phi * \phi$ by means of eight postulates that every operator $*$ should satisfy, and similarly for the contraction $\Phi \div \phi$. The underlying principle of these postulates is that of *minimal change*, meaning that the agent shall modify her beliefs as little as possible. These postulates are typically referred to as AGM postulates by the authors' initials.

The AGM postulates stipulate that for any belief set Φ and any formulas ϕ, ψ, υ, the following conditions are met.

AGM1. If $\Phi * \phi \models \psi$ and $\models (\psi \supset \upsilon)$, then $\Phi * \phi \models \upsilon$.

AGM2. $\Phi * \phi \models \phi$.

AGM3. If $\Phi * \phi \models \psi$, then $\Phi + \phi \models \psi$.

AGM4. If $\Phi \not\models \neg\phi$ and $\Phi + \phi \models \psi$, then $\Phi * \phi \models \psi$.

AGM5. If $\Phi \not\models \textsc{false}$, then $\Phi * \phi \not\models \textsc{false}$.

AGM6. If $\models (\phi \equiv \psi)$, then $\Phi * \phi \models \upsilon$ iff $\Phi * \psi \models \upsilon$.

AGM7. If $\Phi * (\phi \land \psi) \models \upsilon$, then $(\Phi * \phi) + \psi \models \upsilon$.

AGM8. If $\Phi * \phi \not\models \neg\psi$ and $(\Phi * \phi) + \psi \models \upsilon$, then $\Phi * (\phi \land \psi) \models \upsilon$.

Here we slightly differ in presentation from the usual way, where preferably subset relations are used. For example, the third postulate is typically expressed as $\Phi * \phi \subseteq \Phi + \phi$. As the sets Φ, $\Phi * \phi$, and $\Phi + \phi$ are closed under $\models$, both ways are clearly equivalent.

These postulates are constraints for possible revision operators. Interestingly, there are multiple ways to constructively define $\Phi * \phi$ in accordance with the AGM postulates: selection functions (Alchourrón, Gärdenfors, and Makinson 1985), epistemic entrenchment (Gärdenfors and Makinson 1988), and systems of spheres (Grove 1988; Lewis 1973). Here we focus on system of spheres, which we already encountered in Section 2.1.2.

Roughly, a system of sphere is a collection of sets of possible worlds $\vec{e} = \langle e_1, e_2, \ldots \rangle$ such that $e_p \subseteq e_{p+1}$, where a world is a truth assignment of all propositions. The subset relation ranks these spheres by plausibility: the worlds in the narrowest set are the most plausible. This plausibility ranking is the extra-logical information needed for revision.

Let Φ be a belief set and let $\vec{e} = \langle e_1, e_2, \ldots \rangle$ be a system of spheres centred on Φ, that is,

$$e_1 = \{w \mid w \models \phi \text{ for all } \phi \in \Phi\}.$$

Then the following $\Phi * \phi$ satisfies the AGM postulates (Grove 1988):

$$\Phi * \phi = \{\psi \mid w \models \psi \text{ for all } w \in e_{p^*} \text{ with } w \models \phi\},$$

where e_{p^*} is the narrowest sphere of $\vec{e}$ that is consistent with ϕ, that is, $w \models \phi$ for some $w \in e_{p^*}$, but $w \not\models \phi$ for all $w \in e_p$ and $p < p^*$.

Again, we used slightly different notation and terminology than the belief revision literature, where deductively closed sets of formulas are preferably used instead of the semantic concept of worlds. The reader will encounter the concept of worlds and system of spheres in the upcoming chapters again, where they are introduced more formally.

2.2.2 Postulates for iterated belief revision

The AGM postulates only refer to single revision and leave open the question of how to characterize repeated revision $(\Phi * \phi) * \psi$. In terms of systems of spheres, the trouble is that after the first revision, the extra-logical plausibility ranking on the worlds, which determines the result of revision, is lost. Darwiche and Pearl (1997) and Nayak, Pagnucco, and Peppas (2003) proposed amendments to the AGM postulates to address this shortcoming.

To retain the extra-logical plausibility ranking, Darwiche and Pearl define revision as an operation not only on belief sets, but on *epistemic states*, which augment a belief set with a plausibility structure. An epistemic states can be represented as a system of spheres. We write $\vec{e} \models \psi$ to say that the belief set associated with $\vec{e}$ entails ψ: $\{\phi \mid w \models \phi \text{ for all } w \in e_1\} \models \psi$, or equivalently: $w \models \psi$ for all $w \in e_1$.

In this context, revision operators accordingly map epistemic states to epistemic states. Darwiche and Pearl (1997) demand such operators to satisfy the AGM postulates (with the belief set Φ substituted by an epistemic state $\vec{e}$), plus the following postulates, where $\vec{e}$ is an epistemic state and ϕ, ψ, υ are formulas.

DP1. If $\phi \models \psi$, then $(\vec{e} * \phi) * \psi \models \upsilon$ iff $\vec{e} * \psi \models \upsilon$.

DP2. If $\phi \models \neg\psi$, then $(\vec{e} * \phi) * \psi \models \upsilon$ iff $\vec{e} * \psi \models \upsilon$.

DP3. If $\vec{e} * \psi \models \phi$, then $(\vec{e} * \phi) * \psi \models \phi$.

DP4. If $\vec{e} * \psi \not\models \neg\phi$, then $(\vec{e} * \phi) * \psi \not\models \neg\phi$.

The field of iterated revision is a highly disputed one, and a large variety of postulates has been proposed. While the above Darwiche–Pearl postulates certainly are the most prominent framework for iterated revision, it has been criticized for both, being too permissive and too strong.

An alternative approach has been proposed by Nayak, Pagnucco, and Peppas (2003). They stipulate a new postulate that says that the order of any two consistent pieces of evidence is irrelevant, which is not necessarily true from the Darwiche–Pearl postulates. This is captured by the third postulate of their framework.

NPP1. If $\vec{e} \models \textsc{false}$, then $\vec{e} * \phi \models \psi$ iff $\models (\phi \supset \psi)$.

NPP2. AGM1–AGM6 hold for $\vec{e}$.

NPP3. If $\not\models \neg(\phi \wedge \psi)$ then $(\vec{e} * \phi) * \psi \models \upsilon$ iff $\vec{e} * (\phi \wedge \psi) \models \upsilon$.

NPP4. If $\models (\psi \supset \neg\phi)$ and $\not\models \neg\phi$, then $(\vec{e} * \phi) * \psi \models \upsilon$ iff $\vec{e} * \phi \models \upsilon$.

Originally, Nayak, Pagnucco, and Peppas (2003) modified the definition of the belief revision operator: they understand it as a unary function from sentences to belief sets, which inherently contain the current belief set already, and which are dynamic, that is, revision not only produces a new belief set but also a new revision function. To ease the comparison, we phrased their postulates using epistemic states above. Jin and Thielscher (2007) in turn criticized NPP3 for being overly strong and proposed an alternative extension of the Darwiche–Pearl framework.

On the other hand, it has been claimed that Darwiche–Pearl postulates are too strong. In particular, DP2, which says that of two contradictory pieces of evidence the later one shall prevail over the earlier, is considered to lead to undesirable results and at odds with the AGM postulates (Peppas 2008). Nayak, Pagnucco, and Peppas addressed this their fourth postulate. Delgrande and Jin (2012) remedy this by extending revision to take not only a formula as new evidence but a set of independent formulas.

2.2.3 Revision operators

Along with the large variety of postulate systems, many different operators that adhere to these postulates have been proposed. We first investigate revision operators that do not rely on numeric ranking of possible worlds and thus can be represented, for example, with a system of spheres.

A conceptually very simple revision operator is *natural revision* (Boutilier 1993, 1996). Given an epistemic state $\vec{e}$ and a piece of evidence ϕ, the idea is revise $\vec{e}$ by making the most-plausible worlds that satisfy ϕ more plausible than any other world in $\vec{e}$. Natural revision places only little trust in ϕ: in general, a new piece of evidence ψ may already lead to retraction of ϕ when none of the most-plausible ϕ worlds satisfies ψ.

Deeper trust in new information can be modelled by *lexicographic revision* (Nayak 1994; Nayak, Pagnucco, and Peppas 2003). It leads to a more profound change of the worlds' plausibility ranking: lexicographic revision by ϕ promotes *all* ϕ-worlds over all $\neg\phi$-worlds, while their respective relative orders are retained. That way, for the agent to give up ϕ, a subsequent evidence ψ must be falsified by *all* ϕ-worlds, not just the most-plausible ones as in natural revision.

The revision operators we consider in this thesis are natural and lexicographic revision. An extensive analysis of similar operators can be found in (Rott 2009).

An alternative framework for revision operators are the *ordinal conditional functions* (Spohn 1988). An ordinal conditional function κ maps a possible worlds to a natural number, which can be taken as their plausibility. Some worlds need to have plausibility 0, which marks them out as most plausible. The ranking extends to sentences ϕ by $\kappa(\phi) = \min\{\kappa(w) \mid w \models \phi\}$. Spohn (1988) presents a general scheme to update κ by some new evidence ϕ, so that $\neg\kappa$ is less plausible than m after the revision:

$$(\kappa * (\phi, m))(w) = \begin{cases} \kappa(w) - \kappa(\phi) & \text{if } w \models \phi; \\ \kappa(w) - \kappa(\neg\phi) + m & \text{otherwise.} \end{cases}$$

Spohn denotes $\kappa * (\phi, m)$ as (ϕ, m)-*conditionalization*. By choosing specific values for m, different operators are obtained. For example, Darwiche and Pearl (1997) show that for $m = \kappa(\neg\phi) + 1$, (ϕ, m)-conditionalization satisfies their postulates. Note that similar to lexicographic revision, the relative order among the ϕ-worlds is left unchanged, and likewise for the $\neg\phi$-worlds. Due to the numeric confidence, conditionalization bears some similarity to probabilistic conditioning.

2.2.4 Belief update

Belief revision à la AGM is not suited to reflect how beliefs evolve in the face of physical change. For example, suppose all we believe is InBox(gift) ≡ Broken(gift), and we now deposit the gift in the box. Not differentiating between revision and update, we would revise the knowledge base by InBox(gift). According to AGM, this revision shall be a simple expansion. The resulting revised knowledge base is therefore (InBox(gift) ≡ Broken(gift)) ∧ InBox(gift), which is logically equivalent to InBox(gift) ∧ Broken(gift). The AGM postulates hence let putting the gift into the box have an awkward side-effect: it breaks the gift!

Katsuno and Mendelzon (1991) were the first to realize this shortcoming of the AGM postulates. Instead of modifying the AGM framework, they argue to distinguish between belief revision and belief update, and propose a new set of postulates for update operators $\diamond$. We say a set of sentences Φ is *complete* when for every ϕ, either $\Phi \models \phi$ or $\Phi \models \neg\phi$. Then the postulates for any belief set Φ and any formulas ϕ, ψ, υ, the postulates are as follows.

KM1. If $\Phi \diamond \phi \models \psi$ and $\models (\psi \supset \upsilon)$, then $\Phi \diamond \phi \models \upsilon$.

KM2. $\Phi \diamond \phi \models \phi$.

KM3. If $\Phi \models \phi$, then $\Phi \diamond \phi \models \psi$ iff $\Phi \models \psi$.

KM4. If $\Phi \not\models$ FALSE and $\phi \not\models$ FALSE, then $\Phi \diamond \phi \not\models$ FALSE.

KM5. If $\models (\phi \equiv \psi)$, then $\Phi \diamond \phi \models \upsilon$ iff $\Phi \diamond \psi \models \upsilon$.

KM6. If $\Phi \diamond (\phi \wedge \psi) \models \upsilon$, then $(\Phi \diamond \phi) + \psi \models \upsilon$.

KM7. If $\Phi \diamond \phi \models \psi$ and $\Phi \diamond \psi \models \phi$, then $\Phi \diamond \phi \models \upsilon$ iff $\Phi \diamond \psi \models \upsilon$.

KM8. If Φ is complete and $\Phi \diamond (\phi \vee \psi) \models \upsilon$, then $(\Phi \diamond \phi) \cup (\Phi \diamond \psi) \models \upsilon$.

KM9. $\Phi \diamond \phi \models \upsilon$ iff for all complete $\Psi \supseteq \Phi$, $\Psi \diamond \phi \models \upsilon$.

This presentation is a rewording of the one from (Peppas 2008) in our terminology.

The Katsuno–Mendelzon postulates consider change of a different type than we do in this thesis. In their framework, some formula ϕ is made true by some physical action, but there is no encoded law how this works and which action caused this effect. For example, there is no explicit way of saying that dropping a box breaks every fragile object in it.

By contrast, this thesis is concerned with the model of change known from action theories, whose goal is to explicitly represent how actions affect their environment. In particular, this allows conditional effects such as dropping the box breaking only fragile items in the box. On the other hand, in the Katsuno–Mendelzon framework an action can make a sentence $P \vee Q$ true. In the type of action theory we shall consider in this thesis, the so-called situation calculus basic action theories (Reiter 1991, 2001), such nondeterministic change cannot be modelled.

2.3 Actions and change

Commonsense reasoning requires an understanding of actions and change. For example, it is generally plausible that a fragile object breaks when it is dropped. McCarthy (1963) was the first to recognize this need. He envisioned a logical theory of actions and change, which he called situation calculus, to be part of the (fictional) Advice Taker program (McCarthy 1959).

The situation calculus quickly turned out to be too complex for real-world applications. As a consequence, the first practical planning system STRIPS (Fikes and Nilsson 1972, 1993) employed a much more restricted but also computationally more efficient language. Today's planning systems use a modern and more expressive version of this language called PDDL (Helmert 2006; Hoffmann and Nebel 2001; McDermott et al. 1998).

In the meantime, research on action theories made progress towards computational feasibility as well. A milestone was when Reiter (1991) presented a simple solution for the frame problem, based on earlier work by Haas (1987), Pednault (1989), and Schubert (1989). Related formalisms are the fluent calculus (Thielscher 1998) and the event calculus (Kowalski and Sergot 1989). Thanks to the use of first-order logic these languages are highly expressive. Moreover there is the propositional family of action languages $\mathcal{A}$ (Gelfond and Lifschitz 1993, 1998) based on logic programs.

We are here mostly concerned with Reiter's situation calculus, in particular with a modal variant thereof (Lakemeyer and Levesque 2011). Its predominant position among today's theories of action is reflected by Reiter's monograph (Reiter 2001) and a chapter in the knowledge representation handbook (Lin 2008).

2.3.1 Problems

Since the conception of McCarthy's situation calculus, three foundational problems associated with actions and change have been identified, namely the frame, ramification, and qualification problems. These problems, in one way or another, must be addressed by any action formalism.

The frame problem

The frame problem is to specify the non-effects of an action. The effect of an action is typically limited to a small set of properties and can hence be represented with reasonable effort. Much more trouble is to stipulate what does *not* change, due to the sheer number of non-effects.

For example, it is a reasonable assumption that when we drop the box, everything fragile in the box breaks. But almost everything else would remain the same as before: the same items as before are in the box, the weather is still the same, kangaroos are still hopping around in Australia, and so on.

The problem is one of the most famous in artificial intelligence and was first identified by McCarthy and Hayes (1969). It got its name from the frame axioms which are supposed to capture the non-effects of an action. The naive approach is to specify *frame axioms* for every non-effect. In general, this is not feasible due to the vast number of non-effects. McCarthy (1986) hence proposed the use of circumscription (McCarthy 1980), a form of nonmonotonic reasoning. A simpler solution in the situation calculus was found by Reiter (1991), which uses classical logical equivalence to capture an action's effects and non-effects at once.

The ramification problem

The ramification problem is about the indirect effects an action may have due to domain constraints.

As an example of a domain constraint we might say that we have control over things we have in our hand and everything attached to it. Thus, when we are holding a box, we also have control over the objects in it. But once we drop the box, we also lose control over them as an indirect effect.

The ramification problem was first discussed by Finger (1987). In certain cases ramification constraints can be compiled to direct effects (McIlraith 2000). Lin (1995) proposes another solution using a variant of the situation calculus where state constraints as well as ordinary action effects are encoded with a notion of causality. Thielscher (1997) also uses causal relationships for state constraints in the fluent calculus.

The qualification problem

The qualification problem is to capture the preconditions of an action. Besides a few major preconditions for an action, there is usually a large number of minor qualifications that must satisfied so that an action is possible, but usually these minor qualifications assumed to hold. The problem gets more complicated when state constraints affect the precondition.

For example, we can drop a box only when we are holding it. But there is a long list of other (rather improbable) conditions which make it impossible to drop the box, even if we are holding it: when there is no gravity, when the box is glued to our hands, when we have turned into stone, and the list goes on.

The qualification problem was first observed by McCarthy (1977). He proposed to use circumscription to make action preconditions true "unless something prevents it."

2.3.2 The projection problem

If solving the frame, ramification, and qualification problems ought to be more than an end in themselves, action theories must serve another purpose than just studying their own problems. From an artificial intelligence point of view, arguably the most fundamental task of an action theory is *projection*. Roughly, the projection problem is determine what is true and what is not after some actions.

Reiter (2001) defines projection more formally as follows: given a sequence of actions and a logical formula, the projection problem is to decide if the formula holds after executing the actions. There are two cardinal, dual approaches to the projection problem: regression and progression.

Projection by regression

Regression rewrites the original formula, the query, to roll back the effect of actions. Regression is often a very elegant mechanism to eliminate actions from the reasoning task. On the downside, the regressed query may grow exponentially in the number of actions. The procedure is hence not suited for long-lived systems that amass a huge number of actions.

The idea of regression in the context of artificial intelligence is due to Waldinger (1981). Reiter (1991) introduced goal regression along with his solution to the frame problem for the situation calculus. Reiter's regression operator also carries over to epistemic extensions of the situation calculus (Lakemeyer and Levesque 2011; Scherl and Levesque 2003).

Projection by progression

Progression applies the effects of an action to a knowledge base. Surprisingly, progression is much more complex than regression – in general, progressing a first-order knowledge base requires second-order logic. Fortunately, for specific problem classes progression is first-order definable and even computable.

Interpreting STRIPS operators (Fikes and Nilsson 1972) as a mapping from one first-order theory to another gives rise to the notion of progression. In a seminal paper, Lin and Reiter (1997) gave a semantic account of progression in the situation calculus. They proved that this type of progression is not generally first-order definable. Later Vassos and Levesque (2013) confirmed this negative result for arbitrary models of progression. Computationally more attractive problem classes have been analysed in (Lin and Reiter 1997; Liu and Lakemeyer 2009; Vassos, Lakemeyer, and Levesque 2008). Lakemeyer and Levesque (2009) investigated the progression of knowledge in epistemic situation calculus.

2.3.3 Theories of action

So far we only discussed problems related to action theories. In the following, we give brief introductions to the most popular action formalisms in use today: Reiter's situation calculus, the fluent calculus, the event calculus, and the family of action languages $\mathcal{A}$.

Situation calculus

The situation calculus (McCarthy 1963) was the first action formalism. It introduced the concept of *fluents*, predicates or functions whose value may change over the course of actions. To facilitate this in classical logic, the situation calculus reifies world states, called *situations*, which then occur as additional arguments in fluents.

As mentioned above, the original situation calculus was notoriously computationally infeasible, and it largely remained like that until Reiter's solution of the frame problem and his variant of the situation calculus (Reiter 1991). Whereas McCarthy and Hayes took a situation to be a "complete state of the universe at an instant of time," Reiter models them as a sequences of actions that occurred since some initial situation, that is, as "world histories." Logically, they are represented as the initial situation S_0 and successor situations $\mathrm{do}(a, s)$ resulting from doing action a in situation s.

To solve the frame problem, Reiter assumes a set of positive effect axioms of the form $\gamma_i^+(\vec{x}, a, s) \supset F(\vec{x}, \mathrm{do}(a, s))$ and negative effect axioms $\gamma_j^-(\vec{x}, a, s) \supset \neg F(\vec{x}, \mathrm{do}(a, s))$ that state under which conditions an action a makes F true or false, respectively. A *causal completeness assumption* then allows to condense these axioms to a single *successor-state axiom* of the form

$$F(\vec{x}, \mathrm{do}(a, s)) \equiv \bigvee_i \gamma_i^+(\vec{x}, a, s) \vee \big(F(\vec{x}, s) \wedge \neg \bigvee_j \gamma_j^-(\vec{x}, a, s)\big).$$

Intuitively, this axiom says that $F(\vec{x}, \mathrm{do}(a, s))$ is true iff it was made true by a or it was true before already and was not made false by a. For example, we could have a successor state axiom

$$\mathrm{Broken}(y, \mathrm{do}(a, s)) \equiv \mathrm{Broken}(y, s) \vee \mathrm{InBox}(y, s) \wedge \mathrm{Fragile}(y, s) \wedge a = \mathrm{dropbox}$$

to say that dropping a box breaks the fragile items contained in it.

Successor-state axioms are a remarkably concise representation. Adding a new fluent merely requires to add one new successor-state axiom, and adding a new action merely takes to adjust those fluents' successor-state axioms which are affected by that action.

Successor-state axioms are also key to Reiter's regression operator (Reiter 1991, 2001) for the frame problem. The idea is, roughly, to iteratively replace occurrences of fluents $F(\vec{x}, \mathrm{do}(a, s))$ with the right-hand side of F's successor-state axiom. That way, actions are successively eliminated.

Reiter's solution is based on earlier work by Haas (1987), Pednault (1989), and Schubert (1989), and bears some resemblance to Clark's completion for logic programs (Clark 1978). Successor state axioms are the key ingredient of Reiter's *basic action theories* and his regression operator.

Reiter's situation calculus was a starting point of a whole new research agenda. Among the notable results are the action language Golog (Levesque, Reiter, et al. 1997), intended as a hybrid between traditional programming and planning, and epistemic extensions (Lakemeyer and Levesque 2011; Scherl and Levesque 2003). We present the epistemic situation calculus in detail in Chapter 3.

Fluent calculus

The fluent calculus (Thielscher 1998) is a variant of Reiter's situation calculus which has an explicit representation of world *states*. It inherits Reiter-style situations from the situation calculus, but associates every situation with a state, which in turn is a collection of reified fluents.

Other than successor-state axioms, which are fluent-wise, the axioms in the fluent calculus are action-wise and represent how the state changes when an action is performed. The binary function $\circ$ is used to express which fluents are known to be true in a state. Using the functions State and $\circ$, *state-update axioms* characterize how states are affected by actions in a progression-like fashion. For example,

$$\mathrm{Holds}(\mathrm{Fragile}(\mathrm{gift}) \land \mathrm{InBox}(\mathrm{gift}), s) \supset \mathrm{State}(\mathrm{do}(\mathrm{dropbox}, s)) = \mathrm{State}(s) \circ \mathrm{Broken}(\mathrm{gift})$$

says that when the gift is in the box and fragile, then it is broken after dropping the box. Note that this example state-update axiom does not capture the intuitive effect of dropbox, which should be that *every* fragile item in the box breaks, not just the gift, as axiomatized in the example successor-state axiom above.

Indeed representing infinitary effects such as dropbox breaking everything fragile in the box is quite cumbersome in the fluent calculus (Thielscher 1999). Nevertheless it is noteworthy that in the absence of infinitary effects situation calculus basic action theories can be translated to the fluent calculus.

Like in the situation calculus, a number of extensions to the fluent calculus have been developed, including epistemic ones (Jin and Thielscher 2004; Thielscher 2000). With FLUX (Thielscher 2005) there is a programming system similar to Golog.

Event calculus

The event calculus (Kowalski 1992; Kowalski and Sergot 1989) is a more distant relative of Reiter's situation calculus. Unlike the branching structure of situations used in the situation and fluent calculus, the event calculus is based on a narrative of events that happen on a linear and continuous time scale. For example, Happens(dropbox, t) stipulates that the box is dropped at time point t.

Effects of events on this time scale are axiomatized by stipulating which events initiate or terminate certain fluents. For example, the effect of dropping the box would be

$$\mathrm{HoldsAt}(\mathrm{InBox}(y) \land \mathrm{Fragile}(y), t) \land \mathrm{Initiates}(\mathrm{dropbox}, \mathrm{Broken}(y), t).$$

The frame problem is then solved by circumscribing (McCarthy 1980) the foundational predicates Initially, Happens and Happens for the narrative of events and the predicates Initiates and Terminates for the action effects.

A number of different variants of the event calculus exist. The formalism has also been extended to for epistemic reasoning (Miller, Morgenstern, and Patkos 2013).

Action language $\mathcal{A}$

The family of action languages $\mathcal{A}$ (Gelfond and Lifschitz 1993, 1998) differs considerably from the approaches described above. They come in two sorts: action description languages and actions query languages. Unlike the first-order situation, event, and fluent calculus, these languages are propositional.

Action description languages determine a transition system where actions lead to new states. Such transition systems are the semantic primitive of $\mathcal{A}$ and its descendants. For example, the action description of our dropbox action would read as

$$\text{dropbox } \textbf{causes} \text{ Broken } \textbf{if} \text{ InBox, Fragile.}$$

Descendants of the action language $\mathcal{A}$ also allow more complex action descriptions, such as state constraints and defaults.

Action query languages then operate on such transition systems. Given a couple of axioms such as **now** InBox and **now** Fragile, we can infer

$$\textbf{necessarily} \text{ Broken } \textbf{after} \text{ dropbox.}$$

The query is evaluated on the transition system in a straightforward way: the query holds iff Broken is true in all states reached by dropbox from any state initial state that satisfies InBox and Fragile.

2.3.4 Actions and belief revision

A number of belief revision extensions of the situation calculus have been proposed (Delgrande and Levesque 2012; Demolombe and Pozos Parra 2006; Fang and Liu 2013; Fang, Liu, and Wen 2015; Shapiro et al. 2011). The fundamental differences to our proposal are the following: firstly, no other approach allows to fully capture the idea of a knowledge base as we do with only-believing; secondly, with the exception of (Fang, Liu, and Wen 2015) none of them addresses the belief projection problem; thirdly, only Delgrande and Levesque (2012) employ a traditional belief revision scheme; and finally, none of the approaches seems to support quantifying-in.

Despite these differences, the work by Shapiro et al. (2011) is closest to our approach. They too use conditional beliefs to determine the initial beliefs. Without only-believing, however, this is quite cumbersome in practice and often neither more concise nor more intuitive than specifying a plausibility ranking by hand. Another major difference is that Shapiro et al. assume perfectly accurate sensors and thus cannot handle contradictory information.

The next-closest relative is by Delgrande and Levesque (2012). In this formalism, actions can inform the agent that some information is true like in our approach; new information is then incorporated by a revision scheme based on Spohn's ranking functions (Spohn 1988). Among the mentioned approaches, this is the only one that follows a traditional revision scheme like ours. However, the work is focused on modelling fallible actions; it is not concerned with belief projection.

Fang and Liu (2013) use plausibility rankings on worlds and actions; the plausibility ranking then changes according to the executed actions. Both rankings are explicitly specified in the beginning. Fang, Liu, and Wen (2015) also consider progression, however only in the propositional case.

The proposal by Demolombe and Pozos Parra (2006) avoids any plausibility ranking by compiling physical and epistemic effects on predetermined beliefs of interest to special successor-state axioms. As we have shown here, though, ranking the worlds by plausibility exclusively in the semantics is sufficient; plausibilities need not be part of the language.

Another framework to deal with faulty sensors is the Bayesian approach by Bacchus, Halpern, and Levesque (1999). They also use extra action parameters to indicate the real-world outcome, similar to our mimicking of classical sensing described in Section 5.11.

Belief revision, albeit not with physical actions, has also been addressed in dynamic epistemic logic (van Ditmarsch, van der Hoek, and Kooi 2007) by several authors (Aucher 2005; Baltag and Smets 2008; van Benthem 2007; van Ditmarsch 2005). Closest to our work is the approach by van Benthem (2007), where revised beliefs are reduced to initial beliefs in a regression-like fashion.

2.4 Decidable first-order reasoning

The problem of deciding whether a formula is valid in first-order logic is undecidable (Church 1936a,b; Turing 1936). This puts serious bounds on the practical utility of first-order logic. In particular, determining what follows from a knowledge base is undecidable. Starting from this negative result, research evolved in several directions to achieve decidable or even tractable reasoning. There are two possible directions to keep decidable:

- restrict the language expressiveness;
- restrict the inference capability.

Our approach is hybrid; it arguably chiefly belongs to the latter class but also makes syntactic restrictions to the knowledge base.

2.4.1 Restricting the language

The syntactic perspective on decidability in first-order logic is to categorize formulas by syntactic features in classes. Such a class is considered decidable if for every formula that belongs to that class satisfiability is decidable. The following overview is largely based on (Börger, Grädel, and Gurevich 1997), the standard reference for decidability and undecidability of syntactic fragments of first-order logic.

Prefix-vocabulary classes

A classical approach of syntactic restriction is by investigating prefix-vocabulary classes. Formulas in prenex normal form, that is, formulas where quantifiers occur only at the outermost level, are categorized by their quantifier string, the number and arity of function and predicate symbols they mention, and whether they mention equality. An example of a decidable class is the Bernays–Schönfinkel class of formulas, which are in prenex normal form with a quantifier prefix of the form $\exists^*\forall^*$, that is, arbitrarily many existential quantifiers are followed by arbitrarily many universal quantifiers, and mention no functions but possibly equality.

Adding or taking away equality does make a difference for other classes. The Gödel–Kalmár–Schütte class contains the formulas with quantifier prefixes of the form $\exists^*\forall\forall\exists^*$ that mention neither functions nor equality. This class is decidable. But adding equality makes the class undecidable in the presence of a single binary predicate.

As Grädel, Kolaitis, and Vardi (1997) put it, today “the dividing line between decidability and undecidability for all prefix-vocabulary classes” is identified: there are nine minimal undecidable classes without functions and equality, seven minimal undecidable classes with functions or equality, and seven maximal decidable classes (Börger, Grädel, and Gurevich 1997).

Bounded-variable logics

An alternative way is to restrict the number of variables. These sentences need not be in prenex normal form; in fact, the same variable may be quantified over and over again.

The fragment of first-order logic restricted to two variables and no functions is known to be decidable (Mortimer 1975; Scott 1962). However, three variables and one binary predicate already lead to undecidability. An interesting and still decidable extension especially for knowledge representation is the two-variable logic with counting quantifiers $\exists^{\leq n}$, $\exists^{\geq n}$, which require the existence of at least or at most n distinct objects.

The decidability of two-variable logic connects first-order logic with propositional modal logic (Grädel, Kolaitis, and Vardi 1997) and the prototypical description logic $\mathcal{ALC}$ (Nardi and Brachman 2003). Typical description logics are more expressive than propositional logic but aim to avoid the complexity of first-order logic. The three main modelling primitives are *concepts*, *roles*, and *individuals*, which correspond to unary predicates, binary predicates, and constants in first-order logic.

Interestingly, the problem of satisfiability is simpler in $\mathcal{ALC}$ (where it is PSPACE-complete (Baader and Nutt 2003)) than in the two-variable fragment of first-order logic (where it is NEXPTIME-complete (Grädel, Kolaitis, and Vardi 1997)). A considerable zoo of description logics builds on $\mathcal{ALC}$ and its decidability and on the two-variable logic with equality or counting quantifiers (Sattler, Calvanese, and Molitor 2003); these logics populate a wide area of the efficiency-vs-decidability spectrum.

Bounded-extension logics

The decidability results for most (but not all) prefix-vocabulary classes as well as for the two-variable fragment are closely connected to finite-model properties (Börger, Grädel, and Gurevich 1997; Libkin 2013). A class of formulas is said to have the finite-model property when all satisfiable formulas have a model with a finite universe of discourse.

To exemplify the relevance of finite-model properties, consider a formula from the Bernays–Schönfinkel class, say $\exists x_1 \ldots \exists x_m \forall y_1 \ldots \forall y_n \phi$ for quantifier-free ϕ. Any model of this formula contains objects that witness the existentially quantified $x_1, \ldots, x_m$. Limiting the universe of discourse to these $\leq m$ objects then yields a finite substructure that satisfies the formula. So a simple decision procedure for the Bernays–Schönfinkel class is to generate all nonisomorphic structures with $\leq m$ individuals and check if any of them satisfies the formula (Libkin 2013).

A finite-model property can of course also be enforced axiomatically by bounding the extensions of all predicates of interest to contain no more than N different tuples. In many realistic scenarios such a bound appears reasonable. For example, in our gift-giving scenario we might say that our box can contain at most one object at a time. Based on this idea, De Giacomo, Lespérance, and Patrizi (2016) introduce *bounded* basic action theories in Reiter's situation calculus, where in any reachable situation every fluent has at most N tuples in its extension. It is remarkable that boundedness only concerns the extensions; the universe of discourse remains potentially infinite.

Boundedness implies that it is sufficient to consider certain finite models in the individual situations, and hence verifying first-order formulas locally is decidable. De Giacomo, Lespérance, and Patrizi further prove decidability for an expressive class of temporal formulas that includes controlled quantification across situations.

They also provide sufficient criteria under which actions preserve the boundedness condition provided that the initial situation is bounded. One method is to simply condition action executability on the resulting situation being bounded. For instance, we could allow the agent to put a new object into the box only when the box is empty. Other ways to obtain boundedness are to ensure that actions do not enable more fluents than they disable, or that tuples fade away gradually.

2.4.2 Restricting inference

An alternative way to achieve decidability is to limit the reasoning capabilities. Besides undecidability, there is a second motivation to do so: the problem of *logical omniscience* (Hintikka 1975). In the possible-worlds semantics, when the agent knows α, and α logically implies β, then the she also knows β. To solve the omniscience problem, this closure under logical consequence must be suspended, at least to some extent.

Semantic approaches based on tautological entailment

One of the most influential approaches to the problem of omniscience is the logic of *implicit and explicit belief* due to Levesque (1984a). Implicit belief follows the usual possible-worlds semantics and is omniscient. If ϕ is the agent's knowledge base, then implicit belief contains all consequences of ϕ, for this characterizes what the world would be like if ϕ was actually true. This is not to say that the agent actually is aware of these consequences.

For that purpose, Levesque introduces explicit belief. As more worlds lead to fewer beliefs, Levesque introduces additional impossible worlds by resorting to a four-valued semantics, where an atom may be true, or false, or true and false at once, or neither. Such four-valued semantics first occurred in relevance logic (Anderson and Belnap 1975;

Belnap 1977). In Levesque's logic, $\mathbf{B}P \land \mathbf{B}(P \supset Q) \land \neg\mathbf{B}Q$ is satisfiable, where $\mathbf{B}$ is the modal operator for explicit belief, that is, belief is not closed under implication. This is contrary to any classical modal logic that satisfies the K axiom from modal logic. Moreover, $\mathbf{B}P \land \neg\mathbf{B}(P \land (Q \lor \neg Q))$ is satisfiable, so belief is not closed under equivalence.

Based on (Levesque 1984a) and tautological entailment, a number of further semantically grounded approaches emerged. Lakemeyer (1994, 1996) and Patel-Schneider (1990) extend the notion of implicit belief to the first-order case with weakened existential quantifiers. Delgrande (1995) adds belief contexts which can be reasoned about independently. Schaerf and Cadoli (1995) also build on Levesque's ideas for their theory approximation from below and above, which we describe below.

Semantic approaches based on subsumption, unit propagation, and case splits

Another more recent thread of research by Liu, Lakemeyer, and Levesque investigates limited belief in a first-order language (Lakemeyer and Levesque 2013, 2014, 2016; Liu 2006; Liu, Lakemeyer, and Levesque 2004). It is based on unit propagation, subsumption, and case splits as inference mechanism. The fundamental structure of the syntactically flavoured semantics are sets of ground clauses, so-called *setups*, which are closed under unit propagation and subsumption. The semantics is essentially defined inductively, with ground clauses being the base case. A setup satisfies a ground clause when it is contained in the setup's closure under unit propagation and subsumption.

Besides the setup, a natural number $k \in \{0, 1, 2, \ldots\}$ is part of the model to indicate the maximum allowed *effort* which may be put into proving that the setup satisfies the formula. Roughly, k specifies the number of allowed case splits. In the earlier approaches (Lakemeyer and Levesque 2013; Liu 2006; Liu, Lakemeyer, and Levesque 2004), a case split means to select some clause c from the setup and to verify that the setup augmented by every literal $\ell \in c$ satisfies the formula. In a more recent variant (Lakemeyer and Levesque 2014), a case split means to select an arbitrary literal ℓ and consider the setup first augmented by ℓ and then by $\overline{\ell}$. Lakemeyer and Levesque (2016) extend the latter notion to equality literals with functions.

For so-called proper$^+$ knowledge bases (Lakemeyer and Levesque 2002) reasoning about beliefs in these limited logics is sound (but incomplete) with respect to the traditional possible-worlds semantics. Liu (2006) and Liu, Lakemeyer, and Levesque (2004) also give a tractability result for the propositional case and fixed k. The major restriction of proper$^+$ knowledge bases is that existential quantifiers are not allowed. This restriction leads to a one-to-one correspondence between setups and the knowledge base and thus reduces logical entailment to model checking. Interestingly, the recent integration of functions (Lakemeyer and Levesque 2016) allows to represent existentials by means of Skolemization.

Klassen, McIlraith, and Levesque (2015) recently introduced a logic of limited belief in a fashion similar to (Lakemeyer and Levesque 2014). They propose a neighbourhood semantics, which elegantly avoids the syntactic flavour of setups. However, this approach is restricted to propositional logic.

Our approach to limited reasoning in Chapters 6 and 7 follows the idea of limited reasoning based on setups, unit propagation, subsumption, and case splits. As we will see, incomplete reasoning alone is not sufficient for conditional beliefs, and we will devise a complete semantics in the spirit of (Lakemeyer and Levesque 2014).

Other approaches

Not all approaches to the logical omniscience problem are semantically motivated. In an early proposal by Konolige (1986), the agent has a set of basic beliefs represented by logical sentences, and a (typically incomplete) set of deduction rules infers additional beliefs. The trouble with that is, as Levesque (1984a) argues, that the deduction rules need to ensure obvious inferences, such as that $(\alpha \vee \beta)$ implies $(\beta \vee \alpha)$. But even if that is given, the deduction lacks a semantical justification and it is not clear where to draw the line between obvious and non-obvious deductions.

Vardi (1986) represents belief as sets of sentences, each of which is modelled by its satisfying worlds. Fagin and Halpern (1987) introduce concepts of awareness and local reasoning. Awareness is supposed to capture that an agent can only have beliefs about something she is aware of. Local reasoning intends to model that people do not consider all issues at once, similar to (Delgrande 1995). To this end, not just one but multiple clusters of possible worlds are considered, and a sentence is believed when it is true in one of them. Both approaches have in common that they suspend closure under logical consequence. On the other hand, they are not satisfying as belief is still closed under logical equivalence.

Halpern, Moses, and Vardi (1994) and Kaplan and Schubert (2000) propose computational approaches, where the deduction of beliefs is captured by some terminating algorithm. These approaches are clearly very general – for example, Halpern, Moses, and Vardi (1994) can capture the proposals from Konolige (1986) and Levesque (1984a), at least under reasonable assumptions. On the downside, algorithmic approaches seem to be easy to use only as long as the applied algorithms are easy to reason about.

Unsound reasoning

By far the majority of existing approaches on decidable reasoning aim for sound but incomplete inferences. In some scenarios, however, complete but unsound reasoning is useful to *dis*prove statements. A particular use-case relevant for this thesis is to soundly determine whether a formula is consistent with some background knowledge.

In recent years, unsound reasoning has gained some interest in the theorem-proving community, as is illustrated by a series of "Workshops on Disproving: Non-Theorems, Non-Validity, Non-Provability" held at CADE and IJCAR conferences 2004–2007.

Among the more generally applicable approaches from these workshops is an unsound variant of resolution by Lynch (2004). Assuming a theorem prover that incrementally generates inferences (like resolution), the idea is to add new propositions which are possibly no logical consequences but ensure that the theorem prover halts.

Above we already mentioned the approach for theory approximation by Schaerf and Cadoli (1995). It supports both sound or complete reasoning by using tautological entailment similar to Levesque (1984a). In a propositional language where all formulas are in negation normal form, that is, negations only occur in front of atoms, their goal is to approximate, for given Φ, the theory $\{\phi \mid \Phi \models \phi\}$ by subsets and supersets. To this end, for a given set of propositional variables S, they define two consequence relations, $\models^3_S$ and $\models^1_S$, which are sound and complete, respectively, with respect to classical entailment. Semantically, this is achieved by two kinds of interpretations: while propositions from S are assigned either true or false as usual, interpretations for $\models^3_S$ may assign propositions not in S both true and false at once, and interpretations for $\models^1_S$ may not assign to these propositions any truth value at all. A (negated) atom is considered to be true under an interpretation when the interpretation maps it to true (false). Then $\models^3_S$ is sound, $\models^1_S$ is complete, and both are monotonic in S. Formally, for every Φ, ϕ, and $S \subseteq S'$,

$$\Phi \models^3_S \phi \text{ only if } \Phi \models^3_{S'} \phi \text{ only if } \Phi \models \phi \text{ only if } \Phi \models^1_{S'} \phi \text{ only if } \Phi \models^1_S \phi.$$

Schaerf and Cadoli (1995) also extended their approach to propositional modal logic and to fragments of first-order logic that correspond to simple description logics. Finger and Wassermann (2007) improve the approximation from above of $\models^1_S$ for propositional logic.

In Chapter 6, we introduce two semantics, which are sound, complete, and monotonic in a sense similar to $\models^1_S$ and $\models^3_S$. The complete semantics in a way pursues the same intuitive idea as Lynch (2004) does.

3 Logical Foundations

This chapter presents the logical foundations of this thesis. To begin with, we introduce the first-order logical language $\mathcal{L}$, which is the basis for every following logic in this thesis. While very similar to classical predicate logic, $\mathcal{L}$ has a distinguishing feature called standard names. Standard names not only simplify the semantics compared to classical logic, but will also prove fundamental for later results in this thesis, such as the representation theorems in Chapters 4 and 5 and the limited semantics in Chapters 6 and 7.

However, like classical logic, $\mathcal{L}$ is not well-suited to represent knowledge, especially the lack thereof. For this purpose we adopt the logic of only-knowing $\mathcal{OL}$, a modal extension of $\mathcal{L}$ due to Levesque (1984b). We present $\mathcal{OL}$ in this chapter for future reference. Later, in Chapter 4, we introduce an extension of $\mathcal{OL}$ to the more general case of conditional belief.

$\mathcal{OL}$ is a purely static languages. To prepare for our later work on beliefs in dynamic systems in Chapter 5, we present the epistemic situation calculus $\mathcal{ES}$ (Lakemeyer and Levesque 2011). It is an extension of $\mathcal{OL}$ that accommodates actions similar to Reiter's situation calculus but in the spirit of modal logic.

This chapter is based on (Levesque and Lakemeyer 2001) for $\mathcal{L}$ and $\mathcal{OL}$, and, with a little modification, on (Lakemeyer and Levesque 2011) for $\mathcal{ES}$.

3.1 Standard names

First-order logic differs from propositional logic in that it allows to represent objects. For example, a constant called gift may be used to represent a present, and InBox(gift) could be used to say that the gift is in the box. Yet, it is not possible in classical first-order logic to express *which* object the gift is. $\mathcal{L}$, which we introduce in the next sections, provides a non-classical feature called (standard) names to do just that. Standard names are special constants which allow us to syntactically refer to each and every object from the universe of discourse. Effectively, this means we fix the universe of discourse to be the countably infinite set of names $\#1, \#2, \#3, \ldots$

What is the benefit of standard names? After all, in Tarskian semantics (Tarski 1935, 1944) of first-order logic the domain is not fixed and may be any non-empty set, including uncountable ones (Kleene 2002). But standard names also have at least two significant advantages for our purposes.

Firstly, as we shall see, they allow us to handle quantification by substitution, which considerably simplifies the semantics. This is not possible in classical first-order logic because there the objects from the domain are no syntactic elements. It should be pointed out, though, that quantification by substitution has drawn philosophical criticism (Kripke 1976).

Secondly, standard names are useful to express belief about things. Namely, they allow us to distinguish between the following statements.

- We believe there is some x in the box.
- There is an x which we believe to be in the box.

The former expresses believing *that*, also known as de dicto belief in philosophy: perhaps we have no idea what x is. The latter, by contrast, represents believing *what* and is known as de re belief: here we know (or believe to know) what x is. Clearly, de re belief is stronger than de dicto belief.

Standard names can be thought of as special constants that satisfy the unique name assumptions and an infinitary version of domain closure.

3.2 The language $\mathcal{L}$

Definition 3.2.1 (Levesque and Lakemeyer 2001) The symbols of $\mathcal{L}$ are taken from the following vocabulary:

- infinitely many standard names $\#1, \#2, \ldots$, written schematically as n;
- infinitely many first-order variables, written schematically as x;
- infinitely many function symbols, written schematically as g;
- infinitely many predicate symbols, written schematically as P;
- connectives and other symbols: $=$, $\vee$, $\neg$, $\exists$, round brackets, comma.

Each function or predicate symbol has an *arity*, which indicates how many arguments it takes. Identifiers may be decorated with subscripts or superscripts.

Examples for predicate symbols are the unary InBox or Broken, and an example for a function symbol is the constant gift.

Definition 3.2.2 (Levesque and Lakemeyer 2001) The set of *terms* of $\mathcal{L}$ is the least set which includes all variables, standard names, and $g(t_1, \ldots, t_k)$ where g is a k-ary function symbol and $t_1, \ldots, t_k$ are terms. A term that contains no variables is called *ground*. A ground term that contains only a single function symbol is called *primitive*.

Definition 3.2.3 (Levesque and Lakemeyer 2001) The set of *formulas* of $\mathcal{L}$ is the least set such that

- $P(t_1, \ldots, t_k)$ is a formula where P is a predicate symbol and the t_i are terms;
- $(t_1 = t_2)$ is a formula where t_1 and t_2 are terms;
- $\neg\alpha$, $(\alpha \vee \beta)$, and $\exists x\, \alpha$ are formulas where α and β are formulas and x is a variable.

A formula of the form $P(t_1, \ldots, t_k)$ is called *atomic* or just an *atom*. A formula that contains no variables is called *ground*. A ground atom whose arguments $t_1, \ldots, t_k$ are standard names is called *primitive*. An occurrence of a variable x in α is *free* if that occurrence is not in a subformula of α of the form $\exists x\, \beta$. By α^x_t we denote the result of substituting t for all free occurrences of x in α. A formula that contains no free variable is called a *sentence*.

A formula $(\alpha \vee \beta)$ is called a disjunction, and $\exists$ is called an existential quantifier. For convenience, we define abbreviations to express inequality, tautology, contradiction, conjunctions, universals, material implications, and equivalence:

- $(t_1 \neq t_2)$ stands for $\neg(t_1 = t_2)$;
- TRUE stands for $\exists x (x = x)$;
- FALSE stands for $\neg$TRUE;
- $(\alpha \wedge \beta)$ stands for $\neg(\neg\alpha \vee \neg\beta)$;
- $\forall x\, \alpha$ stands for $\neg\exists x\, \neg\alpha$;
- $(\alpha \supset \beta)$ stands for $(\neg\alpha \vee \beta)$;
- $(\alpha \equiv \beta)$ stands for $((\alpha \supset \beta) \wedge (\beta \supset \alpha))$.

Occasionally we use $\vec{t}$ as an abbreviation for $t_1, \ldots, t_k$. We often omit brackets to ease readability. Then the convention is that the unary operators $\neg$, $\exists$, and $\forall$ bind strongest;

$\land$ binds stronger than $\lor$; $\lor$ binds stronger than $\supset$; $\supset$ binds stronger than $\equiv$. Moreover, unless said otherwise, we assume free variables to be universally quantified with maximal scope.

3.3 The semantics of $\mathcal{L}$

To define truth of formulas, some semantic primitive is needed. In classical first-order logic, truth of a formula is defined with respect to a structure, which consists of a domain of discourse and an interpretation function that maps predicate and function symbols to relations and functions over that domain, and an assignment function that maps variables to domain elements. With $\mathcal{L}$, things are much simpler thanks to standard names.

Definition 3.3.1 (Levesque and Lakemeyer 2001) A *world* w is a function

- from the primitive terms $g(n_1, \ldots, n_k)$ to standard names;
- from the primitive atoms $P(n_1, \ldots, n_k)$ to truth values $\{0, 1\}$.

Hence, a world can be used to determine the value of a term.

Definition 3.3.2 (Levesque and Lakemeyer 2001) The *denotation* $w(t)$ of a term t is defined as follows:

- $w(n) = n$ for every standard name n;
- $w(g(t_1, \ldots, t_k)) = w[g(n_1, \ldots, n_k)]$ where $n_i = w(t_i)$ and g is a function symbol.

Since quantification can be handled by substituting standard names for the variable, the truth of a formula in a world is easily defined.

Definition 3.3.3 (Levesque and Lakemeyer 2001) The truth relation $\models$ of $\mathcal{L}$ is defined with respect to a world w:

$\mathcal{L}$1. $w \models P(t_1, \ldots, t_k)$ iff $w[P(n_1, \ldots, n_k)] = 1$ where $n_i = w(t_i)$;

$\mathcal{L}$2. $w \models (t_1 = t_2)$ iff n_1 and n_2 are identical names where $n_i = w(t_i)$;

$\mathcal{L}$3. $w \models \neg\alpha$ iff $w \not\models \alpha$;

$\mathcal{L}$4. $w \models (\alpha \lor \beta)$ iff $w \models \alpha$ or $w \models \beta$;

$\mathcal{L}$5. $w \models \exists x\, \alpha$ iff $w \models \alpha^x_n$ for some name n.

As usual, we use the symbol $\models$ in the following also to denote logical entailment: we write $\Sigma \models \alpha$ iff all worlds w which satisfy all sentences in Σ also satisfy α, that is, $w \models \beta$ for all $\beta \in \Sigma$ implies that $w \models \alpha$. When Σ is empty, we just write $\models \alpha$. When Σ is the singleton containing β, we abbreviate $\beta \models \alpha$. We also sometimes identify a finite set of sentences with their conjunction.

3.4 Relationship to classical first-order logic

An important question is how big the difference of $\mathcal{L}$ to classical first-order logic is. The most notable difference is perhaps that $\mathcal{L}$ is not compact. Compactness is a corollary of of Gödel's completeness theorem (Gödel 1929; Kleene 2002) and means that a set of sentences is satisfiable iff every finite subset is satisfiable. In $\mathcal{L}$, there is a simple counterexample: $\{\exists x \neg P(x), P(\#1), P(\#2), P(\#3), \ldots\}$ is clearly unsatisfiable, but every finite subset is satisfiable.

However, for formulas without standard names, $\mathcal{L}$ comes very close to classical logic. The following theorem illustrates this.

Theorem 3.4.1 (Levesque and Lakemeyer 2001) *Let α contain no standard names and no =. Then $\models \alpha$ iff α is a valid sentence in classical first-order logic.*

Why do classical logic and $\mathcal{L}$ differ on =? The reason again are standard names: they assert an infinite but countable domain. While classical first-order logic cannot distinguish between countable and uncountable infinite domains by the Löwenheim-Skolem theorem (Kleene 2002), a formula may very well require every model to have a finite domain. For example, $\exists x_1 \ldots \exists x_k \forall x (x = x_1 \lor \ldots \lor x = x_k)$ ensures the domain has no more than k elements in classical predicate logic with equality; it is easy to see that this formula is unsatisfiable in $\mathcal{L}$. Let δ_k be the negation of that sentence and $\Delta = \{\delta_k \mid k \geq 1\}$ be a theory that asserts an infinite domain. Then any sentence α that mentions no standard names is valid in $\mathcal{L}$ iff Δ entails α in classical predicate logic with equality (Levesque and Lakemeyer 2001). (Note that the interpretation of = is fixed in predicate logic with equality to be the identity relation (Kleene 2002).)

As a corollary of Theorem 3.4.1 and the famous results by Church (1936a,b) and Turing (1936) it follows that the decision problem for $\mathcal{L}$ is undecidable.

Corollary 3.4.2 *Satisfiability in $\mathcal{L}$ is undecidable.*

This of course limits the practical utility of $\mathcal{L}$ and any language that subsumes it, which includes all logics presented in this and the following two chapters. To remedy this, we investigate limited first-order reasoning in Chapters 6 and 7, which is decidable for a large class of problems.

3.5 Modelling knowledge

Ordinary first-order logic is a fairly standard knowledge-representation language, so the question may arise: why not use it also to represent knowledge and belief? Roughly, the problem with $\mathcal{L}$ is that it does not allow fine-grained enough control over what is known and, more importantly, what is *not* known. For example, we might know there is something in a box, but we do not know what it is. The formula $\exists x\, \mathrm{InBox}(x)$ could only capture the first part. So the question arises whether there is a way in $\mathcal{L}$ to capture the second part as well.

One way to express that the identity of x is unknown would be to use another predicate $\mathrm{Known}_{\mathrm{InBox}}$ and to say $\exists x\, (\mathrm{InBox}(x) \land \neg\mathrm{Known}_{\mathrm{InBox}}(x))$. Unfortunately, though, this approach does not scale well. Suppose we know that either $\#1$ or $\#2$ is in the box. The formula $\mathrm{Known}_{\mathrm{InBox}}(\#1) \lor \mathrm{Known}_{\mathrm{InBox}}(\#2)$ does not represent this appropriately: the formula is true if $\mathrm{Known}_{\mathrm{InBox}}(\#1)$ or $\mathrm{Known}_{\mathrm{InBox}}(\#2)$ is true, but neither should be, since neither $\#1$ or $\#2$ is known to be in the box! Hence another predicate $\mathrm{Known}_{\mathrm{InBox}(\#1)\lor\mathrm{InBox}(\#2)}$ would be necessary. In general, we would need such a predicate for every formula that expresses incomplete knowledge.

Another solution that comes to mind is a three-valued semantics, where a third value represents "unknown" besides the binary truth values. But the composition of two unknowns is unclear. In the above example, we would have $w[\mathrm{InBox}(n)] = \text{unknown}$ for all names n. But on the other hand, $w \models \exists x\, \mathrm{InBox}(x)$, so the unknowns somehow should combine to knowledge. It is unclear how such a semantics could look like.

The standard tool to overcome such issues is to consider multiple *possible* worlds. The concept is due to Kripke (1959) and Hintikka (1962). The intuition is that since the agent has only incomplete knowledge, she considers many different worlds possible. Knowledge is what is true in all these worlds. For example, if we had two possible worlds, one of which satisfies $\mathrm{InBox}(\#1)$ and the other $\mathrm{InBox}(\#2)$, we would know $\mathrm{InBox}(\#1) \lor \mathrm{InBox}(\#2)$, but neither $\mathrm{InBox}(\#1)$ nor $\mathrm{InBox}(\#2)$. Syntactically, this is expressed with a modal operator: $\mathbf{K}(\mathrm{InBox}(\#1) \lor \mathrm{InBox}(\#2)) \land \neg\mathbf{K}\mathrm{InBox}(\#1) \land \neg\mathbf{K}\mathrm{InBox}(\#2)$.

$\mathcal{OL}$ uses possible worlds as the semantic model of knowledge, and we will stick with this for the most part of this thesis. An alternative way would be to reify possible worlds, thus avoiding a new semantics. For one thing, however, it seems counterintuitive to represent a clearly semantic concept like worlds syntactically. For another, the additional indirection would make reasoning in this theory unnecessarily complex.

An important case of knowledge frequently encountered in knowledge representation are knowledge bases. A knowledge base is special in that it intuitively captures the agent's knowledge *to its full extent*. Here, the lack of knowledge we want to express is even infinite: everything is unknown unless it follows from the knowledge base. Even if we reified possible worlds as objects in ordinary first-order logic, this would require a meta-logical knowledge closure to express all these unknowns. In $\mathcal{OL}$, by contrast, it is captured with a modal operator for only-knowing: $\mathbf{O}(\mathrm{InBox}(\#1) \lor \mathrm{InBox}(\#2))$ stipulates that all the agent knows is $\mathrm{InBox}(\#1) \lor \mathrm{InBox}(\#2)$, everything else is *not* known (including but not limited to $\mathrm{InBox}(\#1)$ and $\mathrm{InBox}(\#2)$).

3.6 The language $\mathcal{OL}$

Definition 3.6.1 (Levesque and Lakemeyer 2001) The symbols of $\mathcal{OL}$ are the same as for $\mathcal{L}$ (Definition 3.2.1) plus **K** and **O**. The *terms* are the same as in $\mathcal{L}$ (Definition 3.2.2). The *formulas* are formed by the same rules as $\mathcal{L}$ (Definition 3.2.3) plus

- $\mathbf{K}\alpha$ and $\mathbf{O}\alpha$ are a formulas if α is a formula.

A formula that mentions no **K** or **O** is called *objective*. A formula that mentions function and predicate symbols only within **K** or **O** is called *subjective*.

$\mathbf{K}\alpha$ may be read as "α is known." Using a modal operator for knowledge allows us to distinguish between de dicto and de re knowledge: knowing that there is some x in the box is represented as $\mathbf{K}\exists x\,\mathrm{InBox}(x)$, whereas knowing an object in the box can be written as $\exists x\,\mathbf{K}\mathrm{InBox}(x)$. When a variable is quantified outside of the belief modality as in de re knowledge, this is also referred to as *quantifying-in*. It also allows for introspection: $\mathbf{KK}\alpha$ means that we know that we know α.

$\mathbf{O}\alpha$ means "α is *all* that is known." This concept is useful in knowledge representation because when specifying a knowledge base, one typically assumes that it exhaustively represents the agent's knowledge. Only-knowing also is related to autoepistemic logic (Levesque 1990; Moore 1985) and Reiter's default logic (Lakemeyer and Levesque 2005; Reiter 1980).

As a notational convention, we will use υ, ϕ, ψ for objective sentences.

3.7 The semantics of $\mathcal{OL}$

The semantics of $\mathcal{OL}$ inherits Definitions 3.3.1 and 3.3.2 for worlds and the denotation of terms, respectively. We can hence immediately proceed to give the semantics of the language.

Definition 3.7.1 (Levesque and Lakemeyer 2001) The truth relation $\models$ of $\mathcal{OL}$ is defined with respect to a set of worlds e and a world w:

$\mathcal{OL}$1. $e, w \models P(t_1, \ldots, t_k)$ iff $w[P(n_1, \ldots, n_k)] = 1$ where $n_i = w(t_i)$;

$\mathcal{OL}$2. $e, w \models (t_1 = t_2)$ iff n_1 and n_2 are identical names where $n_i = w(t_i)$;

$\mathcal{OL}$3. $e, w \models \neg\alpha$ iff $e, w \not\models \alpha$;

$\mathcal{OL}$4. $e, w \models (\alpha \vee \beta)$ iff $e, w \models \alpha$ or $e, w \models \beta$;

$\mathcal{OL}$5. $e, w \models \exists x\, \alpha$ iff $e, w \models \alpha^x_n$ for some name n;

$\mathcal{OL}$6. $e, w \models \mathbf{K}\alpha$ iff for all w', if $w' \in e$, then $e, w' \models \alpha$;

$\mathcal{OL}$7. $e, w \models \mathbf{O}\alpha$ iff for all w', $w' \in e$ iff $e, w' \models \alpha$.

Except for the addition of Rules $\mathcal{OL}$6 and $\mathcal{OL}$7 and the new parameter e, the semantics matches the one of $\mathcal{L}$ from Definition 3.3.3. Note that the rule for only-knowing is very similar to the one for ordinary knowledge, but also requires the converse direction. Thus, $\mathbf{O}\alpha$ implies $\mathbf{K}\alpha$ and additionally maximizes the number of possible worlds in e.

In the following, we allow ourselves to omit e or w sometimes. Then $e \models \alpha$ is to say $e, w \models \alpha$ for all w. Analogously, $w \models \alpha$ stands for $e, w \models \alpha$ for all e. We often use these abbreviations for objective and subjective α, respectively.

It is straightforward to show that knowledge is closed under modus ponens and both positively and negatively introspective.

Theorem 3.7.2 (Levesque and Lakemeyer 2001)

(i) $\models \mathbf{K}\alpha \wedge \mathbf{K}(\alpha \supset \beta) \supset \mathbf{K}\beta$;

(ii) $\models \mathbf{K}\alpha \supset \mathbf{K}\mathbf{K}\alpha$;

(iii) $\models \neg\mathbf{K}\alpha \supset \mathbf{K}\neg\mathbf{K}\alpha$.

In the terminology of modal logic, this makes $\mathbf{K}\alpha$ a K45 operator (Fagin, Halpern, et al. 1995).

It is immediate that only-knowing is stronger than knowing.

Theorem 3.7.3 $\models \mathbf{O}\alpha \supset \mathbf{K}\alpha$.

The next property says that we can conjoin everything that is known.

Theorem 3.7.4 (Levesque and Lakemeyer 2001) $\models \mathbf{O}\alpha \wedge \mathbf{K}\beta \supset \mathbf{O}(\alpha \wedge \beta)$.

Of particular interest are entailments of the form $\mathbf{O}\alpha \models \mathbf{K}\beta$, which corresponds to querying the knowledge base α whether β is known. The following theorem is of great importance for such reasoning tasks.

Theorem 3.7.5 (Levesque and Lakemeyer 2001)
Let ϕ be objective. Then there is a unique e such that $e \models \mathbf{O}\phi$.

Is is also easy to see that Theorem 3.7.3 can be strengthened to the following result.

Theorem 3.7.6 *Let ϕ be objective. Then $e \models \mathbf{O}\phi$ iff e is maximal such that $e \models \mathbf{K}\phi$.*

Only-knowing is thus useful to uniquely determine an agent's knowledge. Checking whether $\mathbf{O}\phi$ entails $\mathbf{K}\alpha$ thus boils down to model-checking $e \models \mathbf{K}\alpha$, where $e \models \mathbf{O}\phi$ is unique.

As an example, let α be $\exists x\, \mathrm{InBox}(x)$. Then $e \models \mathbf{O}\alpha$ iff $e = \{w \mid w[\mathrm{InBox}(n)] = 1$ for some name $n\}$. We therefore have $\mathbf{O}\alpha \models \mathbf{K}\exists x\, \mathrm{InBox}(x)$ since every $w \in e$ satisfies $\mathrm{InBox}(n)$ for some name n. However, $\mathbf{O}\alpha \models \neg\exists x\, \mathbf{K}\mathrm{InBox}(x)$, since for all names n there is some $w \in e$ such that $w \models \neg\mathrm{InBox}(n)$.

This example also illustrates the difference between knowing and only-knowing. Clearly, $\mathbf{K}\alpha \models \mathbf{K}\exists x\, \mathrm{InBox}(x)$ as well. This is because for every $e' \models \mathbf{K}\alpha$, $e' \subseteq e$. However, since e' is not necessarily maximal, $\mathbf{K}\alpha \not\models \neg\exists x\, \mathbf{K}\mathrm{InBox}(x)$. Only-knowing does maximize e and thus appropriately represents the agent's unknowns.

The following is an immediate consequence of the unique-model property.

Corollary 3.7.7 (Levesque and Lakemeyer 2001)
Let ϕ be objective and σ be subjective. Then $\mathbf{O}\phi \models \sigma$ or $\mathbf{O}\phi \models \neg\sigma$.

3.8 Modelling actions

Ordinary first-order logic is static, just like $\mathcal{OL}$. A popular way to bring dynamics to logic is to endow functions and predicates with an additional argument that reifies the current world state. In the situation calculus (McCarthy 1963; Reiter 1991, 2001), this parameter is called *situation*, and predicates or functions whose value changes depending on the situation are called *fluent*.

While there are other interpretations (McCarthy and Hayes 1969), Reiter takes a situation as sequential history of executed actions $a_1, \ldots, a_k$, represented as a term of the form $\mathrm{do}(a_k, \mathrm{do}(a_{k-1}, \ldots \mathrm{do}(a_1, s) \ldots))$. Knowledge and sensing can be modelled in this theory following the idea of possible worlds: a special fluent predicate serves as accessibility relation of possible situations (Scherl and Levesque 2003); and when the agent senses that $\phi(s)$ holds in the actual situation s, only those situations s' which also satisfy $\phi(s')$ are stay accessible.

Reiter's situation calculus is a dialect of classical first-order logic, which however needs second-order logic to axiomatize the space of situations appropriately. Especially when it comes to epistemic reasoning, reification and Tarskian semantics make reasoning quite tedious at times.

Lakemeyer and Levesque (2004, 2011) propose an amalgamation of the situation calculus with $\mathcal{OL}$, called $\mathcal{ES}$ for epistemic situation calculus. What makes $\mathcal{ES}$ special is that it retains the spirit of $\mathcal{OL}$: situations are not reified but a purely semantic concept, and actions occur as modal operators $[a]$ and $\Box$ (unlike in an earlier proposal (Lakemeyer and Levesque 1998)). For example, in the original situation calculus saying that any fragile x in the box is broken after dropping the box needs a formula $\forall s \forall x (\text{InBox}(x, s) \land \text{Fragile}(x, s) \supset \text{Broken}(x, \text{do}(\text{dropbox}, s))$. In $\mathcal{ES}$, the same statement is expressed as $\Box \forall x (\text{InBox}(x) \land \text{Fragile}(x) \supset [\text{dropbox}]\text{Broken}(x))$. Actions go very well along with epistemic features in this language.

The situation calculus itself does not stipulate which (physical or epistemic) effects an action has. All it says is that executing a specific action in a certain situation always leads to the same successor situation; actions are deterministic in this sense. An action's effects may very well be nondeterministic, though. However, in the most prevalent kind of theory used in Reiter's situation calculus and its descendants, the so-called basic action theories, action effects are indeed deterministic. Uncertainty about the domain hence needs to be encoded in the agent's initial knowledge and beliefs.

3.9 The language $\mathcal{ES}$

The language $\mathcal{ES}$, in contrast to $\mathcal{OL}$, is multi-sorted: standard names come in sorts *object* and *actions*. There are also two different types of predicate symbols: fluent ones may change their truth value as the result of actions, rigid ones do not. We introduce the symbols and syntax of $\mathcal{ES}$ from scratch.

Definition 3.9.1 The symbols of $\mathcal{ES}$ are taken from the following vocabulary:

- infinitely many object standard names $\#1, \#2, \ldots$;
- infinitely many first-order variables, written schematically as x,
  - of sort object, written schematically as y;
  - of sort action, written schematically as a;
- infinitely many function symbols

  - of sort object, written schematically as g;
  - of sort action, written schematically as A;
- infinitely many predicate symbols
  - of type fluent, written schematically as F;
  - of type rigid, written schematically as R;
- connectives and other symbols: $=$, $\vee$, $\neg$, $\exists$, $\Box$, **K**, **O**, round and square brackets, comma.

Each function or predicate symbol has an *arity* which indicates how many arguments it takes. Identifiers may be decorated with subscripts or superscripts. There shall be two special unary fluent predicates called Poss and SF.

Intuitively, Poss(t) represents the precondition of the action t, and SF(t) represents the sensing result of action t. For example, if t is the action of checking if the box is empty, SF(t) would be true iff the box is empty. Of course, preconditions and sensing results just like action effects are not fixed in the logic but to be defined axiomatically.

Definition 3.9.2 Standard names, written schematically as n, come in two sorts:

- the object standard names are $^{\#}1, ^{\#}2, \ldots$;
- the action standard names are of the form $A(n_1, \ldots, n_k)$ where A is an action function symbol and the n_i are object standard names.

The set of *terms* of sort object or action is the least set such that

- every variable and standard name is a term of the corresponding sort;
- $g(t_1, \ldots, t_k)$ is an object term if g is an object function symbol and the t_i are terms;
- $A(t_1, \ldots, t_k)$ is an action term if A is an action function symbol and the t_i are object terms.

A term that contains no variables is called *ground*. A ground term that contains only a single function symbol is called *primitive*.

The differences of the version of $\mathcal{ES}$ we present here and the one from (Lakemeyer and Levesque 2011) are already visible.

- We renamed the modalities *Know* and *OKnow* to **K** and **O**, respectively, in order to be consistent with $\mathcal{OL}$.

- We omit fluent functions here. This is merely to ease the presentation; functions can be simulated in first-order logic with equality anyway, so our restriction means no loss in expressivity.
- Our variant considers no second-order logic. Again, this eases the presentation. We will return to this issue later in Section 5.6.
- Finally, in (Lakemeyer and Levesque 2011), action standard names are defined analogously to object standard names. In contrast, here they are primitive action terms. While it may seem awkward at first sight, this helps us avoid a counterintuitive behaviour of the semantics in (Lakemeyer and Levesque 2011).

Definition 3.9.3 The set of *formulas* of $\mathcal{ES}$ is the least set such that

- $P(t_1, \dots, t_k)$ is a formula where P is a predicate symbol and the t_i are terms;
- $(t_1 = t_2)$ is a formula where t_1 and t_2 are terms of the same sort;
- $\neg\alpha$, $(\alpha \lor \beta)$, $\exists x\, \alpha$ are formulas where α and β are formulas and x is a variable;
- $[t]\alpha$ and $\Box\alpha$ are formulas where α is a formula and t is an action term;
- $\mathbf{K}\alpha$ and $\mathbf{O}\alpha$ are formulas where α is a formula.

A formula of the form $P(t_1, \dots, t_k)$ is called *atomic* or just an *atom*. A formula that contains no variables is called *ground*. A ground atom whose arguments $t_1, \dots, t_k$ are standard names is called *primitive*. A formula that contains no free variable is called a *sentence*.

The essential addition over $\mathcal{OL}$ are the expressions $[t]\alpha$ and $\Box\alpha$, both of which concern actions. The former means "after doing t, α is true," the latter means "after any sequence of actions, α holds true."

3.10 The semantics of $\mathcal{ES}$

To account for actions, we need to extend the concept of worlds from $\mathcal{L}$ and $\mathcal{OL}$ (Definition 3.3.1). A world shall not only determine a momentary snapshot but also determine future states.

Definition 3.10.1 An *action sequence* is the empty sequence $\langle\rangle$ or the concatenation $z \cdot n$ of an action sequence z and an action standard name n. A *world* w is a function

- from the primitive object terms $g(n_1, \dots, n_k)$ to object standard names;

- from the primitive rigid atoms $R(n_1, \ldots, n_k)$ to truth values $\{0, 1\}$;
- from the primitive fluent atoms $F(n_1, \ldots, n_k)$ and action sequences to truth values $\{0, 1\}$.

Following Reiter's terminology and his concept of situation terms (Reiter 2001), we occasionally refer to an action sequence as *situation*, and to the empty sequence as the initial situation.

Definition 3.10.2 The *denotation* $w(t)$ of a term t is defined as follows:

- $w(n) = n$ for every standard name n;
- $w(g(t_1, \ldots, t_k)) = w[g(n_1, \ldots, n_k)]$ where $n_i = w(t_i)$ and g is an object function symbol;
- $w(A(t_1, \ldots, t_k)) = A(n_1, \ldots, n_k)$ where $n_i = w(t_i)$ and A is an action function symbol;

For example, if gift is an object constant and unbox is a unary action symbol, we could have $w(\text{gift}) = {}^{\#}5$, and then $w(\text{unbox}(\text{gift})) = \text{unbox}({}^{\#}5)$.

As mentioned above, actions not only have effects but can also produce new information through sensing. The sensing result of an action n is represented by the atom $\text{SF}(n)$. Semantically, the sensing effect of n is reflected by dropping all possible worlds that disagree with the real world's value of $\text{SF}(n)$.

Definition 3.10.3 We write $w' \simeq_z w$ to say that w' agrees with w on the sensing throughout action sequence z, which is defined inductively by

- $w' \simeq_{\langle\rangle} w$ for all w' and w;
- $w' \simeq_{z \cdot n} w$ iff $w' \simeq_z w$, $w'[\text{Poss}(n), z] = 1$, and $w'[\text{SF}(n), z] = w[\text{SF}(n), z]$.

Intuitively, in $w' \simeq_z w$ the first world w' is a possible world, and w is the actual world where the sensing happened. Besides requiring that both worlds agree on $\text{SF}(n)$, the relation also requires n to be a legal action in w' (in some variants of $\mathcal{ES}$ the latter requirement is omitted, for example, in (Lakemeyer and Levesque 2004)).

Lakemeyer and Levesque (2011) thirdly require in $w' \simeq_z w$ both worlds to agree on rigid terms and atoms. In their formalization, action terms are interpreted analogously to object terms and mapped to action standard names, which are atomic entities just like object standard names. With the additional constraint, they ensure that the standard names in z refer to the same actions in both the actual and the possible worlds. (Claßen

(2013) hence weakens the requirement to only hold for action terms.) This requirement brings trouble when we want to extend Levesque's representation theorem to $\mathcal{ES}$, as we will do in Section 5.9. We briefly elaborate on the problem in Section 5.12. The reader familiar with the representation theorem may already verify that the formula $((n = A) \supset \mathbf{K}(n = A))$ for an action standard name n and an action constant A is valid in $\mathcal{ES}$, but the classical representation theorem would reduce the sentence to $((n = A) \supset \text{FALSE})$.

Definition 3.10.4 The truth relation $\models$ of $\mathcal{ES}$ is defined with respect to a set of worlds e, a world w, and an action sequence z:

$\mathcal{ES}1$. $e, w, z \models R(t_1, \ldots, t_k)$ iff
R is rigid and $w[R(n_1, \ldots, n_k)] = 1$ where $n_i = w(t_i)$;

$\mathcal{ES}2$. $e, w, z \models F(t_1, \ldots, t_k)$ iff
F is fluent and $w[F(n_1, \ldots, n_k), z] = 1$ where $n_i = w(t_i)$;

$\mathcal{ES}3$. $e, w, z \models (t_1 = t_2)$ iff n_1 and n_2 are identical names where $n_i = w(t_i)$;

$\mathcal{ES}4$. $e, w, z \models \neg\alpha$ iff $e, w, z \not\models \alpha$;

$\mathcal{ES}5$. $e, w, z \models (\alpha \vee \beta)$ iff $e, w, z \models \alpha$ or $e, w, z \models \beta$;

$\mathcal{ES}6$. $e, w, z \models \exists x\, \alpha$ iff $e, w, z \models \alpha^x_n$ for some name n of the same sort as x;

$\mathcal{ES}7$. $e, w, z \models [t]\alpha$ iff $e, w, z \cdot n \models \alpha$ where $n = w(t)$;

$\mathcal{ES}8$. $e, w, z \models \Box\alpha$ iff $e, w, z \cdot z' \models \alpha$ for every action sequence z';

$\mathcal{ES}9$. $e, w, z \models \mathbf{K}\alpha$ iff for all $w' \simeq_z w$, if $w' \in e$, then $e, w', z \models \alpha$;

$\mathcal{ES}10$. $e, w, z \models \mathbf{O}\alpha$ iff for all $w' \simeq_z w$, $w' \in e$ iff $e, w', z \models \alpha$.

Lakemeyer and Levesque (2004, 2009, 2011) have proposed several alternative definitions for only-knowing. The trouble with the above Rule $\mathcal{ES}10$, which is taken from (Lakemeyer and Levesque 2004), is that the unique-model property of only-knowing from Theorem 3.7.5 does not hold true after actions. For example, consider $\mathrm{SF}(n) \wedge [n]\mathbf{O}\text{TRUE}$. Then any model e contains all worlds that satisfy $\mathrm{SF}(n)$. But Rule $\mathcal{ES}10$ also allows e to contain an arbitrary number of additional worlds that falsify $\mathrm{SF}(n)$.

Lakemeyer and Levesque (2009) fix this issue by progressing all worlds by z. This way, they retain the unique-model property, so that they can capture all that is known after actions. More precisely, their definition is

$\mathcal{ES}$9'. $e, w, z \models \mathbf{K}\alpha$ iff for all w', if $w' \in e^w_z$, then $e^w_z, w', \langle\rangle \models \alpha$;

$\mathcal{ES}$10'. $e, w, z \models \mathbf{O}\alpha$ iff for all w', $w' \in e^w_z$ iff $e^w_z, w', \langle\rangle \models \alpha$;

where the progression of a set of worlds is $e^w_z = \{w'_z \mid w' \in e \text{ and } w' \simeq_z w\}$ and for a single world $w'_z[F(\vec{n}), z'] = w'[F(\vec{n}), z \cdot z']$ and $w'_z[R(\vec{n})] = w'[R(\vec{n})]$.

Alternatively, one could progress e and w already in Rules $\mathcal{ES}$7 and $\mathcal{ES}$8. Then the z parameter is no longer needed for $\models$. We devise such a semantics for $\mathcal{ESB}$, our variant of $\mathcal{ES}$, in Chapter 5.

3.11 Discussion

This chapter introduced the logical foundations of this thesis: the non-modal language $\mathcal{L}$, the logic of only-knowing $\mathcal{OL}$, and finally the epistemic situation calculus $\mathcal{ES}$, which are subsumed by each other. $\mathcal{L}$ extends classical first-order logic with a feature called standard names which allow to represent the identity of objects. $\mathcal{OL}$ extends this language with modalities for knowing and only-knowing. $\mathcal{ES}$ further adds actions and sensing in the spirit of Reiter's situation calculus.

While knowledge needs not necessarily be correct, there are still important limitations which we address in the upcoming chapters.

For one thing, knowledge is unconditional. For example, we might believe that someone is Italian, but if not, she presumably is Australian. The only sort of conditional statement supported by the logics from this chapter is material implication. The sentence "if she is not Italian, then α" can only be translated to $\neg\text{Italian} \supset \alpha$. Since we know she is Italian, this material implication is vacuously true. To remedy this, the next chapter introduces $\mathcal{BO}$, an extension of $\mathcal{OL}$ to conditional belief that allows to express such beliefs. We prefer the term "belief" over "knowledge" for such conditionals, as it suggests that the agent the possibility into account that is beliefs are wrong. We continue to use "knowledge" in the spirit of $\mathcal{OL}$ where the corresponding information is assumed to be correct.

For another, knowledge is indefeasible. For example, when the agent knows ϕ and then senses $\neg\phi$ in $\mathcal{ES}$, the agent is in a state of logical inconsistency, in which she knows *everything*. Arguably, this is not useful in practical considerations. In Chapter 5 we introduce $\mathcal{ESB}$, which amalgamates $\mathcal{BO}$ with actions like $\mathcal{ES}$ does with $\mathcal{OL}$, but deals more reasonably with contradictory information by giving up beliefs in an appropriate way.

4 Conditional Belief and Only-Believing

This chapter introduces a logic of conditional belief called $\mathcal{BO}$. Conditional beliefs are ubiquitous in our daily lives when we reason about different contingencies. Often, they are of the form "if *some premise* holds, then presumably *some consequent* is true."

What makes such conditionals special is that the premise may be (or believed to be) counterfactual, and still an agent could reasonably consider what would follow if the premise was true. In $\mathcal{OL}$, such a statement cannot be represented appropriately: the only form of conditional in $\mathcal{OL}$ is the material implication, which is vacuously true for a false premise.

$\mathcal{BO}$ is a descendant of $\mathcal{OL}$ and inherits many of its ideas, but extends it with the notion of conditional belief. In particular, it generalizes the concept of only-knowing to conditional belief; we refer to this conditional variant of only-knowing as *only-believing*. By an embedding theorem we will see that indeed $\mathcal{BO}$ soundly extends $\mathcal{OL}$. We also investigate the close relationship of only-believing to Pearl's System Z (Pearl 1990).

The presentation of $\mathcal{BO}$ is based on (Schwering and Lakemeyer 2014, 2015). Some of the longer proofs can be found in Appendix A.

4.1 Conditional belief versus knowledge

To see why conditional belief is a more general concept than knowledge, consider the following example. (We prefer this example over Example 1.1.1 to illustrate the results in this chapter and later in Chapters 6 and 7, as it involves no actions.)

Example 4.1.1 Suppose we expect a guest for dinner. We don't know much about her preferred diet yet, but we do have some (somewhat narrow-minded) beliefs:

- Most Australians are not Italians, and vice-versa.
- Australians usually eat kangaroo meat.
- We believe our guest is Italian or a vegetarian, and
  if she is not Italian, she presumably is Australian.

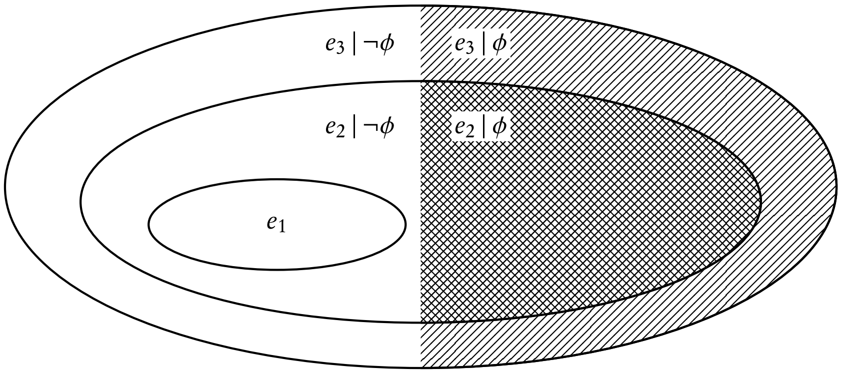


Figure 4.1: A system of spheres. The hatched area indicates ϕ-worlds; the most-plausible ones occur in e_2, and e_3 contains additional ones; there is none among the most-plausible worlds e_1. The conditional "if ϕ is true, then presumably ψ is also true" holds when the worlds in the double-hatched area satisfy ψ.

- We know that kangaroo is meat, and that vegetarians do not eat meat.

Given this conditional knowledge base, do we expect our guest to be a vegetarian in case she is not Italian?

Monotonic reasoning would suggest so: our belief of her being Italian or a vegetarian yields that she must be a vegetarian if not Italian. But she also must be Australian, hence eat kangaroo, and thus be a non-vegetarian – that is, in a monotonic logic *everything* is believed if she is not Italian, because the beliefs are inconsistent with that contingency.

Conditional beliefs do not trap into inconsistency that easily. They detect that the premise "not Italian" is inconsistent with the most-plausible scenarios, and therefore go on to look for *less-plausible* scenarios where the premise holds, and check the consequent in these scenarios. In our example, we hence believe that if the guest is not Italian, she presumably is an Australian kangaroo-eater, but not a vegetarian.

Perhaps the most popular way to represent conditional beliefs is by a system of spheres due to Lewis (1973) and Grove (1988). Every sphere is a set of possible worlds, like the e in the semantics of $\mathcal{OL}$. The idea is that one starts out with a narrow set of possible worlds, which is contained by larger spheres. Such a system is depicted in Figure 4.1. A conditional "if α, then presumably β" holds when the material implication $(\alpha \supset \beta)$ holds at the innermost sphere consistent with α. The logic we present next follows this model.

4.2 The language $\mathcal{BO}$

Definition 4.2.1 The symbols of $\mathcal{BO}$ are the same as for $\mathcal{L}$ (Definition 3.2.1) plus curly brackets, $\Rightarrow$, **B**, and **O**. The *terms* are the same as in $\mathcal{L}$ (Definition 3.2.2). The *formulas* are formed by the same rules as $\mathcal{L}$ (Definition 3.2.3) plus

- $\mathbf{B}(\alpha_1 \Rightarrow \beta_1)$ and $\mathbf{O}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ are formulas if α_i, β_i are formulas.

A formula that mentions no **B** or **O** is called *objective*. A set $\{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ is called objective when all ϕ_i and ψ_i are objective. A formula that mentions function and predicate symbols only within **B** or **O** is called *subjective*.

We read the conditional belief $\mathbf{B}(\alpha \Rightarrow \beta)$ as "if α is true, then presumably β is also true," or alternatively as counterfactual "if α was true, then β would also be true." We call $\mathbf{O}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ the *only-believing* operator. It generalizes Levesque's only-knowing from $\mathcal{OL}$ and is read as "the conditionals $\alpha_i \Rightarrow \beta_i$ are *all* that is believed."

We define the following abbreviations:

- $\mathbf{B}\alpha$ stands for $\mathbf{B}(\textsc{true} \Rightarrow \alpha)$;
- $\mathbf{K}\alpha$ stands for $\mathbf{B}(\neg\alpha \Rightarrow \textsc{false})$.

$\mathbf{B}\alpha$ and $\mathbf{K}\alpha$ are read as "α is believed" and "α is known," respectively. We also adopt the other logical abbreviations from $\mathcal{L}$. Like with other unary operators, **B** and **K** shall bind stronger than the logical connectives.

Before we proceed with the semantics, let us see how Example 4.1.1 can be expressed in $\mathcal{BO}$.

Example 4.2.2 Let the predicates Aussie, Italian, Veggie represent that the guest is Australian, Italian, a vegetarian, respectively; Eats(x) that x is among her preferred diet; Meat(x) that x is meat. Let roo be a standard name representing kangaroo meat. Then all we believe is

- Aussie $\Rightarrow$ ¬Italian and Italian $\Rightarrow$ ¬Aussie;
- Aussie $\Rightarrow$ Eats(roo);
- $\textsc{true} \Rightarrow$ Italian $\vee$ Veggie and ¬Italian $\Rightarrow$ Aussie;
- $\neg\text{Meat(roo)} \Rightarrow \textsc{false}$ and $\neg\forall x(\text{Veggie} \wedge \text{Meat}(x) \supset \neg\text{Eats}(x)) \Rightarrow \textsc{false}$.

The question "if she is not Italian, is she presumably not a vegetarian?" then boils down to whether $\mathbf{O}\Gamma$ entails $\mathbf{B}(\neg\text{Italian} \Rightarrow \neg\text{Veggie})$, where Γ is the set that contains the above conditionals.

4.3 The semantics of $\mathcal{BO}$

The semantics of $\mathcal{BO}$ inherits from $\mathcal{L}$ Definitions 3.3.1 and 3.3.2 of worlds and of the denotation of terms, respectively. Before we proceed with the semantics of the language, we need to formalize the notion of an epistemic state. It follows the concept of system of spheres mentioned above and depicted in Figure 4.1. To keep matters simple, we assume only finitely many different spheres, which are consecutively numbered.

Definition 4.3.1 An *epistemic state* $\vec{e}$ is an infinite sequence of sets of worlds e_p, $p \in \mathbb{P} = \{1, 2, \ldots\}$, that

- is concentric, that is, $e_p \subseteq e_{p+1}$ for all $p \in \mathbb{P}$;
- converges, that is, $e_q = e_p$ for some $q \in \mathbb{P}$ and all $p \geq q$.

We use $\langle e_1, \ldots, e_q \rangle$ as a short notation for $\vec{e}$ when it converges at level q or earlier.

Note that every $\vec{e}$ can be expressed as $\langle e_1, \ldots, e_q \rangle$ for some $q \in \mathbb{P}$; q does not need to be minimal, though; for example, $\langle e_1, e_2 \rangle = \langle e_1, e_2, e_2 \rangle$.

An epistemic state induces a ranking of worlds and sentences by their *plausibility*. The plausibility of a world is the plausibility of the first sphere that contains said world. The plausibility of a sentence corresponds to the most-plausible world that satisfies that sentence; we denote the plausibility of α in an epistemic state $\vec{e}$ by $\lfloor \vec{e} \mid \alpha \rfloor$. In case there is no such world in $\vec{e}$, $\lfloor \vec{e} \mid \alpha \rfloor$ cannot be a natural number. For that purpose, we use $\infty \notin \mathbb{P}$ to represent an "undefined" plausibility, with the understanding that $p < \infty$ and $\infty + p = p + \infty = \infty$ for all $p \in \mathbb{P}$ and $\infty + \infty = \infty$. Thus, $\lfloor \vec{e} \mid \alpha \rfloor = \infty$ indicates that all worlds in $\vec{e}$ satisfy $\neg\alpha$. To avoid confusion, we always make explicit when an expression may take the value ∞.

Definition 4.3.2 The truth relation $\models$ of $\mathcal{BO}$ is defined with respect to an epistemic state $\vec{e}$ and a world w:

$\mathcal{BO}1$. $\vec{e}, w \models P(t_1, \ldots, t_k)$ iff $w[P(n_1, \ldots, n_k)] = 1$ where $n_i = w(t_i)$;

$\mathcal{BO}2$. $\vec{e}, w \models (t_1 = t_2)$ iff n_1 and n_2 are identical names where $n_i = w(t_i)$;

$\mathcal{BO}3$. $\vec{e}, w \models \neg\alpha$ iff $\vec{e}, w \not\models \alpha$;

$\mathcal{BO}4$. $\vec{e}, w \models (\alpha \vee \beta)$ iff $\vec{e}, w \models \alpha$ or $\vec{e}, w \models \beta$;

$\mathcal{BO}5$. $\vec{e}, w \models \exists x\, \alpha$ iff $\vec{e}, w \models \alpha^x_n$ for some name n;

$\mathcal{BO}6$. $\vec{e}, w \models \mathbf{B}(\alpha \Rightarrow \beta)$ iff
for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \alpha \rfloor$ and $w' \in e_p$, then $\vec{e}, w' \models (\alpha \supset \beta)$;

$\mathcal{BO}7$. $\vec{e}, w \models \mathbf{O}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ iff

$$\text{for all } p \in \mathbb{P},\ w' \in e_p \text{ iff } \vec{e}, w' \models \textstyle\bigwedge_{i:\lfloor \vec{e} \mid \alpha_i \rfloor \geq p}(\alpha_i \supset \beta_i);$$

where $\lfloor \vec{e} \mid \alpha \rfloor = \min\{p \mid p = \infty \text{ or } \vec{e}, w \models \alpha \text{ for some } w \in e_p\}$ denotes the plausibility of α in $\vec{e}$.

We allow ourselves to omit $\vec{e}$ or w when writing $\vec{e}, w \models \alpha$ by the same convention as for $\mathcal{OL}$. In particular, for subjective sentences σ we may just write $\vec{e} \models \sigma$.

Moreover, we will subscript the $\models$ symbol with the name of the logic to avoid ambiguity when necessary. For example, $\models_{\mathcal{OL}}$ refers to truth of $\mathcal{OL}$ from Definition 3.7.1, and $\models_{\mathcal{BO}}$ refers to truth in $\mathcal{BO}$ as defined above.

4.4 Properties of conditional belief

The conditional belief operator $\mathbf{B}(\alpha \Rightarrow \beta)$ is used to form queries about the agent's beliefs. It expresses the agent's belief that if α was true, then β would be true as well. Or in terms of possible worlds, the most-plausible α-worlds must satisfy β as well. This operator turns out to be a quite versatile tool. Besides the general conditional or counterfactual reading, the following intuitions can be captured with it.

- $\mathbf{B}(\text{TRUE} \Rightarrow \alpha)$ represents ordinary belief in α: it holds when all most-plausible worlds satisfy α. It is therefore abbreviated by $\mathbf{B}\alpha$.
- $\mathbf{B}(\neg\alpha \Rightarrow \text{FALSE})$ captures the usual semantics of indefeasible knowledge of α: all worlds at all spheres satisfy α. It is therefore abbreviated by $\mathbf{K}\alpha$.
- $\mathbf{B}(\alpha \vee \beta \Rightarrow \neg\beta)$ asserts that α is strictly more plausible than β: the first $(\alpha \vee \beta)$-worlds must be $\neg\beta$-worlds.
- $\neg\mathbf{B}(\alpha \Rightarrow \neg\beta)$ says that β would be considered possible if α were true: among the most-plausible α-worlds at least one is a β-world. In particular, $\neg\mathbf{B}(\alpha \Rightarrow \text{FALSE})$ and $\neg\mathbf{K}\neg\alpha$ express that there is at least one α-world.

In this section we examine a few properties of conditional belief. En route, we shall familiarize ourselves with the formalism.

The following alternative formulation of its semantics is sometimes more convenient to work with than Rule $\mathcal{BO}6$.

Theorem 4.4.1 *$\vec{e} \models \mathbf{B}(\alpha \Rightarrow \beta)$ iff $\lfloor \vec{e} \mid \alpha \rfloor = \infty$ or $\vec{e}, w \models (\alpha \supset \beta)$ for all $w \in e_{\lfloor \vec{e} \mid \alpha \rfloor}$.*

Proof. For the only-if direction let $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \beta)$. Then by Rule $\mathcal{BO}6$, for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \alpha \rfloor$ and $w \in e_p$, then $\vec{e}, w \models (\alpha \supset \beta)$. If $\lfloor \vec{e} \mid \alpha \rfloor = \infty$, the right-hand side of the theorem trivially holds. Otherwise $\vec{e}, w \models (\alpha \supset \beta)$ for all $w \in e_{\lfloor \vec{e} \mid \alpha \rfloor}$, and the right-hand side holds again.

For the if direction first let $\lfloor \vec{e} \mid \alpha \rfloor = \infty$. Then $\vec{e}, w \not\models \alpha$ for all $w \in e_p$ and $p \in \mathbb{P}$. Hence $\vec{e}, w \models (\alpha \supset \beta)$ for all $w \in e_p$ and $p \in \mathbb{P}$, and so $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \beta)$ by Rule $\mathcal{BO}6$. Now let $\lfloor \vec{e} \mid \alpha \rfloor \neq \infty$ and $\vec{e}, w \models (\alpha \supset \beta)$ for all $w \in e_{\lfloor \vec{e} \mid \alpha \rfloor}$. By the concentricity constraint in Definition 4.3.1, $e_1 \subseteq \ldots \subseteq e_{\lfloor \vec{e} \mid \alpha \rfloor}$. Thus for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \alpha \rfloor$ and $w \in e_p$, then $\vec{e}, w \models (\alpha \supset \beta)$, which by Rule $\mathcal{BO}6$ gives $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \beta)$. □

Another easy exercise is to confirm that $\mathbf{K}\alpha$ indeed expresses knowledge of α as claimed above.

Theorem 4.4.2 *$\vec{e} \models \mathbf{K}\alpha$ iff $\vec{e}, w \models \alpha$ for all $w \in e_p$ and $p \in \mathbb{P}$.*

Proof. For the only-if direction, let $\vec{e} \models \mathbf{K}\alpha$. By Rule $\mathcal{BO}6$, for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \neg\alpha \rfloor$ and $w \in e_p$, then $\vec{e}, w \models (\neg\alpha \supset \textsc{false})$, which simplifies to $\vec{e}, w \models \alpha$ (*). We show by induction on p that $p \leq \lfloor \vec{e} \mid \neg\alpha \rfloor$ for all $p \in \mathbb{P}$, which immediately gives us the right-hand side of the theorem. The base case holds trivially. For the induction step, suppose $p \leq \lfloor \vec{e} \mid \neg\alpha \rfloor$. Then $\vec{e}, w \models \alpha$ for all $w \in e_p$ by (*), and thus $\lfloor \vec{e} \mid \neg\alpha \rfloor > p$, that is, $p + 1 \leq \lfloor \vec{e} \mid \neg\alpha \rfloor$.

Conversely, let $\vec{e}, w \models \alpha$ for all $w \in e_p$ and $p \in \mathbb{P}$. Then $\lfloor \vec{e} \mid \neg\alpha \rfloor = \infty$, and by Rule $\mathcal{BO}6$, $\vec{e} \models \mathbf{K}\alpha$. □

Next, we prove that $\mathbf{B}(\alpha \vee \beta \Rightarrow \neg\beta)$ says that α is more plausible than β.

Theorem 4.4.3 *$\vec{e} \models \mathbf{B}(\alpha \vee \beta \Rightarrow \neg\beta)$ iff $\lfloor \vec{e} \mid \alpha \rfloor < \lfloor \vec{e} \mid \beta \rfloor$ or $\lfloor \vec{e} \mid \alpha \rfloor = \lfloor \vec{e} \mid \beta \rfloor = \infty$.*

Proof. $\vec{e} \models \mathbf{B}(\alpha \vee \beta \Rightarrow \neg\beta)$ iff (by Theorem 4.4.1) $\lfloor \vec{e} \mid \alpha \vee \beta \rfloor = \infty$ or $\vec{e}, w \models (\alpha \vee \beta \supset \neg\beta)$ for all $w \in e_{\lfloor \vec{e} \mid \alpha \vee \beta \rfloor}$. The former is equivalent to $\lfloor \vec{e} \mid \alpha \rfloor = \lfloor \vec{e} \mid \beta \rfloor = \infty$. The latter holds iff $\vec{e}, w \not\models \beta$ for all $w \in e_{\lfloor \vec{e} \mid \alpha \vee \beta \rfloor}$ iff (by concentricity of $\vec{e}$) $\lfloor \vec{e} \mid \beta \rfloor > \lfloor \vec{e} \mid \alpha \vee \beta \rfloor = \lfloor \vec{e} \mid \alpha \rfloor$. □

The following theorem establishes several general properties of the conditional belief operator. As usual, neither transitivity nor monotonicity nor contraposition hold for conditional beliefs (Properties (i), (ii), (iii)). Knowledge and belief are closed under modus ponens from material implications and from counterfactual conditionals (Properties (iv), (v), (vi)), and what is believed is a subset of what is known (Property (vii)). The abbreviations $\mathbf{B}\alpha$ and $\mathbf{K}\alpha$ both are positively and negatively introspective (Properties (vii), (viii), (ix)). Hence, they are K45 operators (Fagin, Halpern, et al. 1995). The Barcan formula is satisfied as well (Property (x)) and the agent is moreover omniscient (Property (xi)). Somewhat surprising is perhaps Property (xii): when a conditional is nested in another conditional's consequent, then the outer conditional's antecedent is irrelevant to the inner conditional. Alternatively one could condition the nested belief on the outer conditional's antecedent as well. However, our simple semantics is advantageous when it comes to the representation theorem and later, in Chapter 5, belief regression.

Theorem 4.4.4

(i) $\not\models \mathbf{B}(\alpha \Rightarrow \beta) \land \mathbf{B}(\beta \Rightarrow \gamma) \supset \mathbf{B}(\alpha \Rightarrow \gamma)$*;*

(ii) $\not\models \mathbf{B}(\alpha \Rightarrow \gamma) \supset \mathbf{B}(\alpha \land \beta \Rightarrow \gamma)$*;*

(iii) $\not\models \mathbf{B}(\alpha \Rightarrow \beta) \equiv \mathbf{B}(\neg\beta \Rightarrow \neg\alpha)$*;*

(iv) $\models \mathbf{B}\alpha \land \mathbf{B}(\alpha \supset \beta) \supset \mathbf{B}\beta$*;*

(v) $\models \mathbf{K}\alpha \land \mathbf{K}(\alpha \supset \beta) \supset \mathbf{K}\beta$*;*

(vi) $\models \mathbf{B}\alpha \land \mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{B}\beta$*;*

(vii) $\models \mathbf{K}\alpha \supset \mathbf{B}\alpha$*;*

(viii) $\models \mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{K}\mathbf{B}(\alpha \Rightarrow \beta)$*;*

(ix) $\models \neg\mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{K}\neg\mathbf{B}(\alpha \Rightarrow \beta)$*;*

(x) $\models \forall x\, \mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{B}(\alpha \Rightarrow \forall x\, \beta)$ *where x does not occur freely in α;*

(xi) $\models \mathbf{K}\alpha$ *if* $\models \alpha$*;*

(xii) $\models \mathbf{B}(\alpha \Rightarrow \mathbf{B}(\beta \Rightarrow \gamma)) \land \neg\mathbf{K}\neg\alpha \supset \mathbf{B}(\beta \Rightarrow \gamma)$.

Proof. (i) We show that $\vec{e} \not\models \mathbf{B}(\neg R \Rightarrow \textsc{true}) \land \mathbf{B}(\textsc{true} \Rightarrow R) \supset \mathbf{B}(\neg R \Rightarrow R)$ for $\vec{e}$ with $e_1 = \{w \mid w \models R\}$ and $e_2 = \{w \mid w \models \textsc{true}\}$. Firstly, $\vec{e} \models \mathbf{B}(\neg R \Rightarrow \textsc{true})$ iff (by Theorem 4.4.1) $\lfloor \vec{e} \mid \neg R \rfloor = \infty$ or $w \models \neg R \supset \textsc{true}$ for all $w \in e_{\lfloor \vec{e} \mid \neg R \rfloor}$, which trivially holds. Secondly, $\vec{e} \models \mathbf{B}(\textsc{true} \Rightarrow R)$ iff (by Theorem 4.4.1) $\lfloor \vec{e} \mid \textsc{true} \rfloor = \infty$ or $w \models \textsc{true} \supset R$ for all $w \in e_{\lfloor \vec{e} \mid \textsc{true} \rfloor}$, which holds by definition of e_1. However, $\vec{e} \models \mathbf{B}(\neg R \Rightarrow R)$ iff $\lfloor \vec{e} \mid \neg R \rfloor = \infty$ or $w \models \neg R \supset R$ for all $w \in e_{\lfloor \vec{e} \mid \neg R \rfloor}$, which is false because $\lfloor \vec{e} \mid \neg R \rfloor = 2$ and $w \not\models R$ for some $w \in e_2$.

(ii) Let R and $\vec{e}$ be as in the previous case. We showed that $\vec{e} \models \mathbf{B}(\textsc{true} \Rightarrow R)$, but $\vec{e} \not\models \mathbf{B}(\neg R \Rightarrow R)$, so clearly strengthening the premise in $\mathbf{B}(\textsc{true} \Rightarrow R)$ by $\neg R$ renders it false: $\vec{e} \not\models \mathbf{B}(\textsc{true} \land \neg R \Rightarrow R)$.

(iii) Again let R and $\vec{e}$ be as in the first case. We show $\vec{e} \not\models \mathbf{B}(\textsc{true} \Rightarrow R) \equiv \mathbf{B}(\neg R \Rightarrow \neg\textsc{true})$, which is just what the abbreviation $\mathbf{B}R \equiv \mathbf{K}R$ stands for. In the first case we already showed that $\vec{e} \models \mathbf{B}(\textsc{true} \Rightarrow R)$. However, $w \not\models R$ for some $w \in e_2$, so $\vec{e} \not\models \mathbf{K}R$ by Theorem 4.4.2.

(iv) We show that $\vec{e} \models \mathbf{B}\alpha \wedge \mathbf{B}(\alpha \supset \beta) \supset \mathbf{B}\beta$ for all $\vec{e}$. Let $\vec{e} \models \mathbf{B}\alpha \wedge \mathbf{B}(\alpha \supset \beta)$. We need to show that $\vec{e} \models \mathbf{B}\beta$, which by Theorem 4.4.1 holds iff $\lfloor \vec{e} \mid \text{TRUE} \rfloor = \infty$ or $\vec{e}, w \models \beta$ for all $w \in e_{\lfloor \vec{e} \mid \text{TRUE} \rfloor}$. Suppose $\lfloor \vec{e} \mid \text{TRUE} \rfloor \neq \infty$, for otherwise $\vec{e} \models \mathbf{B}\beta$ follows trivially. From the antecedent $\vec{e} \models \mathbf{B}\alpha \wedge \mathbf{B}(\alpha \supset \beta)$ we obtain by Theorem 4.4.1 that $\vec{e}, w \models \alpha \wedge (\alpha \supset \beta)$ for all $w \in e_{\lfloor \vec{e} \mid \text{TRUE} \rfloor}$. Thus $\vec{e}, w \models \beta$ for all $w \in e_{\lfloor \vec{e} \mid \text{TRUE} \rfloor}$, and so $\vec{e} \models \mathbf{B}\beta$.

(v) We show that $\vec{e} \models \mathbf{K}\alpha \wedge \mathbf{K}(\alpha \supset \beta) \supset \mathbf{K}\beta$ for all $\vec{e}$. Let $\vec{e} \models \mathbf{K}\alpha \wedge \mathbf{K}(\alpha \supset \beta)$. By Theorem 4.4.2, $\vec{e}, w \models \alpha$ and $\vec{e}, w \models (\alpha \supset \beta)$ for all $w \in e_p$ and $p \in \mathbb{P}$. Hence, $\vec{e}, w \models \beta$ for all $w \in e_p$ and $p \in \mathbb{P}$, and so $\vec{e}, w \models \mathbf{K}\beta$ by Theorem 4.4.2.

(vi) It suffices to show $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{B}(\alpha \supset \beta)$ for all $\vec{e}$; the property then follows with Property (iv). According to Rule $\mathcal{BO}6$ we have

- $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \beta)$ iff
  for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \alpha \rfloor$ and $w' \in e_p$, then $\vec{e}, w' \models (\alpha \supset \beta)$;
- $\vec{e} \models \mathbf{B}(\alpha \supset \beta)$ iff
  for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \text{TRUE} \rfloor$ and $w' \in e_p$, then $\vec{e}, w' \models \text{TRUE} \supset (\alpha \supset \beta)$.

We show that the right-hand side of the first line subsumes the right-hand side of the second. Clearly, $\vec{e}, w' \models \text{TRUE} \supset (\alpha \supset \beta)$ iff $\vec{e}, w' \models (\alpha \supset \beta)$. So it only remains to be shown that the second line's if-condition is at least as strong as the first line's. It suffices to show $\lfloor \vec{e} \mid \alpha \rfloor \geq \lfloor \vec{e} \mid \text{TRUE} \rfloor$, which clearly since for any w with $w \models \alpha$, also $w \models \text{TRUE}$.

(vii) Let $\vec{e} \models \mathbf{K}\alpha$. Then by Theorem 4.4.2, $\vec{e}, w \models \alpha$ for all $w \in e_p$ and $p \in \mathbb{P}$, particularly when $p \leq \lfloor \vec{e} \mid \text{TRUE} \rfloor$. Thus by Rule $\mathcal{BO}6$, $\vec{e} \models \mathbf{B}\alpha$.

(viii) Let $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \beta)$. Then $\vec{e}, w \models \mathbf{B}(\alpha \Rightarrow \beta)$ for arbitrary w, and particularly for all $w \in e_p$ and $p \in \mathbb{P}$. Thus by Theorem 4.4.2, $\vec{e} \models \mathbf{KB}(\alpha \Rightarrow \beta)$.

(ix) Let $\vec{e} \not\models \mathbf{B}(\alpha \Rightarrow \beta)$. Then similar to the above, $\vec{e}, w \not\models \mathbf{B}(\alpha \Rightarrow \beta)$ for arbitrary w, and particularly for all $w \in e_p$ and $p \in \mathbb{P}$. Thus by Theorem 4.4.2, $\vec{e} \models \mathbf{K}\neg\mathbf{B}(\alpha \Rightarrow \beta)$.

(x) Let $\vec{e} \models \forall x \mathbf{B}(\alpha \Rightarrow \beta)$. By Rules $\mathcal{BO}3$, $\mathcal{BO}5$, and $\mathcal{BO}6$, for all standard names n, for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \alpha \rfloor$ and $w \in e_p$, then $\vec{e}, w \models \alpha \supset \beta^x_n$. Reintroducing the quantifier by Rules $\mathcal{BO}3$ and $\mathcal{BO}5$ in front of β yields that for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \alpha \rfloor$ and $w \in e_p$, then $\vec{e}, w \models \alpha \supset \forall x \beta$. Thus by Rule $\mathcal{BO}6$, $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \forall x \beta)$.

(xi) Let $\vec{e}, w \models \alpha$ for all $\vec{e}, w$. Then $\vec{e}, w \models \neg\alpha \supset \text{FALSE}$ for all $w \in e_p$ and $p \in \mathbb{P}$ for all $\vec{e}$, so by Theorem 4.4.2, $\vec{e} \models \mathbf{K}\alpha$ follows.

(xii) Let $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \mathbf{B}(\beta \Rightarrow \gamma)) \wedge \neg\mathbf{K}\neg\alpha$. The first assumption implies by Theorem 4.4.1 that $\lfloor \vec{e} \mid \alpha \rfloor \neq \infty$ or $\vec{e}, w \models \alpha \supset \mathbf{B}(\beta \Rightarrow \gamma)$ for all $w \in e_{\lfloor \vec{e} \mid \alpha \rfloor}$. The second assumption implies by Theorem 4.4.2 that $\lfloor \vec{e} \mid \alpha \rfloor \neq \infty$, and thus $\vec{e}, w \models \alpha$ for some $w \in e_{\lfloor \vec{e} \mid \alpha \rfloor}$. Hence $\vec{e}, w \models \alpha \wedge (\alpha \supset \mathbf{B}(\beta \Rightarrow \gamma))$ for that w, so $\vec{e} \models \mathbf{B}(\beta \Rightarrow \gamma)$. □

Before we turn to the unique-model property of only-believing in the next section, we observe some relations with ordinary conditional belief. To begin with, only-believing is stronger than ordinary conditional belief.

Theorem 4.4.5 $\models \mathbf{O}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\} \supset \bigwedge_i \mathbf{B}(\alpha_i \Rightarrow \beta_i)$.

Proof. Suppose $\vec{e} \models \mathbf{O}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$. By Rule $\mathcal{BO}7$, for every $p \in \mathbb{P}$, $w \in e_p$ iff $\vec{e}, w \models \bigwedge_{i:\lfloor \vec{e} \mid \alpha_i \rfloor \geq p} (\alpha_i \supset \beta_i)$. Hence for every $p \in \mathbb{P}$, if $w \in e_p$ and $p \leq \lfloor \vec{e} \mid \alpha_i \rfloor$, then $\vec{e}, w \models (\alpha_i \supset \beta_i)$. By Rule $\mathcal{BO}6$, $\vec{e} \models \mathbf{B}(\alpha_i \Rightarrow \beta_i)$. □

In the case of objective α_i, β_i there is also a converse relation of conditional belief and only-believing, as we shall see in Theorem 4.6.2.

Moreover, conditional belief can be conjoined to the only-believing.

Theorem 4.4.6 *Let* $\Gamma = \{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$.
Then $\models \mathbf{O}\Gamma \wedge \mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{O}\Gamma \cup \{\alpha \Rightarrow \beta\}$.

Proof. Let $\vec{e} \models \mathbf{O}\Gamma \wedge \mathbf{B}(\alpha \Rightarrow \beta)$. Then for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \alpha \rfloor$ and $w \in e_p$, $\vec{e}, w \models (\alpha \supset \beta)$. To prove $\vec{e} \models \mathbf{O}\Gamma \cup \{\alpha \Rightarrow \beta\}$ we show the right-hand side of Rule $\mathcal{BO}7$. For the only-if direction, for all $p \in \mathbb{P}$ and $w \in e_p$, we have by assumption $\vec{e}, w \models \bigwedge_{i:\lfloor \vec{e} \mid \alpha_i \rfloor \geq p} (\alpha_i \supset \beta_i) \wedge \bigwedge_{\lfloor \vec{e} \mid \alpha \rfloor \geq p} (\alpha \supset \beta)$. Conversely, for every $p \in \mathbb{P}$, if $\vec{e}, w \models \bigwedge_{i:\lfloor \vec{e} \mid \alpha_i \rfloor \geq p} (\alpha_i \supset \beta_i) \wedge \bigwedge_{\lfloor \vec{e} \mid \alpha \rfloor \geq p} (\alpha \supset \beta)$, then $\vec{e}, w \models \bigwedge_{i:\lfloor \vec{e} \mid \alpha_i \rfloor \geq p} (\alpha_i \supset \beta_i)$, which implies $w \in e_p$. □

Note that there are analogous results for only-knowing in $\mathcal{OL}$, namely Theorems 3.7.3 and Theorem 3.7.4.

4.5 Unique-model property of only-believing

The idea behind only-believing is to determine all the agent believes – which particularly includes what she is ignorant of. It is therefore a convenient way to specify what the agent believes from a (conditional) knowledge base.

Here we show that only-believing always has a *unique* model provided the conditionals are objective. The proof proceeds by two lemmas. The first says that there is at most one model, and the second says that there always is at least one. The proof of the second lemma also shows how such a model can be determined.

Lemma 4.5.1 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be objective.*
If $\vec{e} \models \mathbf{O}\Gamma$ *and* $\vec{e}\,' \models \mathbf{O}\Gamma$, *then* $\vec{e} = \vec{e}\,'$.

Proof. Let $\vec{e} \models \mathbf{O}\Gamma$ and $\vec{e}\,' \models \mathbf{O}\Gamma$. We show by induction on $p \in \mathbb{P}$ that $e_p = e'_p$ and that $\lfloor \vec{e} \mid \phi_i \rfloor > p$ iff $\lfloor \vec{e}\,' \mid \phi_i \rfloor > p$ for all i. For the base case consider $p = 1$. By Rule $\mathcal{BO}7$,

$w \in e_1$ iff $w \models \bigwedge_{1 \le i \le m}(\phi_i \supset \psi_i)$ iff $w \in e'_1$. Thus $e_1 = e'_1$, and $\lfloor \vec{e} \mid \phi_i \rfloor > 1$ iff $w \not\models \phi_i$ for all $w \in e_1 = e'_1$ iff $\lfloor \vec{e}' \mid \phi_i \rfloor > 1$.

For the induction step suppose the statement holds for $p-1$. By induction, $\lfloor \vec{e} \mid \phi_i \rfloor \ge p$ iff $\lfloor \vec{e}' \mid \phi_i \rfloor \ge p$ for all i (*). By Rule $\mathcal{BO}7$, $w \in e_p$ iff $w \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \ge p}(\phi_i \supset \psi_i)$ iff (by (*)) $w \models \bigwedge_{i:\lfloor \vec{e}' \mid \phi_i \rfloor \ge p}(\phi_i \supset \psi_i)$ iff $w \in e'_p$. Thus $e_p = e'_p$, and $\lfloor \vec{e} \mid \phi_i \rfloor > p$ iff $w \not\models \phi_i$ for all $w \in e_p = e'_p$ iff $\lfloor \vec{e}' \mid \phi_i \rfloor > p$. □

Lemma 4.5.2 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be objective.*
Then there is an $\vec{e} = \langle e_1, \ldots, e_{m+1} \rangle$ *such that* $\vec{e} \models \mathbf{O}\Gamma$.

Proof. Let $\vec{e} = \langle e_1, \ldots, e_{m+1} \rangle$, where $e_p = \{w \mid w \models \bigwedge_{i:\lfloor \langle e_1,\ldots,e_{p-1} \rangle \mid \phi_i \rfloor \ge p}(\phi_i \supset \psi_i)\}$ where $\langle\rangle$ shall stand for $\langle\{\}\rangle$. This is well-defined as the right-hand side for e_p only refers to $e_1, \ldots, e_{p-1}$. Note that $\lfloor \langle e_1, \ldots, e_{p-1} \rangle \mid \alpha \rfloor \ge p$ iff $\lfloor \vec{e} \mid \alpha \rfloor \ge p$ for any objective α (*). To see that for all i either $\lfloor \vec{e} \mid \phi_i \rfloor \le m$ or $\lfloor \vec{e} \mid \phi_i \rfloor = \infty$ (**), suppose there is a "hole" in the plausibility ranking, that is, there is some p and i such that $p + 1 = \lfloor \vec{e} \mid \phi_i \rfloor \ne \infty$, and $\lfloor \vec{e} \mid \phi_j \rfloor \ne p$ and for all j. Then $w \in e_p$ iff (by (*)) $w \models \bigwedge_{k:\lfloor \vec{e} \mid \phi_k \rfloor \ge p}(\phi_k \supset \psi_k)$ iff (since p is a hole) $w \models \bigwedge_{k:\lfloor \vec{e} \mid \phi_k \rfloor \ge p+1}(\phi_k \supset \psi_k)$ iff $w \in e_{p+1}$. Then $w \models \phi_i$ for some $w \in e_{p+1} = e_p$, which contradicts the assumption $p + 1 = \lfloor \vec{e} \mid \phi_i \rfloor$ and thus confirms (**). By (*) and (**), $\vec{e}$ satisfies Rule $\mathcal{BO}7$. □

Together, Lemmas 4.5.1 and 4.5.2 constitute the unique-model property of only-believing, a fundamental property of only-believing.

Theorem 4.5.3 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be objective.*
Then there is a unique $\vec{e} = \langle e_1, \ldots, e_{m+1} \rangle$ *such that* $\vec{e} \models \mathbf{O}\Gamma$.

Proof. By Lemma 4.5.2, $\vec{e}$ exists, and by Lemma 4.5.1, it unique. □

An immediate corollary is that only-believing determines the truth of subjective formulas.

Corollary 4.5.4 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be objective and* σ *be subjective.*
Then $\mathbf{O}\Gamma \models \sigma$ *and* $\mathbf{O}\Gamma \models \neg\sigma$.

To conclude this section, let us illustrate by means of Example 4.1.1 how only-believing uniquely determines the agent's beliefs and how the corresponding epistemic state can be generated.

Example 4.5.5 Let Γ contain the conditionals from the formalization in Example 4.2.2. The first sphere e_1 of $\vec{e} \models \mathbf{O}\Gamma$ contains all worlds that satisfy all materialized conditionals from Γ:

$$e_1 = \{w \mid w \models (\neg\text{Aussie} \lor \neg\text{Italian}) \land (\neg\text{Aussie} \lor \text{Eats(roo)}) \land \\ (\text{Italian} \lor \text{Veggie}) \land (\text{Italian} \lor \text{Aussie}) \land \mu\}$$

where $\mu = \text{Meat(roo)} \land \forall x(\text{Veggie} \land \text{Meat}(x) \supset \neg\text{Eats}(x))$ represents our knowledge about meat and vegetarians.

For the next sphere, we need to figure out the plausibilities $\lfloor \vec{e} \mid \phi \rfloor$ for the conditionals $\phi \Rightarrow \psi \in \Gamma$. To begin with, we need to answer if $\lfloor \vec{e} \mid \text{Aussie} \rfloor \geq 2$, that is, if e_1 is inconsistent with Aussie. To give the answer, we can split on Veggie: from Veggie we obtain $\neg$Eats(roo) (by μ) and thus $\neg$Aussie; on the other hand, from $\neg$Veggie we infer Italian and thus $\neg$Aussie; so indeed e_1 is inconsistent with Aussie, that is, $\lfloor \vec{e} \mid \text{Aussie} \rfloor \geq 2$. By the same argument, $\lfloor \vec{e} \mid \neg\text{Italian} \rfloor \geq 2$. It is moreover easy to see that e_1 is consistent and thus $\lfloor \vec{e} \mid \text{Italian} \rfloor = \lfloor \vec{e} \mid \text{TRUE} \rfloor = 1$. Hence the conditionals Aussie $\Rightarrow$ $\neg$Italian, Aussie $\Rightarrow$ Eats(roo), $\neg$Italian $\Rightarrow$ Aussie, plus the knowledge about meat and vegetarians determine the second sphere:

$$e_2 = \{w \mid w \models (\neg\text{Aussie} \lor \neg\text{Italian}) \land (\neg\text{Aussie} \lor \text{Eats(roo)}) \land (\text{Italian} \lor \text{Aussie}) \land \mu\}.$$

Again we need to check which premises are consistent with e_2, and only the remaining conditionals determine the next sphere e_3. It is easy to see that $\lfloor \vec{e} \mid \text{Aussie} \rfloor = \lfloor \vec{e} \mid \neg\text{Italian} \rfloor = 2$, so for the third and last sphere:

$$e_3 = \{w \mid w \models \mu\}.$$

Since $\vec{e}$ is the unique model of $\mathbf{O}\Gamma$ in $\mathcal{BO}$, it determines our beliefs. For example, $\mathbf{O}\Gamma \models \mathbf{B}(\neg\text{Italian} \Rightarrow \neg\text{Veggie})$ since $\lfloor \vec{e} \mid \neg\text{Italian} \rfloor = 2$ and $w \models \neg\text{Italian} \supset \neg\text{Veggie}$ for all $w \in e_2$.

Finally we remark that the unique-model property does not extend to subjective formulas. For example, $\mathbf{O}\{\text{TRUE} \Rightarrow \neg\mathbf{B}\text{FALSE}\}$ has two models, namely $\langle\{\}\rangle$ and $\langle W\rangle$, where W denotes the set of all worlds.

Let us first verify $\langle\{\}\rangle \models \mathbf{O}\{\text{TRUE} \Rightarrow \neg\mathbf{B}\text{FALSE}\}$. It is immediate that $\lfloor \langle\{\}\rangle \mid \text{TRUE} \rfloor = \infty$, so by Rule $\mathcal{BO}7$ we merely need to prove that $w \in \{\}$ iff $\langle\{\}\rangle, w \models \neg\mathbf{B}\text{FALSE}$. The only-if direction is vacuously true. For the converse, observe that $\langle\{\}\rangle \models \mathbf{B}\text{FALSE}$, so for every $w \notin \{\}$, $\langle\{\}\rangle \not\models \neg\mathbf{B}\text{FALSE}$.

Next we verify $\langle W\rangle \models \mathbf{O}\{\text{TRUE} \Rightarrow \neg\mathbf{B}\text{FALSE}\}$. Clearly, $\lfloor \langle W\rangle \mid \text{TRUE} \rfloor = 1$, so by Rule $\mathcal{BO}7$ we need to show that $w \in W$ iff $\langle W\rangle, w \models \neg\mathbf{B}\text{FALSE}$. The if direction is vacuously true. For the converse, $\langle W \rangle \models \neg\mathbf{B}\text{FALSE}$ iff $w' \models \text{TRUE}$ for some $w' \in W$, which is true.

4.6 Relationship to $\mathcal{OL}$

Many of the above results about $\mathcal{BO}$ have counterparts of in $\mathcal{OL}$: only-knowing and only-believing imply ordinary knowledge and belief, respectively (Theorems 3.7.3 and 4.4.5); knowledge and belief can be conjoined with only-knowing and only-believing, respectively (Theorems 3.7.4 and 4.4.6); the knowledge and belief operators are K45 operators (Theorems 3.7.2 and 4.4.4); and in the objective case both only-knowing and only-believing have a unique model (Theorems 3.7.5 and 4.5.3) and thus determine the truth of subjective sentences (Corollaries 3.7.7 and 4.5.4). This section further examines the close relationship between $\mathcal{OL}$ and $\mathcal{BO}$.

Only-believing expresses *all* the agent believes, that is, it maximizes the non-beliefs. Intuitively, this should go along with maximizing the epistemic state, that is, with taking into consideration as many possible worlds as possible. For only-knowing, this intuition was already confirmed in Theorem 3.7.6. The next theorem shows that it also holds true for only-believing objective conditionals.

Definition 4.6.1 We say $\vec{e}$ is *maximal* with $\vec{e} \models \sigma$ for a subjective sentence σ when no worlds can be added to any plausibility sphere without falsifying σ, that is, $\vec{e}\,' \not\models \sigma$ for all $\vec{e}\,'$ with $e'_p \supseteq e_p$ for all $p \in \mathbb{P}$ and $e'_{p'} \supsetneq e_{p'}$ for some $p' \in \mathbb{P}$.

Theorem 4.6.2 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be objective. Then* $\vec{e} \models \mathbf{O}\Gamma$ *iff* $\vec{e}$ *is maximal such that* $\vec{e} \models \bigwedge_i \mathbf{B}(\phi_i \Rightarrow \psi_i)$.

Proof. For the if direction suppose $\vec{e} \models \mathbf{B}(\phi_i \Rightarrow \psi_i)$ for all i and $\vec{e}$ is maximal with that property. Let $p \in \mathbb{P}$ and w be a world. By Rule $\mathcal{BO}$6, if $p \le \lfloor \vec{e} \mid \phi_i \rfloor$ and $w \in e_p$, then $w \models \phi_i \supset \psi_i$ for all i. Since $\vec{e}$ is maximal, if $p \le \lfloor \vec{e} \mid \phi_i \rfloor$ and $w \notin e_p$, then $w \not\models \phi_i \supset \psi_i$ for some i. Thus $w \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \ge p} (\phi_i \supset \psi_i)$ iff $w \in e_p$. Hence $\vec{e} \models \mathbf{O}\Gamma$.

For the only-if direction suppose $\vec{e} \models \mathbf{O}\Gamma$. By Theorem 4.4.5, $\vec{e} \models \bigwedge_i \mathbf{B}(\phi_i \Rightarrow \psi_i)$. Suppose $\vec{e}$ is not maximal. Then there is some $\vec{e}\,'$ such that $\vec{e}\,' \models \bigwedge_i \mathbf{B}(\phi_i \Rightarrow \psi_i)$ and $e'_p \supseteq e_p$ for all $p \in \mathbb{P}$ and $e'_{p'} \supsetneq e_{p'}$ for some $p' \in \mathbb{P}$. By the if direction, $\vec{e}\,' \models \mathbf{O}\Gamma$, and by Lemma 4.5.1 $\vec{e} = \vec{e}\,'$. Contradiction. □

For that theorem too there is a counterpart in $\mathcal{OL}$, namely Theorem 3.7.6. Given that only-believing and only-knowing apparently have a lot in common, the question arises whether $\mathcal{BO}$ subsumes the logic $\mathcal{OL}$. Indeed the answer is affirmative. We make this precise by embedding $\mathcal{OL}$ in $\mathcal{BO}$.

Definition 4.6.3 Let $\sharp$ be the function from $\mathcal{OL}$ formulas to $\mathcal{BO}$ formulas which is defined by $(\mathbf{O}\alpha)^\sharp = \mathbf{O}\{\neg\alpha^\sharp \Rightarrow \text{FALSE}\}$, and $(\mathbf{K}\alpha)^\sharp = \mathbf{K}\alpha^\sharp$, and inductively for the other operators: $\phi^\sharp = \phi$ for objective ϕ; $(\neg\alpha)^\sharp = \neg\alpha^\sharp$; $(\alpha \vee \beta)^\sharp = (\alpha^\sharp \vee \beta^\sharp)$; $(\exists x\, \alpha)^\sharp = \exists x\, \alpha^\sharp$.

Theorem 4.6.4 *Let α be a sentence of $\mathcal{OL}$. Then $\models_{\mathcal{OL}} \alpha$ iff $\models \alpha^\sharp$.*

The proof can be found in Appendix A.1. It is surprisingly tedious because care must be taken to preserve equivalence when translating a system of spheres to a single set of worlds.

4.7 Relationship to System Z

Pearl (1990) introduced System Z with the goal of tractable reasoning about conditionals. It is fundamentally based on a unique ranking of the conditionals, called Z-ordering. As we shall see in this section, the Z-ordering is essentially equivalent to the ranking imposed by only-believing. Moreover, Pearl's notions of 0- and 1-entailment can be characterized in $\mathcal{BO}$.

System Z is not a logical language but a meta-logical framework. To accord with System Z, we assume for the rest of this section that ϕ, ψ are objective and propositional and Γ contains finitely many conditionals $\phi \Rightarrow \psi$. We now give a few definitions from (Pearl 1990); we only adapt them to our syntax.

Definition 4.7.1 (Pearl 1990) Γ *tolerates* $\phi \Rightarrow \psi$ iff $\bigwedge_{\phi' \Rightarrow \psi' \in \Gamma}(\phi' \supset \psi') \wedge \phi \wedge \psi$ is satisfiable in classical propositional logic. Γ is *consistent* iff for every $\Gamma' \subseteq \Gamma$, there is some $\phi \Rightarrow \psi \in \Gamma'$ such that $\Gamma' \setminus \{\phi \Rightarrow \psi\}$ tolerates $\phi \Rightarrow \psi$.

Lemma 4.7.2 *Γ tolerates $\phi \Rightarrow \psi$ iff $w \models \phi \wedge \bigwedge_{\phi' \Rightarrow \psi' \in \Gamma \cup \{\phi \Rightarrow \psi\}}(\phi' \supset \psi')$ for some w.*

Proof. Since all formulas are assumed to be propositional, they are satisfiable in classical logic iff they are satisfiable in $\mathcal{L}$ by Theorem 3.4.1. The lemma thus follows immediately from Definition 4.7.1. $\square$

Definition 4.7.3 (Pearl 1990) Let $\Gamma_i = \{\phi \Rightarrow \psi \in \Gamma \mid \Gamma \setminus (\Gamma_0 \cup \ldots \Gamma_{i-1}) \text{ tolerates } \phi \Rightarrow \psi\}$ be defined inductively on i.

- For $\phi \Rightarrow \psi \in \Gamma$, the *Z-rank* is defined by $Z(\phi \Rightarrow \psi) = i$ iff $\phi \Rightarrow \psi \in \Gamma_i$.
- For a world w, $Z(w) = \min\{i \mid w \models \bigwedge_{Z(\phi \Rightarrow \psi) \geq i \text{ and } \phi \Rightarrow \psi \in \Gamma}(\phi \supset \psi)\}$.
- For a formula ϕ, $Z(\phi) = \min\{Z(w) \mid w \models \phi\}$.

The following theorem relates the Z-ordering with only-believing and plausibilities.

Theorem 4.7.4 *Let $\vec{e} \models \mathbf{O}\Gamma$, which exists and is unique by Theorem 4.5.3.*

(i) *Γ is inconsistent iff $\lfloor \vec{e} \mid \phi \rfloor = \infty$ for some $\phi \Rightarrow \psi \in \Gamma$.*

(ii) *If Γ is consistent, then $\lfloor \vec{e} \mid \phi \rfloor = Z(\phi \Rightarrow \psi) + 1$ for every $\phi \Rightarrow \psi \in \Gamma$.*

(iii) *If Γ is consistent, then $\min\{p \mid w \in e_p\} = Z(w) + 1$.*

(iv) *If Γ is consistent and ϕ is satisfiable, then $\lfloor \vec{e} \mid \phi \rfloor = Z(\phi) + 1$.*

The proof is in Appendix A.2. Next, let us consider Pearl's notions of 0- and 1-entailment.

Definition 4.7.5 (Pearl 1990) 0-entailment and 1-entailment in the context of Γ are defined as follows:

- $\phi \vdash_0 \psi$ iff $\Gamma \cup \{\phi \Rightarrow \neg\psi\}$ is inconsistent;
- $\phi \vdash_1 \psi$ iff $Z(\phi \wedge \psi) < Z(\phi \wedge \neg\psi)$.

Unfortunately, Definition 4.7.3 leaves the value of Z undefined in some cases. Firstly, it says nothing about the value of $Z(\phi \Rightarrow \psi)$ in case $\phi \Rightarrow \psi \notin \Gamma_i$. Similarly, $Z(w)$ is undefined for some w in case Γ is inconsistent. And $Z(\phi)$ is hence only defined for consistent Γ and satisfiable ϕ.

Requiring Γ to be consistent seems acceptable. But the restriction of $Z(\phi)$ to satisfiable ϕ also means that, for example, $\psi \vdash_1 \psi$ is undefined. To alleviate this, we assume for the following theorem that Pearl implicitly defined $\min\{\} = \infty$. Under that assumption, $Z(w)$ and $Z(\phi)$ are well-defined for all w and ϕ provided that Γ is consistent. Then we can show that 1-entailment corresponds to conditional belief in $\mathcal{BO}$.

Lemma 4.7.6 *Let ϕ be unsatisfiable. Then $\lfloor \vec{e} \mid \phi \rfloor = \infty = Z(\phi)$.*

Proof. By assumption, there is no w such that $w \models \phi$. Then clearly $\lfloor \vec{e} \mid \phi \rfloor = \infty$. Moreover, $Z(\phi) = \min\{\} = \infty$. □

Theorem 4.7.7 *Let Γ be consistent. Then $\phi \vdash_1 \psi$ iff $\mathbf{O}\Gamma \models \neg\mathbf{K}\neg\phi \wedge \mathbf{B}(\phi \Rightarrow \psi)$.*

Proof. $\phi \vdash_1 \psi$ iff $Z(\phi \wedge \psi) < Z(\phi \wedge \neg\psi)$ iff (by Theorem 4.7.4 and Lemma 4.7.6) $\lfloor \vec{e} \mid \phi \wedge \psi \rfloor < \lfloor \vec{e} \mid \phi \wedge \neg\psi \rfloor$ iff $\lfloor \vec{e} \mid \phi \rfloor \neq \infty$ and $w \models (\phi \supset \psi)$ for all $w \in e_{\lfloor \vec{e} \mid \phi \rfloor}$ iff $w \models \phi$ for some $w \in e_p$ and $p \in \mathbb{P}$, and $w \models (\phi \supset \psi)$ for all $w \in e_{\lfloor \vec{e} \mid \phi \rfloor}$ iff (by Theorems 4.4.2 and 4.4.1) $\vec{e} \models \neg\mathbf{K}\neg\phi \wedge \mathbf{B}(\phi \Rightarrow \psi)$. □

There seems to be no equivalent to 0-entailment in $\mathcal{BO}$. Given that Pearl (1990) himself criticizes 0-entailment for being "extremely conservative" and thus missing out on many intuitive consequences, this is probably not a big issue. Still, the following theorem states bounds of 0-entailment in $\mathcal{BO}$.

Theorem 4.7.8 *Let* Γ *be consistent.*
Then $\mathbf{O}\Gamma \models \mathbf{K}(\phi \supset \psi)$ *implies* $\phi \vdash_0 \psi$ *implies* $\mathbf{O}\Gamma \models \neg\mathbf{K}\neg\phi \wedge \mathbf{B}(\phi \Rightarrow \psi)$.

Proof. For the first implication suppose $\mathbf{O}\Gamma \models \mathbf{K}(\phi \supset \psi)$. By Theorem 4.5.3, $\vec{e} \models \mathbf{O}\Gamma$ exists and is unique. Then $\vec{e} \models \mathbf{K}(\phi \supset \psi)$. By Theorem 4.4.2, $w \models (\phi \supset \psi)$ for all $w \in e_p$. Since Γ is consistent, there is some e_p which contains all worlds by Theorem 4.7.4. Hence, $\phi \supset \psi$ is a tautology, and so $\phi \wedge (\phi \supset \neg\psi)$ is unsatisfiable. Thus by Lemma 4.7.2, there is no $\Gamma' \subseteq \Gamma$ that tolerates $\phi \Rightarrow \neg\psi$, so $\phi \vdash_0 \psi$.

As for the second claim, $\phi \vdash_0 \phi$ only if (Pearl 1990) $\phi \vdash_1 \psi$ iff (by Theorem 4.7.7) $\mathbf{O}\Gamma \models \neg\mathbf{K}\neg\phi \wedge \mathbf{B}(\phi \Rightarrow \psi)$. □

4.8 Representation theorem

In this section we investigate if and how entailments involving conditional beliefs can be reduced to non-modal reasoning. In particular, we want to reduce the problem $\mathbf{O}\Gamma \models \mathbf{B}(\alpha \Rightarrow \psi)$ to non-modal first-order entailments.

The concept is due to Levesque (1984b), who developed it for $\mathcal{OL}$. The idea is as follows. Suppose at the first sphere of an epistemic state $\vec{e}$ it is believed that at object #5 is broken, but everything else is not: $e_1 = \{w \mid w \models \gamma_1\}$ where $\gamma_1 = \text{Broken}(x) \equiv x = \#5$. Obviously, $\vec{e} \models \mathbf{B}\exists x \text{Broken}(x)$, because #5 is broken: $\models \gamma_1 \supset \exists x \text{Broken}(x)$ holds. It gets more tricky when x is quantified *outside* of $\mathbf{B}$. Then we need to find a standard name n that we can substitute for x so that $\models \gamma_1 \supset \text{Broken}(x)^x_n$. Obvious choices to try are the standard names that occur in the knowledge base and in the query. Here this is just #5, and indeed $\models \gamma_1 \supset \text{Broken}(x)^x_{\#5}$ comes out true. In general, however, this is not enough, and also objects *not* occurring in the knowledge base and query must be tested. For example, to show $\vec{e} \models \exists x \mathbf{B}\neg\text{Broken}(x)$ we could test #7. Luckily, we do not need to test all (infinitely many) names: Levesque showed that the names from the knowledge base and query plus a single additional one suffice already. The intuitive reason is that names that do not occur in the knowledge base or the query cannot be distinguished.

The above sketch only refers to reasoning on a single sphere, but an epistemic state may have many different spheres. Luckily, their number is bounded by the number of belief conditionals, provided they are objective, according to Theorem 4.5.3. This allows us to generate a formula which deals with every sphere in the above fashion, and check whether this formula is valid. We formalize this idea in the rest of this section.

To begin with, we adopt Levesque's trick of substituting finitely many standard names for free variables.

Definition 4.8.1 (Levesque and Lakemeyer 2001) Let ϕ be an objective sentence and ψ

be an objective formula. Then $\mathrm{RES}[\![\psi, \phi]\!]$ is defined as follows:

- if ψ has no free variables, then

$$\mathrm{RES}[\![\psi, \phi]\!] = \begin{cases} \text{TRUE} & \text{if} \models (\phi \supset \psi); \\ \text{FALSE} & \text{otherwise;} \end{cases}$$

- if x is a free variable in ψ and
  - $\mathcal{N}$ contains the standard names occurring in ϕ or ψ,
  - n' is a new standard name not in $\mathcal{N}$,

  then

$$\mathrm{RES}[\![\psi, \phi]\!] = \bigvee_{n \in \mathcal{N}} ((x = n) \wedge \mathrm{RES}[\![(\psi^x_n), \phi]\!]) \vee \\ \bigwedge_{n \in \mathcal{N}} ((x \neq n) \wedge \mathrm{RES}[\![(\psi^x_{n'}), \phi]\!]^{n'}_x).$$

The case for free variables x tries all names from ϕ and ψ plus another one, n'. Notice that n' is eventually replaced by x again, so that no new standard name is introduced by $\mathrm{RES}[\![\psi, \phi]\!]$. Given an $\mathcal{OL}$ knowledge base $\mathbf{O}\phi$, Levesque would then replace every occurrence $\mathbf{K}\psi$ in the query (perhaps with free variables) with $\mathrm{RES}[\![\psi, \phi]\!]$.

For us, it is not that simple because our knowledge base $\mathbf{O}\Gamma$ consists not just of a single sentence but of usually multiple conditionals, so we cannot simply read off ϕ from the knowledge base. Similarly, it is not immediate how one would form the ψ from a conditional belief in the query. So how could ϕ and ψ look like in our setting? Firstly, it is fundamental is that $\mathbf{O}\Gamma$ for objective Γ can be represented with objective formulas and their number is bounded by Γ as well.

Definition 4.8.2 Let $\Gamma = \{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$. Let $\vec{e} \models \mathbf{O}\Gamma$. An *objective representation* $\vec{\gamma}$ of $\mathbf{O}\Gamma$ is an infinite sequence of objective sentences γ_p, $p \in \mathbb{P}$, such that $e_p = \{w \mid w \models \gamma_p\}$ for all $p \in \mathbb{P}$. We write $\langle \gamma_1, \ldots, \gamma_q \rangle$ if $\gamma_q = \gamma_p$ for all $p \geq q$.

So γ_p represents what is believed at the pth sphere. Luckily, the unique-model property for objective Γ from Theorem 4.5.3 essentially carries over to objective representations. Moreover, as shown in the proof of Lemma 4.5.2, first-order reasoning suffices to determine an objective representation.

Lemma 4.8.3 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be objective and* $\vec{\gamma} = \langle \gamma_1, \ldots, \gamma_{m+1} \rangle$ *be defined inductively by* $\gamma_p = \bigwedge_{i:\gamma_0 \models \neg\phi_i, \ldots, \gamma_{p-1} \models \neg\phi_i} (\phi_i \supset \psi_i)$. *Then* $\vec{\gamma}$ *is an objective represen-*

tation of **O**Γ*, and for every other objective representation* $\vec{\gamma}\,'$, $\models \gamma_p \equiv \gamma'_p$ *for all* $p \in \mathbb{P}$.

Proof. The construction of $\vec{\gamma}$ precisely reflects $\vec{e}$ from Lemma 4.5.2. Since that $\vec{e}$ is unique by Lemma 4.5.1, and by construction $e_p = \{w \mid w \models \gamma_p\}$, $\vec{\gamma}$ is an objective representation and unique (modulo logical equivalence). □

The idea is that the γ_p of an objective representation of **O**Γ will take the place of ϕ in RES$[\![\psi, \phi]\!]$. But what about ψ? Recall that **O**$\Gamma \models$ **B**$(\alpha \Rightarrow \beta)$ iff for all $p \in \mathbb{P}$, if $\neg\alpha$ holds at all spheres $p' < p$, then $\alpha \supset \beta$ holds at sphere p. Provided that Γ and ϕ', ψ' are objective, we can reformulate whether **O**$\Gamma \models$ **B**$(\phi' \Rightarrow \psi')$ using Lemma 4.8.3: for an objective representation $\vec{\gamma}$ of **O**Γ, the entailment **O**$\Gamma \models$ **B**$(\phi' \Rightarrow \psi')$ holds iff for all $1 \leq p \leq m+1$, if RES$[\![\neg\phi', \gamma_{p'}]\!]$ is valid for all $p' < p$, then RES$[\![\phi' \supset \psi', \gamma_p]\!]$ is valid. We can thus define a procedure $\|\alpha\|_{\vec{\gamma}}$ to eliminate all **B**$(\phi' \Rightarrow \psi')$ from α. To cope with non-objective ϕ' or ψ', we simply apply $\|\cdot\|_{\vec{\gamma}}$ recursively from the inside.

Definition 4.8.4 Let α be a formula without **O** and let $\vec{\gamma} = \langle \gamma_1, \ldots, \gamma_q \rangle$ be objective sentences. Then $\|\alpha\|_{\vec{\gamma}}$ is defined inductively:

- $\|\alpha\|_{\vec{\gamma}} = \alpha$ if α is an objective formula;
- $\|\neg\alpha\|_{\vec{\gamma}} = \neg\|\alpha\|_{\vec{\gamma}}$;
- $\|(\alpha_1 \vee \alpha_2)\|_{\vec{\gamma}} = (\|\alpha_1\|_{\vec{\gamma}} \vee \|\alpha_2\|_{\vec{\gamma}})$;
- $\|\exists x\, \alpha\|_{\vec{\gamma}} = \exists x\, \|\alpha\|_{\vec{\gamma}}$;
- $\|\mathbf{B}(\alpha \Rightarrow \beta)\|_{\vec{\gamma}} = \bigwedge_{p=1}^{q} \big(\big(\bigwedge_{p'=1}^{p-1} \mathrm{RES}[\![\|\neg\alpha\|_{\vec{\gamma}}, \gamma_{p'}]\!] \big) \supset \mathrm{RES}[\![\|(\alpha \supset \beta)\|_{\vec{\gamma}}, \gamma_p]\!] \big)$.

We use $\|\alpha\|_{\mathbf{O}\Gamma}$ as an abbreviation for $\|\alpha\|_{\vec{\gamma}}$ where $\vec{\gamma}$ is the objective representation of **O**Γ from Lemma 4.8.3.

With these definitions, we can eliminate modal operators from formulas. We hence obtain the following result, which generalizes the representation theorem from (Levesque 1984b; Levesque and Lakemeyer 2001) to conditional beliefs.

Theorem 4.8.5 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be objective and* α *be without* **O**. *Then* **O**$\Gamma \models \alpha$ *iff* $\models \|\alpha\|_{\mathbf{O}\Gamma}$.

These results are proved in Appendix B.5, where we consider the more general representation theorem for $\mathcal{ESB}$ that is to be introduced in the next chapter.

Let us illustrate the representation theorem with a simple example; a more elaborate one that involves introspection and quantifying-in is shown in the next chapter.

Example 4.8.6 Let Γ contain the conditionals from the formalization in Example 4.2.2. We show that **O**$\Gamma \models$ **B**$(\neg$Italian $\Rightarrow \neg$Veggie$)$. From Example 4.5.5 we can read off an objective representation $\vec{\gamma}$ of $\mathbf{O}\Gamma$, which can be rewritten equivalently as

$$\gamma_1 = \neg\text{Aussie} \land \text{Italian} \land \mu;$$
$$\gamma_2 = (\text{Aussie} \supset \neg\text{Italian} \land \text{Eats(roo)}) \land (\neg\text{Italian} \supset \text{Aussie}) \land \mu;$$
$$\gamma_3 = \mu,$$

where once again $\mu = \text{Meat(roo)} \land \forall x(\text{Veggie} \land \text{Meat}(x) \supset \neg\text{Eats}(x))$ represents our knowledge about meat and vegetarians.

By Theorem 4.8.5 we need to determine validity of $\|\mathbf{B}(\neg\text{Italian} \Rightarrow \neg\text{Veggie})\|_{\vec{\gamma}}$, which expands to

$$\begin{aligned}\text{RES}[\![\neg\text{Italian} \supset \neg\text{Veggie}, \gamma_1]\!] \;\land \\ (\text{RES}[\![\neg\neg\text{Italian}, \gamma_1]\!] \supset \text{RES}[\![\neg\text{Italian} \supset \neg\text{Veggie}, \gamma_2]\!]) \land \\ (\text{RES}[\![\neg\neg\text{Italian}, \gamma_1]\!] \land \text{RES}[\![\neg\neg\text{Italian}, \gamma_2]\!] \supset \text{RES}[\![\neg\text{Italian} \supset \neg\text{Veggie}, \gamma_3]\!]).\end{aligned}$$

The first line checks for belief at the first sphere, the second line checks the second sphere, and the third line the third and last sphere.

Clearly $\gamma_1 \models (\neg\text{Italian} \supset \neg\text{Veggie})$, so $\text{RES}[\![\neg\text{Italian} \supset \neg\text{Veggie}, \gamma_1]\!]$ is TRUE. Likewise, $\text{RES}[\![\neg\neg\text{Italian}, \gamma_1]\!]$ is TRUE, and $\text{RES}[\![\neg\text{Italian} \supset \neg\text{Veggie}, \gamma_2]\!]$ is TRUE as well since $\gamma_2 \models \neg\text{Italian} \supset \neg\text{Veggie}$; hence the second line is TRUE $\supset$ TRUE. Finally, since $\gamma_2 \not\models \text{Italian}$, $\text{RES}[\![\neg\neg\text{Italian}, \gamma_2]\!]$ is FALSE, so the third line is TRUE $\land$ FALSE $\supset \ldots$ (the consequent is irrelevant as the antecedent is unsatisfiable).

Altogether, by Theorem 4.8.5, we have that $\mathbf{O}\Gamma \models \mathbf{B}(\neg\text{Italian} \Rightarrow \neg\text{Veggie})$ iff TRUE $\land$ (TRUE $\supset$ TRUE) $\land$ (TRUE $\land$ FALSE $\supset \ldots$) is valid, which clearly is the case. That way, we proved $\mathbf{O}\Gamma \models \mathbf{B}(\neg\text{Italian} \Rightarrow \neg\text{Veggie})$ without any non-modal reasoning whatsoever.

4.9 Discussion

In this chapter we presented the logic $\mathcal{BO}$, which extends Levesque's logic of only-knowing $\mathcal{OL}$ to accommodate conditional beliefs. Many typical properties of conditional belief are satisfied in $\mathcal{BO}$ (Theorem 4.4.1). Our definition of conditional belief is also expressive enough to capture indefeasible knowledge. In fact, $\mathcal{OL}$ can be embedded in $\mathcal{BO}$ (Theorem 4.6.4).

The arguably most important result from this chapter is the unique-model property of only-believing for objective conditionals (Theorem 4.5.3). As a consequence, only-believing an objective conditional knowledge base completely determines the agent's beliefs. In fact, only-believing minimizes the agent's beliefs and maximizes the non-beliefs (Theorem 4.6.2) – just like only-knowing does for knowledge. In light of these results it is fair to say: only-believing is to conditional belief what only-knowing is to knowledge. In other words, not only does $\mathcal{BO}$ subsume $\mathcal{OL}$ (Theorem 4.6.4), but it handles conditional belief in the same spirit as $\mathcal{OL}$ handles knowledge.

Only-believing also turns out to be related to System Z (Theorems 4.7.4 and 4.7.7). In a way, $\mathcal{BO}$ incorporates the ideas of System Z in a single logical language, whereas System Z is more of a meta-logical toolbox. $\mathcal{BO}$ is also more general, as its behaviour is also well-defined for inconsistent (in the sense of Definition 4.7.1) conditionals and features first-order logic including quantifying-in and introspection.

Finally we extended Levesque's representation theorem for $\mathcal{OL}$ to conditional beliefs. Besides the unique model of only-believing, our simple semantics of conditional belief made this possible: recall that in an iterated conditional like $\mathbf{B}(\alpha \Rightarrow \mathbf{B}(\beta \Rightarrow \gamma))$ the nested belief $\mathbf{B}(\beta \Rightarrow \gamma)$ is not conditioned on α (Property (xii) in Theorem 4.4.4). This is different from other accounts of conditional belief. Levi (1988) argues on philosophical grounds that $\mathbf{B}(\alpha \Rightarrow \mathbf{B}(\beta \Rightarrow \gamma))$ should be equivalent to $\mathbf{B}(\alpha \wedge \beta \Rightarrow \gamma)$. And Boutilier (1993) defines truth of $\mathbf{B}(\alpha \Rightarrow \beta)$ as truth of $\mathbf{B}\beta$ after revision by α. (We did the same in (Schwering, Lakemeyer, and Pagnucco 2015).) Such semantics would have made the reduction of conditional beliefs to non-modal reasoning (Definition 4.8.4) probably much more complicated or perhaps even impossible, because the reduction works its way from the innermost beliefs to the outside, which in Boutilier's semantics at least would clash with the revisions from the outer beliefs. In the next chapter, our simple semantics will help us a second time, namely with the regression theorem. Not conditioning nested beliefs on an outer belief's antecedent hence seems to be worth to be at odds with other formalisms. (In fact, Levi himself seems not to bother much, as we writes he has "always been mystified why so many serious authors have thought that the problem of iterated conditionals is so important.")

The connections of $\mathcal{BO}$ with $\mathcal{OL}$ and System Z open many interesting follow-up questions. For example, Levesque (1990) and Lakemeyer and Levesque (2005) have shown that only-knowing has close ties to autoepistemic logic (Moore 1985) and Reiter's default logic (Reiter 1980). It would be interesting how these relationships carry over to only-believing. Such an investigation could possibly lead to a logic for reasoning about contingencies in autoepistemic and/or default logic.

Another open problem is to develop a proof theory for $\mathcal{BO}$. Existing proof theories for the propositional fragment of $\mathcal{OL}$ (Levesque and Lakemeyer 2001) and for different accounts of conditional logic (Lewis 1973; Stalnaker 1968) could provide a good starting point. As for the first-order case, the following negative result inherited from $\mathcal{OL}$ sets limits to any proof theory of $\mathcal{BO}$: no sound and complete axiom system of $\mathcal{BO}$ can be recursive. The problem is that any sound and complete proof theory of $\mathcal{BO}$ needs to be able to prove $\mathbf{O}\{\text{TRUE} \Rightarrow \text{TRUE}\} \supset \neg\mathbf{B}\neg\phi$ for objective ϕ iff ϕ is satisfiable; yet the satisfiable formulas are not recursively enumerable in first-order logic, so the proof theory cannot be recursive.

Many further questions arise from the connections originating from System Z. In particular, System Z is related to Pearl's work on probabilistic inference (Pearl 1990, 2014), and it seems plausible to build on that to facilitate probabilistic reasoning in $\mathcal{BO}$. To this end, one might annotate conditionals with probabilities. Belle, Lakemeyer, and Levesque (2016) similarly annotate sentences, but do not allow for conditional probabilities.

An earlier approach to semantically capture System Z is due to Boutilier (1991). Boutilier employs two modal operators to refer to the accessible and the inaccessible worlds in a Kripke structure. In this sense, he draws on the idea of only-knowing, which can be understood as a combination of two modalities: ordinary knowing (at least) and knowing at most. The semantics of Boutilier's logic is quite technical, and Boutilier mentions as future work a "connective analogous to Levesque's **O** operator" in order to combine conditionals with the "semantic clarity of only-knowing." $\mathcal{BO}$ addresses this need with its only-believing operator. Otherwise, it is not easy to see how much $\mathcal{OL}$ and $\mathcal{BO}$ have in common with Boutilier's logic. For example, it is unclear whether a unique-model property comparable to Theorem 4.5.3 holds. A comparison of the logics is left for future work.

This chapter extensively covered the problem of querying a conditional knowledge base, represented as the entailment problem $\mathbf{O}\Gamma \models \mathbf{B}(\alpha \Rightarrow \beta)$. We were not concerned with the question how beliefs change in the face of physical actions or when new information is received and beliefs need be revised appropriately. The interaction of conditional belief with change in the sense of Reiter's situation calculus is the focus of the next chapter.

In terms of Levesque's functional view of knowledge representation (Levesque 1984b; Levesque and Lakemeyer 2001), the present chapter was concerned with *asking*, and the next one deals with *telling*. Defining and studying according operations **TELL** and **ASK** on epistemic states is another perspective of future work.

Reasoning in $\mathcal{BO}$ is of course undecidable because it subsumes the first-order logic $\mathcal{L}$. While the representation theorem eliminates the modal operators, the complexity of first-order logic remains. To alleviate this, we introduce a limited variant of $\mathcal{BO}$ in

Chapter 7.

5 Actions and Belief Revision

In this chapter we investigate how conditional beliefs and classical revision interact with actions. The logic, called $\mathcal{ESB}$ for epistemic situation calculus with beliefs, inherits the conditional belief features from $\mathcal{BO}$ and complements them with situation calculus-style actions. Actions can have two different kinds of effects:

- the physical effect of, say, dropping a box is that fragile items in the box break;
- the epistemic effect of hearing a clink upon dropping the box is that it makes us believe something inside the box broke.

The notion of actions follows the epistemic situation calculus $\mathcal{ES}$ (Lakemeyer and Levesque 2011) presented in Chapter 3. Rather than adopting the approach of classical sensing à la $\mathcal{ES}$, however, we introduce a concept of informing where new information is incorporated by classical revision techniques.

Our particular focus in this chapter is on the *projection problem* for beliefs, which refers to determining what is believed after a sequence of actions occurs. Surprisingly, the belief projection problem in the context of conditional beliefs and/or belief revision remained open for a long time. We present two solutions. The first one is by backward reasoning: formulas are rewritten in order to roll back any actions; this procedure is called *regression*. The alternative is to reason forward: the knowledge base is revised and updated according to actions that occur; this approach is called *progression*.

Besides these main results, we also lift the representation theorem from the previous chapter to $\mathcal{ESB}$. Finally we situate $\mathcal{ESB}$ within the popular belief revision postulate systems and compare informing to sensing in $\mathcal{ES}$.

The presentation of $\mathcal{ESB}$ is based on (Schwering and Lakemeyer 2014, 2015; Schwering, Lakemeyer, and Pagnucco 2015). The results of this chapter require some lengthy proofs, which are given in Appendix B to keep the presentation clear.

5.1 Informing versus sensing

Informing and sensing are two related but different models of how an agent may obtain new information from the outside world. The classical model (Scherl and Levesque 2003) is to let any action sense if a specific formula holds in the real world. Thus sensing actions answer yes–no questions such as "is the gift broken?"; these answers are definitive and cannot be revised. $\mathcal{ES}$ follows this approach, too.

In $\mathcal{ESB}$, we use a more lightweight concept which we call informing. Here, an action simply informs the agent about a suspected fact, possibly without any legitimization from the real world. For example, a clink tells the agent "the gift is broken," but nobody verified this information and it might be actually wrong. This information is inherent to the clink alone, it is independent of what is true in the real world.

Why would we give up the tried and tested concept of sensing? The trouble with sensing is that it is not well-suited for contradictory sensings. Technically, when an action n is performed, the set of possible worlds is thinned out by removing all worlds which disagree with the actual world on the value of $\mathrm{SF}(n)$. For one thing, this means that when two subsequent sensings contradict each other, no possible world is left. Such a logically inconsistent state is highly undesirable as the agent likely is incapable of any reasonable action. For another, even if contradictory information could be handled in a useful way, some error model of would be necessary. Probably it would be represented with two different axioms defining SF, one for the actual world and the other for the possible worlds. Nevertheless it is not clear how these axioms should look like in general.

$\mathcal{ESB}$ addresses the need to handle contradictory information by using (iterated) belief revision. That way implausible information can be displaced. They are not lost once and for all, though, but instead they can be reinstated if new evidence suggests so. The epistemic state of $\mathcal{ESB}$ is hence not just a set of worlds but a system of spheres as in $\mathcal{BO}$. New information is incorporated into this system using classical belief revision schemes, namely natural revision (Boutilier 1993, 1996) or lexicographic revision (Nayak 1994; Nayak, Pagnucco, and Peppas 2003). Natural revision is suitable for less reliable information, lexicographic revision on the other hand leads to strong belief in the new information. For example, the clinking noise is perhaps a weak indicator that something inside the box broke. On the other hand, when we see someone unboxing an object or even do it ourselves, it is quite a convincing that the object actually was in the box, unless perhaps we hallucinate. We hence refer to natural revision as *weak revision* and to lexicographic revision as *strong revision*.

Formally, informing is modelled with a special atom $\mathrm{IF}(n)$ that represents the information carried by the action n. Whenever n is executed, the epistemic state is revised by the information IF(n) according to the revision scheme associated with n. In particular, this means to bring the most-plausible worlds that satisfy IF(n) to the new first sphere, so that afterwards IF(n) is believed.

The second problem with classical sensing mentioned above is to axiomatize the sensor error model. Informing "solves" this problem as a side-effect of its simplicity: there is no error model (in general). (We show in Section 5.11 that informing can mimic sensing à la $\mathcal{ES}$, in which case the problems of classical sensing arise again, of course.) Right or wrong information can freely flow in, without anybody or anything checking its truth or falsity.

Informing has a specific exogenous flavour. In classical sensing, the sensing action is clearly under the agent's control, and its outcome is predetermined by the actual world. By contrast, a clink, for example, is usually not performed by an agent, but rather happens as a consequence of the agent's actions such as dropping the box. Reiter attributes such exogenous actions to "nature" (Reiter 2001). Unwrapping a gift box and finding an object n could be modelled with an action unbox(n). Here, the parameter n intuitively is not under the agent's control but rather an exogenous input: nature fills in which object n was in the box. Bacchus, Halpern, and Levesque (1999) use action parameters in a similar fashion to model noisy sensing results in their probabilistic model.

5.2 The language $\mathcal{ESB}$

The language $\mathcal{ESB}$ is a combination of $\mathcal{ES}$ and $\mathcal{BO}$. The reader will hence recognize much of the following definition.

Definition 5.2.1 The symbols of $\mathcal{ESB}$ are the same as for $\mathcal{ES}$ (Definition 3.9.1) minus **K** plus curly brackets, $\Rightarrow$, and **B**. There shall be a unary special fluent IF (instead of SF). The *terms* are the same as in $\mathcal{ES}$ (Definition 3.9.2), except that action functions come in two subsorts, namely *weak-revision* and *strong-revision* actions. The *formulas* are formed by the same rules as $\mathcal{ES}$ (Definition 3.9.3) with the rule for $\mathbf{K}\alpha$ and $\mathbf{O}\alpha$ replaced with

- $\mathbf{B}(\alpha_1 \Rightarrow \beta_1)$ and $\mathbf{O}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ are formulas if α_i, β_i are formulas.

A formula that mentions no **B** or **O** is called *objective*. A set $\{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ is called objective when all ϕ_i and ψ_i are objective. A formula that mentions function and predicate symbols only within **B** or **O** is called *subjective*. A formula that mentions no $[t]$ or $\square$ operators is called *static*.

Perhaps a few clarifying words about sorts and subsorts are in order. Every function symbol is to be of sort object or action, and every action function is to be of sort weak- or strong-revision. As a consequence, every standard name is of exactly one of the following sorts: object, weak-revision action, or strong-revision action. Variables, however, exist for every sort and subsort; in particular, there are generic action variables, which range over weak-revision and strong-revision actions.

Comparing the syntax of $\mathcal{OL}$ and $\mathcal{ES}$ (Definitions 3.6.1 and 3.9.3) on the one hand and of $\mathcal{BO}$ and $\mathcal{ESB}$ (Definitions 4.2.1 and 5.2.1) on the other shows that $\mathcal{ESB}$ extends $\mathcal{BO}$ with actions the same way as $\mathcal{ES}$ extends $\mathcal{OL}$. The action operators and the conditional belief operators thus have the same meaning in $\mathcal{ESB}$ as in $\mathcal{ES}$ and $\mathcal{BO}$, respectively. We also inherit the abbreviations $\mathbf{B}\alpha$ and $\mathbf{K}\alpha$ from $\mathcal{BO}$.

The newly introduced predicate symbol IF is used to represent the information carried by every action. For example, if t is the action that represents a clink, $\mathrm{IF}(t)$ would be true iff something is broken. The subsort of t indicates whether the information shall be trusted weakly or strongly, which will be semantically reflected in the revision scheme.

We use the same logical abbreviations as for $\mathcal{BO}$. Like with other unary operators, $[t]$ shall bind stronger than any other operators; $\square$, however, shall bind weakest. For example, $\square[a]\mathrm{Broken}(y) \equiv \mathrm{Broken}(y) \lor \mathrm{InBox}(y) \land \mathrm{Fragile}(y) \land a = \mathrm{dropbox}$ abbreviates $\forall a \forall y \square(([a]\mathrm{Broken}(y)) \equiv \big(\mathrm{Broken}(y) \lor ((\mathrm{InBox}(y) \land \mathrm{Fragile}(y)) \land (a = \mathrm{dropbox}))\big))$.

Let us assume that standard names, variables, and function symbols of sort object in $\mathcal{ESB}$ are exactly the standard names, variables, and function symbols of $\mathcal{L}$, and similarly that the rigid predicate symbols in $\mathcal{ESB}$ are the predicate symbols of $\mathcal{L}$. Then clearly every formula of $\mathcal{L}$ or $\mathcal{BO}$ is also a formula of $\mathcal{ESB}$.

5.3 The semantics of $\mathcal{ESB}$

Since $\mathcal{ESB}$ combines the features of $\mathcal{BO}$ and $\mathcal{ES}$, it is no surprise that we can adopt most semantic concepts from them. Action sequences and worlds are defined the same way as in $\mathcal{ES}$ (Definition 3.10.1). An epistemic state is a system of spheres as in $\mathcal{BO}$ (Definition 4.3.1), except that the worlds are worlds in the sense of $\mathcal{ES}$ (Definition 3.10.1). The denotation of a term is also defined as in $\mathcal{ES}$ (Definition 3.10.2).

We define the semantics of $\mathcal{ESB}$ without the additional z parameter for the action sequence used in $\mathcal{ES}$. Instead of keeping the history of executed actions in the model, we simply progress worlds immediately whenever an action occurs.

Definition 5.3.1 The *progression* of a world w by an action standard name n is a world $w \gg n$ such that

- $(w \gg n)[g(n_1, \ldots, n_k)] = w[g(n_1, \ldots, n_k)]$ for all object function symbols g;
- $(w \gg n)[R(n_1, \ldots, n_k)] = w[R(n_1, \ldots, n_k)]$ for all rigid predicate symbols R;
- $(w \gg n)[F(n_1, \ldots, n_k), z] = w[F(n_1, \ldots, n_k), n \cdot z]$ for all fluent predicate symbols F and action sequences z.

We abbreviate $w \gg n_1 \gg \ldots \gg n_k$ by $w \gg \langle n_1, \ldots, n_k \rangle$.

It is immediate that $w \gg n$ is uniquely defined. Having that, we can define the semantics of objective sentences.

Definition 5.3.2 The truth relation $\models$ of $\mathcal{ESB}$ for objective sentences is defined with respect to a world w:

$\mathcal{ESB}$1'. $w \models R(t_1, \ldots, t_k)$ iff
R is rigid and $w[R(n_1, \ldots, n_k)] = 1$ where $n_i = w(t_i)$;

$\mathcal{ESB}$2'. $w \models F(t_1, \ldots, t_k)$ iff
F is fluent and $w[F(n_1, \ldots, n_k), \langle\rangle] = 1$ where $n_i = w(t_i)$;

$\mathcal{ESB}$3'. $w \models (t_1 = t_2)$ iff n_1 and n_2 are identical names where $n_i = w(t_i)$;

$\mathcal{ESB}$4'. $w \models \neg\alpha$ iff $w \not\models \alpha$;

$\mathcal{ESB}$5'. $w \models (\alpha \vee \beta)$ iff $w \models \alpha$ or $w \models \beta$;

$\mathcal{ESB}$6'. $w \models \exists x\, \alpha$ iff $w \models \alpha^x_n$ for some name n of the same sort as x;

$\mathcal{ESB}$7'. $w \models [t]\alpha$ iff $w \gg n \models \alpha$ where $n = w(t)$;

$\mathcal{ESB}$8'. $w \models \Box\alpha$ iff $w \gg z \models \alpha$ for every action sequence z.

Only Rules $\mathcal{ESB}$2', $\mathcal{ESB}$7', and $\mathcal{ESB}$8' differ from the corresponding rules for $\mathcal{ES}$ in Definition 3.10.4. The difference is because the action sequence is no part of the model in $\mathcal{ESB}$; instead, worlds are progressed immediately when an action is executed. Both approaches are equivalent, but ours is easier to work with as we shall see when we give semantics to the belief modalities.

Progressing an epistemic state by an action is more complicated, because the information content of the action needs to be considered. That is, we need to first revise the epistemic state before we progress its individual worlds. The following definition introduces the necessary tools.

Definition 5.3.3 For an epistemic state $\vec{e}$ and an objective sentence ϕ, we define

- $\lfloor \vec{e} \rfloor = \min\{p \mid p = \infty \text{ or } e_p \neq \{\}\}$;
- $\lceil \vec{e} \rceil = \max\{p \mid p = 1 \text{ or } e_{p-1} \neq e_p\}$.
- $\vec{e} \mid \phi = \langle e_1 \cap W, \ldots, e_{\lceil \vec{e} \rceil} \cap W \rangle$ where $W = \{w \mid w \models \phi\}$ for objective ϕ.

Intuitively, $\lfloor \vec{e} \rfloor$ denotes the first non-empty level of $\vec{e}$, and $\lceil \vec{e} \rceil$ refers to the last distinct level. Notice that $\lfloor \vec{e} \rfloor = \infty$ when the epistemic state is empty, whereas always $\lceil \vec{e} \rceil \in \mathbb{P}$ since $\vec{e}$ converges. We will always make explicit when $\lfloor \vec{e} \rfloor$ can take the value ∞. Restricting $\vec{e}$ to only ϕ-worlds by writing $\vec{e} \mid \phi$ is useful in combination with $\lfloor \vec{e} \rfloor$ to capture the plausibility of ϕ in $\vec{e}$: $\lfloor \vec{e} \mid \phi \rfloor$. For now, this expression is only defined for objective sentences ϕ, for we have not yet introduced the semantics of beliefs; but in a few moments we can generalize $\vec{e} \mid \phi$ to arbitrary formulas. The reader may have noticed that we introduced the same notation $\lfloor \vec{e} \mid \alpha \rfloor$ already in Definition 4.3.2 for $\mathcal{BO}$; once we generalize $\vec{e} \mid \phi$ to arbitrary formulas, the old definition of $\lfloor \vec{e} \mid \alpha \rfloor$ from $\mathcal{BO}$ coincides with the new one.

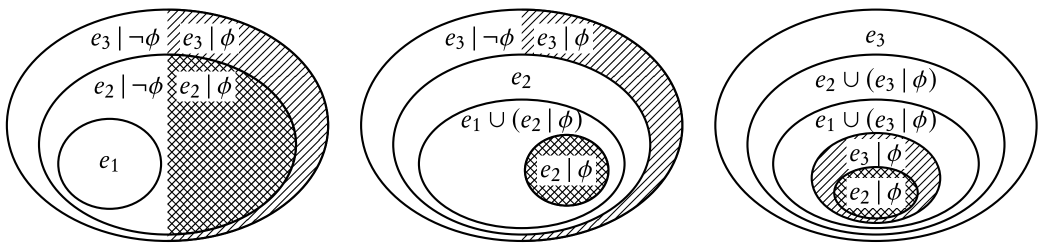
We can now define what it means to revise an epistemic state. Weak revision moves the most-plausible worlds that satisfy the new information to the front. Strong revision, by contrast, promotes *all* worlds that satisfy the new information over all other ones. An illustration is depicted in Figure 5.1. The first sphere is thus the same after a single revision, but the subsequent ones differ. Intuitively, after strong revision the agent is more reluctant to give up belief in the new information again. Strong revision hence leads to stronger belief in the new information than weak revision.

Definition 5.3.4 The *weak revision* $\vec{e} *_{\mathrm{w}} \phi$ of $\vec{e}$ by an objective sentence ϕ is defined as follows:

- if $\lfloor \vec{e} \mid \phi \rfloor = \infty$: $\quad \vec{e} *_{\mathrm{w}} \phi = \langle \{\} \rangle$;
- if $\lfloor \vec{e} \mid \phi \rfloor \neq \infty$: $\quad \vec{e} *_{\mathrm{w}} \phi = \langle W, e_1 \cup W, \ldots, e_{\lceil \vec{e} \rceil} \cup W \rangle$ where $W = (\vec{e} \mid \phi)_{\lfloor \vec{e} \mid \phi \rfloor}$.

The *strong revision* $\vec{e} *_{\mathrm{s}} \phi$ of $\vec{e}$ by an objective sentence ϕ is defined as follows:

- if $\lfloor \vec{e} \mid \phi \rfloor = \infty$: $\quad \vec{e} *_{\mathrm{s}} \phi = \langle \{\} \rangle$;
- if $\lfloor \vec{e} \mid \phi \rfloor \neq \infty$: $\quad \vec{e} *_{\mathrm{s}} \phi = \langle (\vec{e} \mid \phi)_{\lfloor \vec{e} \mid \phi \rfloor}, \ldots, (\vec{e} \mid \phi)_{\lceil \vec{e} \rceil},$
  $(\vec{e} \mid \neg\phi)_{\lfloor \vec{e} \mid \neg\phi \rfloor} \cup W, \ldots, (\vec{e} \mid \neg\phi)_{\lceil \vec{e} \rceil} \cup W \rangle$
  where $W = (\vec{e} \mid \phi)_{\lceil \vec{e} \rceil}$.



(a) An epistemic state $\vec{e}$. (b) Weak revision $\vec{e} *_{\mathrm{w}} \phi$. (c) Strong revision $\vec{e} *_{\mathrm{s}} \phi$.

Figure 5.1: The original epistemic state $\vec{e}$ in (a) has three different spheres $\langle e_1, e_2, e_3\rangle$. Hatched area indicates ϕ-worlds; the most-plausible ones occur in e_2, and e_3 contains additional ones. Weak (natural) revision by ϕ promotes these most-plausible ϕ-worlds, namely $e_2 \mid \phi$, to the first sphere, but leaves the ranking otherwise unchanged, as depicted in (b). Strong (lexicographic) revision promotes all ϕ-worlds, namely $e_2 \mid \phi$ and $e_3 \mid \phi$, over all $\neg\phi$-worlds, but preserves the relative ordering of the ϕ-worlds and $\neg\phi$-worlds, respectively, as depicted (c).

When the revision mechanism is clear from context or irrelevant, we just write $\vec{e} * \phi$. In particular, we mean by $\vec{e} * \mathrm{IF}(n)$ the revision according to the sort of n, that is, $\vec{e} *_{\mathrm{w}} \mathrm{IF}(n)$ if n is a weak-revision action and $\vec{e} *_{\mathrm{s}} \mathrm{IF}(n)$ if n is a strong-revision action. Now it is easy to define the progression of an epistemic state.

Definition 5.3.5 The *progression* of a set of worlds W and of an epistemic state $\vec{e}$ are defined as

- $W \gg n = \{w \gg n \mid w \in W\}$;
- $\vec{e} \gg n = \langle e_1' \gg n, \ldots, e_q' \gg n\rangle$ where $\langle e_1', \ldots, e_q'\rangle = \vec{e} * \mathrm{IF}(n)$.

We abbreviate $\vec{e} \gg n_1 \gg \ldots \gg n_k$ by $\vec{e} \gg \langle n_1, \ldots, n_k\rangle$.

The following lemma says that the revision and the progression of an epistemic state are well-behaved.

Lemma 5.3.6 *$\vec{e} *_{\mathrm{w}} \phi$, $\vec{e} *_{\mathrm{s}} \phi$, and $\vec{e} \gg n$ are epistemic states.*

Proof. Let $\vec{e} = \langle e_1, \ldots, e_q\rangle$ be an epistemic state. Then $e_p \subseteq e_{p+1}$ for all $p \in \mathbb{P}$, and $e_q = e_p$ for all $p \geq q$.

Hence $(\vec{e} *_{\mathrm{w}} \phi)_p \subseteq (\vec{e} *_{\mathrm{w}} \phi)_{p+1}$ for all $p \in \mathbb{P}$, and $(\vec{e} *_{\mathrm{w}} \phi)_{q+1} = (\vec{e} *_{\mathrm{w}} \phi)_p$ for all $p \geq q+1$, so $\vec{e} *_{\mathrm{w}} \phi$ satisfies the concentricity and convergence constraints, and thus is an epistemic state. Likewise, $(\vec{e} *_{\mathrm{s}} \phi)_p \subseteq (\vec{e} *_{\mathrm{s}} \phi)_{p+1}$ for all $p \in \mathbb{P}$, and $(\vec{e} *_{\mathrm{s}} \phi)_{2 \cdot q} = (\vec{e} *_{\mathrm{s}} \phi)_p$ for all $p \geq 2 \cdot q$, so $\vec{e} *_{\mathrm{s}} \phi$ is an epistemic state, too.

Finally consider $\vec{e} \gg n$, which simply progresses the individual worlds in $\vec{e} * \mathrm{IF}(n)$. It is immediate from Definition 5.3.1 that the progression $w \gg n$ of a world w again is a world. Thus $W \gg n$ is a set of worlds if W is one, and if $W \subseteq W'$, then $W \gg n \subseteq W' \gg n$. Therefore, since $\vec{e} * \mathrm{IF}(n)$ is an epistemic state, $\vec{e} \gg n$ is one, too. □

As mentioned above, weak and strong revision of the same epistemic state lead to the same factual beliefs (but typically differ in the counterfactual beliefs).

Lemma 5.3.7 *$(\vec{e} *_{\mathrm{w}} \phi)_1 = (\vec{e} *_{\mathrm{s}} \phi)_1$ and $\lfloor \vec{e} * \phi \,|\, \phi \rfloor = 1$, or $(\vec{e} *_{\mathrm{w}} \phi) = (\vec{e} *_{\mathrm{s}} \phi)$.*

Proof. If $\lfloor \vec{e} \,|\, \phi \rfloor = \infty$, $\vec{e} *_{\mathrm{w}} \phi = \langle \{\} \rangle = \vec{e} *_{\mathrm{s}} \phi$. Otherwise, $(\vec{e} *_{\mathrm{w}} \phi)_1 = (\vec{e} \,|\, \phi)_{\lfloor \vec{e} \,|\, \phi \rfloor} = (\vec{e} *_{\mathrm{s}} \phi)_1 \neq \{\}$ and $\lfloor \vec{e} * \phi \,|\, \phi \rfloor = 1$. □

With these definitions in hand, we can give semantics to the full language.

Definition 5.3.8 The truth relation $\models$ of $\mathcal{ESB}$ is defined with respect to an epistemic state $\vec{e}$ and a world w:

$\mathcal{ESB}1$. $\vec{e}, w \models R(t_1, \ldots, t_k)$ iff
R is rigid and $w[R(n_1, \ldots, n_k)] = 1$ where $n_i = w(t_i)$;

$\mathcal{ESB}2$. $\vec{e}, w \models F(t_1, \ldots, t_k)$ iff
F is fluent and $w[F(n_1, \ldots, n_k), \langle\rangle] = 1$ where $n_i = w(t_i)$;

$\mathcal{ESB}3$. $\vec{e}, w \models (t_1 = t_2)$ iff n_1 and n_2 are identical names where $n_i = w(t_i)$;

$\mathcal{ESB}4$. $\vec{e}, w \models \neg\alpha$ iff $\vec{e}, w \not\models \alpha$;

$\mathcal{ESB}5$. $\vec{e}, w \models (\alpha \vee \beta)$ iff $\vec{e}, w \models \alpha$ or $\vec{e}, w \models \beta$;

$\mathcal{ESB}6$. $\vec{e}, w \models \exists x \, \alpha$ iff $\vec{e}, w \models \alpha^x_n$ for some name n of the same sort as x;

$\mathcal{ESB}7$. $\vec{e}, w \models [t]\alpha$ iff $\vec{e} \gg n, w \gg n \models \alpha$ where $n = w(t)$;

$\mathcal{ESB}8$. $\vec{e}, w \models \Box\alpha$ iff $\vec{e} \gg z, w \gg z \models \alpha$ for every action sequence z;

$\mathcal{ESB}9$. $\vec{e}, w \models \mathbf{B}(\alpha \Rightarrow \beta)$ iff
for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \,|\, \alpha \rfloor$ and $w' \in e_p$, then $\vec{e}, w' \models (\alpha \supset \beta)$;

$\mathcal{ESB}10$. $\vec{e}, w \models \mathbf{O}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ iff
for all $p \in \mathbb{P}$, $w' \in e_p$ iff $\vec{e}, w' \models \bigwedge_{i : \lfloor \vec{e} \,|\, \alpha_i \rfloor \geq p} (\alpha_i \supset \beta_i)$;

where $\vec{e} \,|\, \alpha = \langle e_1 \cap W, \ldots, e_{\lceil \vec{e} \rceil} \cap W \rangle$ for $W = \{w \mid \vec{e}, w \models \alpha\}$ generalizes $\vec{e} \,|\, \phi$ to arbitrary α.

Rules $\mathcal{ESB}$2–$\mathcal{ESB}$8 are just the ones from the objective semantics (Definition 5.3.2) retrofitted with an additional epistemic state $\vec{e}$, which in case of $[t]$ and $\Box$ needs to be progressed on the right-hand side. It is hence immediate that the Definitions 5.3.2 and 5.3.8 agree on the truth of objective sentences. The new Rules $\mathcal{ESB}$9 and $\mathcal{ESB}$10 are the same as for $\mathcal{BO}$ (Definition 4.3.2) and express the same intuition.

Theorem 5.3.9 *Let α be a sentence of $\mathcal{BO}$. Then $\models_{\mathcal{BO}} \alpha$ iff $\models \alpha$.*

The proof is surprisingly tedious; we give it in Appendix B.1.

Lemma 5.3.10 $\models \Box\alpha$ *iff* $\models \alpha$.

Proof. For the only-if direction suppose $\models \Box\alpha$. Then by Rule $\mathcal{ESB}$8, $\vec{e} \gg z, w \gg z \models \alpha$ for all $\vec{e}$, w, and z. In particular, this holds for $z = \langle\rangle$, and since $\vec{e} \gg \langle\rangle = \vec{e}$ and $w \gg \langle\rangle = w$, we have $\vec{e}, w \models \alpha$ for all $\vec{e}$, w, so $\models \alpha$. Conversely, suppose $\models \alpha$. Therefore and by Lemma 5.3.6, for all $\vec{e}$, w, z, we have $\vec{e} \gg z, w \gg z \models \alpha$, and by Rule $\mathcal{ESB}$8 $\vec{e}, w \models \Box\alpha$. Thus $\models \Box\alpha$. □

The proofs of Theorems 4.4.5, 4.4.1, 4.4.2, 4.4.3, 4.4.4, and 4.5.3 all carry over to $\mathcal{ESB}$ without any modification. By Lemma 5.3.10, Theorems 4.4.5, 4.4.4 and 4.5.3 also hold after any sequence of actions. We therefore have the counterparts to the mentioned $\mathcal{BO}$ theorems in $\mathcal{ESB}$.

Theorem 5.3.11 $\models \Box\mathbf{O}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\} \supset \bigwedge_i \mathbf{B}(\alpha_i \Rightarrow \beta_i)$.

Theorem 5.3.12 $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \beta)$ *iff* $\lfloor \vec{e} \mid \alpha \rfloor = \infty$ *or* $\vec{e}, w \models (\alpha \supset \beta)$ *for all* $w \in e_{\lfloor \vec{e} \mid \alpha \rfloor}$.

Theorem 5.3.13 $\vec{e} \models \mathbf{K}\alpha$ *iff* $\vec{e}, w \models \alpha$ *for all* $w \in e_p$ *and* $p \in \mathbb{P}$.

Theorem 5.3.14 $\vec{e} \models \mathbf{B}(\alpha \vee \beta \Rightarrow \neg\beta)$ *iff* $\lfloor \vec{e} \mid \alpha \rfloor < \lfloor \vec{e} \mid \beta \rfloor$ *or* $\lfloor \vec{e} \mid \alpha \rfloor = \lfloor \vec{e} \mid \beta \rfloor = \infty$.

Theorem 5.3.15

(i) $\not\models \Box\mathbf{B}(\alpha \Rightarrow \beta) \wedge \mathbf{B}(\beta \Rightarrow \gamma) \supset \mathbf{B}(\alpha \Rightarrow \gamma)$;

(ii) $\not\models \Box\mathbf{B}(\alpha \Rightarrow \gamma) \supset \mathbf{B}(\alpha \wedge \beta \Rightarrow \gamma)$;

(iii) $\not\models \Box\mathbf{B}(\alpha \Rightarrow \beta) \equiv \mathbf{B}(\neg\beta \Rightarrow \neg\alpha)$;

(iv) $\models \Box\mathbf{B}\alpha \wedge \mathbf{B}(\alpha \supset \beta) \supset \mathbf{B}\beta$;

(v) $\models \Box\mathbf{K}\alpha \wedge \mathbf{K}(\alpha \supset \beta) \supset \mathbf{K}\beta$;

(vi) $\models \Box\mathbf{B}\alpha \wedge \mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{B}\beta$;

(vii) $\models \Box\mathbf{K}\alpha \supset \mathbf{B}\alpha$;

(viii) $\models \Box\mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{K}\mathbf{B}(\alpha \Rightarrow \beta)$;

(ix) $\models \Box\neg\mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{K}\neg\mathbf{B}(\alpha \Rightarrow \beta)$;

(x) $\models \Box\forall x\, \mathbf{B}(\alpha \Rightarrow \beta) \supset \mathbf{B}(\alpha \Rightarrow \forall x\, \beta)$ *where x does not occur freely in α;*

(xi) $\models \Box\mathbf{K}\alpha$ *if* $\models \alpha$;

(xii) $\models \Box\mathbf{B}(\alpha \Rightarrow \mathbf{B}(\beta \Rightarrow \gamma)) \wedge \neg\mathbf{K}\neg\alpha \supset \mathbf{B}(\beta \Rightarrow \gamma)$.

Theorem 5.3.16 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \dots, \phi_m \Rightarrow \psi_m\}$ be objective. Then there is a unique $\vec{e} = \langle e_1, \dots, e_{m+1}\rangle$ such that $\vec{e} \models \mathbf{O}\Gamma$.*

5.4 The belief projection problem

The belief projection problem is to decide if a specific belief holds true after a sequence of actions. Logically this is expressed as an entailment problem: given a knowledge base about the domain's dynamics and the agent's (conditional) beliefs, does a certain formula about actions and beliefs follow? Following (Lakemeyer and Levesque 2011; Reiter 2001), we consider knowledge bases of the following form in this paper.

Definition 5.4.1 Let $\mathcal{F}$ be a finite set of fluent predicate symbols and IF $\notin \mathcal{F}$. A formula is *fluent* when it is objective, static, and all fluent predicate symbols are from $\mathcal{F}$. A *basic action theory* over $\mathcal{F}$ consists of two sets Σ_{dyn} and Σ_{bel}, where

- Σ_{dyn} contains dynamic axioms, namely
  - a sentence $\Box[a]F(x_1, \dots, x_k) \equiv \gamma_F$ for every fluent predicate symbol $F \in \mathcal{F}$ where γ_F is a fluent formula;
  - a single sentence $\Box\text{IF}(a) \equiv \varphi$ where φ is a fluent formula;[1]
- Σ_{bel} contains finitely many conditionals $\phi \Rightarrow \psi$ where ϕ and ψ are fluent sentences.

We identify Σ_{dyn} with the conjunction of its elements and let $\mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$ stand for $\mathbf{O}(\{\neg\Sigma_{\text{dyn}} \Rightarrow \text{FALSE}\} \cup \Sigma_{\text{bel}})$. Then the *belief projection problem* is to decide entailments of the form

$$\mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models [t_1] \dots [t_k]\mathbf{B}(\alpha \Rightarrow \beta).$$

The idea of basic action theories dates back to Reiter (1991, 2001). Sentences of the form $\Box[a]F(x_1, \dots, x_k) \equiv \gamma_F$ are called *successor-state axioms* because they relate the state after an action a to the one before a. They incorporate Reiter's solution to the

[1] Usually, basic action theories also feature a precondition axiom of the form $\text{Poss}(a) \equiv \pi$ for a fluent formula π. It is generally treated very similar to the $\text{IF}(a) \equiv \varphi$ axiom. For simplicity, we omit it here. Also see Section 5.11.

frame problem (Reiter 2001). The *informed-fluent axiom* $\Box \mathrm{IF}(a) \equiv \varphi$ axiomatizes the information an action a tells the agent. The conditional $\neg\Sigma_{\mathrm{dyn}} \Rightarrow \textsc{false}$ in $\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}})$ expresses that the agent knows these dynamic axioms.

There are two principal answers to (belief) projection problems.

Query regression reasons backwards: it rewrites the query to undo the actions and thus reduces reasoning about future situations to reasoning about the initial situation (Reiter 1991, 2001). The goal is hence to rewrite the query $[t_1] \ldots [t_k]\mathbf{B}(\alpha \Rightarrow \beta)$ to a new, static query which is equivalent (modulo $\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}})$) to the original query.

Successor-state axioms are fundamental to roll back an action: they deterministically relate every fluent's value before and after an action. For example, a fluent after an action such as $[n]F(n')$ can be replaced with the right-hand side of F's successor-state axiom $\gamma_F{}^{x}_{n'}{}^{a}_{n}$. This process can be iterated until the initial situation is reached.

As we shall see, this procedure not only works for objective formulas, but also for conditional beliefs. In fact, two theorems will play a role analogous to successor-state axioms: they relate $[n]\mathbf{B}(\alpha \Rightarrow \psi)$ to a belief before n.

Regression is a very elegant mechanism to eliminate actions from the reasoning task. On the downside, the regressed query may grow exponentially in the number of actions. The procedure is hence not suited for long-lived systems that amass a huge number of actions.

Knowledge base progression reasons forward: it applies the action to the knowledge base and thus produces a new "snapshot" of the world (Lin and Reiter 1997; Reiter 2001). That is, we need to determine some updated beliefs Σ'_{bel} such that $\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}})$ entails $[n]\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma'_{\mathrm{bel}})$.

A general way to do so is to rename all fluent symbols in the knowledge based and initialize the original symbols to the correct value after an action using the successor-state axioms. For example, if we have only a single fluent F and a knowledge base $F(n_1) \vee F(n_2)$, then after doing n the new knowledge base can be represented as $(R(n_1) \vee R(n_2)) \wedge \forall x(F(x) \equiv \gamma_F{}^{x}_{n}{}^{F}_{R})$.

Unfortunately, progression is highly complex; it is not first-order definable (Lin and Reiter 1997; Vassos and Levesque 2013). The problem is that in general it is inevitable to introduce new predicate symbols, like we did with R above. These new predicates need to be forgotten again, which can be achieved using second-order logic (Lin and Reiter 1994, 1997): R shall be an existentially second-order variable.

We will see that the sketched general approach to progression can be generalized to conditional knowledge bases using a notion of only-believing extended with means to forget predicates.

Before we study regression and progression in the upcoming sections, let us formalize Example 1.1.1 as a basic action theory and investigate a few example queries.

Example 5.4.2 The scenario comprises a single box that may contain items, which we represent by a fluent predicate $\mathrm{InBox}(n)$. Items can be taken out of the box by action $\mathrm{unbox}(n)$, and the box can be dropped by action dropbox. Dropping the box breaks all fragile items in it, which is formalized using a rigid predicate $\mathrm{Fragile}(n)$ and another fluent predicate $\mathrm{Broken}(n)$. A clinking noise, represented by the action clink, indicates that something in the box seems to be broken: $\exists y(\mathrm{InBox}(y) \land \mathrm{Broken}(y))$. Intuitively clink is exogenous, that is, it is not under the agent's control but she observes (nature executing) a clink. Unboxing an item n through action $\mathrm{unbox}(n)$ tells us that this item was in the box and is not broken: $\mathrm{InBox}(n) \land \neg\mathrm{Broken}(n)$. The successor-state axioms for InBox and Broken and the informed-fluent axiom constitute the dynamic axioms

$$\begin{aligned}\Sigma_{\mathrm{dyn}} = \{&\Box[a]\mathrm{InBox}(y) \equiv \mathrm{InBox}(y) \land a \neq \mathrm{unbox}(y),\\ &\Box[a]\mathrm{Broken}(y) \equiv \mathrm{Broken}(y) \lor \mathrm{InBox}(y) \land \mathrm{Fragile}(y) \land a = \mathrm{dropbox},\\ &\Box\mathrm{IF}(a) \equiv (a = \mathrm{clink} \supset \exists y(\mathrm{InBox}(y) \land \mathrm{Broken}(y))) \land\\ &\qquad\qquad \forall y(a = \mathrm{unbox}(y) \supset \mathrm{InBox}(y) \land \neg\mathrm{Broken}(y))\}.\end{aligned}$$

We still need to decide of which revision sort the actions are. Since a clinking noise is a rather unreliable hint that something is broken, we make clink a weak-revision action. By contrast, when one takes an object out of the box, that object must indeed have been in the box and be in one piece (otherwise one probably hallucinates), so $\mathrm{unbox}(n)$ shall be a strong-revision action. We let dropbox be a strong-revision action, too; since IF(dropbox) is vacuously true the revision has no effect anyway.

Our agent believes that most likely the box is empty; but taking the possibility into account that she may be wrong about that, she believes that in this case only the gift would be in the box. We use the object constant gift to refer to the gift that may or may not be in the box. Note that it is not a standard name, so the agent might have no clue what the gift actually is. She moreover believes that if there was something in the box, it would not be broken yet. Thus we define the initial beliefs as

$$\begin{aligned}\Sigma_{\text{bel}} = \{&\textsc{true} \Rightarrow \forall y\, \neg \text{InBox}(y),\\ &\exists y\, \text{InBox}(y) \Rightarrow \forall y\, (\text{InBox}(y) \equiv y = \text{gift}),\\ &\exists y\, \text{InBox}(y) \Rightarrow \forall y\, (\text{InBox}(y) \supset \neg \text{Broken}(y))\}.\end{aligned}$$

This completes the basic action theory. As we shall see, $\mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$ entails the following queries.

Q1. Initially the agent believes the box is empty: $\mathbf{B}\forall y\, \neg \text{InBox}(y)$.

Q2. After dropping the box, she still believes the box is empty, but also that if something fragile was in the box, it would be broken:
$[\text{dropbox}](\mathbf{B}(\forall y\, \neg \text{InBox}(y)) \wedge \forall y\, \mathbf{B}(\text{InBox}(y) \wedge \text{Fragile}(y) \Rightarrow \text{Broken}(y)))$.

Q3. When a clink occurs after dropping the box, she comes to believe that the gift was in the box, but she has no idea what the gift is:
$[\text{dropbox}][\text{clink}]\mathbf{B}(\text{InBox}(\text{gift}) \wedge \text{Broken}(\text{gift}) \wedge \neg \exists y\, \mathbf{B}\text{gift} = y)$.

Q4. When the object $\#5$ is taken out of the box, she believes that this must be the gift, and that it is not broken after all:
$[\text{dropbox}][\text{clink}][\text{unbox}(\#5)]\exists y\, \mathbf{B}(\text{gift} = y \wedge \neg \text{InBox}(\text{gift}) \wedge \neg \text{Broken}(\text{gift}))$.

We use the latter two queries to illustrate the results of this chapter: regression, progression, and representation theorem. To begin with, we prove these queries semantically.

Example 5.4.3 First we need to determine the epistemic state $\vec{e} \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$. By Theorem 5.3.16, $\vec{e}$ is unique, and using the idea from the proof of Lemma 4.5.2 we generate $\vec{e} = \langle e_1, e_2, e_3 \rangle$. The first level e_1 contains all worlds that satisfy Σ_{dyn} and the material-implication-versions of all conditionals in Σ_{bel}, which simplifies to

$$e_1 = \{w \mid w \models \Sigma_{\text{dyn}} \wedge \forall y\, \neg \text{InBox}(y)\}.$$

Thus $\lfloor \vec{e} \mid \textsc{true} \rfloor = 1$ and $\lfloor \vec{e} \mid \exists y\, \text{InBox}(y) \rfloor > 1$, so the next level e_2 contains all worlds that satisfy Σ_{dyn} and $\exists y\, \text{InBox}(y) \supset \forall y\, (\text{InBox}(y) \equiv y = \text{gift})$ as well as $\exists y\, \text{InBox}(y) \supset \forall y\, (\text{InBox}(y) \supset \neg \text{Broken}(y))$, which simplifies to

$$e_2 = \{w \mid w \models \Sigma_{\text{dyn}} \wedge \forall y\, (\text{InBox}(y) \supset y = \text{gift} \wedge \neg \text{Broken}(y))\}.$$

Hence $\lfloor \vec{e} \mid \exists y \, \mathrm{InBox}(y) \rfloor = 2$, so all following levels contain all worlds that satisfy Σ_{dyn}, that is,

$$e_3 = \{w \mid w \models \Sigma_{\mathrm{dyn}}\}.$$

Example 5.4.4 Now we can prove the above queries.

Q1. $\mathbf{B}\forall y \neg \mathrm{InBox}(y)$.

Clearly, $e_1 \neq \{\}$, so $\lfloor \vec{e} \mid \textsc{true} \rfloor = 1$. Moreover $w \models \forall y \neg \mathrm{InBox}(y)$ for all $w \in e_1$. By Theorem 5.3.12, the query holds.

Q2. $[\mathrm{dropbox}](\mathbf{B}(\forall y \neg \mathrm{InBox}(y)) \wedge \forall y \, \mathbf{B}(\mathrm{InBox}(y) \wedge \mathrm{Fragile}(y) \Rightarrow \mathrm{Broken}(y)))$.

We progress $\vec{e}$ to evaluate this query. The action dropbox makes each Broken(n) true when Fragile(n) and InBox(n) are true. Since dropbox makes a (strong) revision by the vacuously true IF(dropbox), there effectively is no revision. The progressed state $\vec{e} \gg \mathrm{dropbox}$ is thus

$$(\vec{e} \gg \mathrm{dropbox})_1 = \{w \mid w \models \Sigma_{\mathrm{dyn}} \wedge \forall y \neg \mathrm{InBox}(y)\};$$
$$(\vec{e} \gg \mathrm{dropbox})_2 = \{w \mid w \models \Sigma_{\mathrm{dyn}} \wedge \forall y (\mathrm{InBox}(y) \supset y = \mathrm{gift} \wedge (\mathrm{Broken}(y) \equiv \mathrm{Fragile}(y)))\};$$
$$(\vec{e} \gg \mathrm{dropbox})_3 = \{w \mid w \models \Sigma_{\mathrm{dyn}} \wedge \forall y (\mathrm{InBox}(y) \wedge \mathrm{Fragile}(y) \supset \mathrm{Broken}(y))\}.$$

By the same argument as for Q1, $[\mathrm{dropbox}]\mathbf{B}\forall y \neg \mathrm{InBox}(y)$ is true. And for all n, there is some $w \in (\vec{e} \gg \mathrm{dropbox})_2$ such that $w \models \mathrm{InBox}(n) \wedge \mathrm{Fragile}(n)$, and then also $w \models \mathrm{Broken}(n)$, so $[\mathrm{dropbox}]\forall y \, \mathbf{B}(\mathrm{InBox}(y) \wedge \mathrm{Fragile}(y) \Rightarrow \mathrm{Broken}(y))$ holds as well.

Q3. $[\mathrm{dropbox}][\mathrm{clink}]\mathbf{B}(\mathrm{InBox}(\mathrm{gift}) \wedge \mathrm{Broken}(\mathrm{gift}) \wedge \neg \exists y \, \mathbf{B}\mathrm{gift} = y)$.

The action clink does not change the truth value of any fluents, but it triggers a weak revision by $\exists y (\mathrm{InBox}(y) \wedge \mathrm{Broken}(y))$, that is, the most-plausible worlds from $\vec{e} \gg \mathrm{dropbox}$ satisfying this formula constitute the first plausibility level in the revised state $(\vec{e} \gg \mathrm{dropbox}) * \mathrm{clink}$. Thus $(\vec{e} \gg \mathrm{dropbox}) * \mathrm{clink}$ can be written as

$$((\vec{e} \gg \mathrm{dropbox}) * \mathrm{clink})_1 = \{w \mid w \models \Sigma_{\mathrm{dyn}} \wedge \forall y (\mathrm{InBox}(y) \equiv y = \mathrm{gift}) \wedge \mathrm{Broken}(\mathrm{gift}) \wedge \mathrm{Fragile}(\mathrm{gift})\};$$
$$((\vec{e} \gg \mathrm{dropbox}) * \mathrm{clink})_2 = \{w \mid w \models \Sigma_{\mathrm{dyn}} \wedge \forall y (\mathrm{InBox}(y) \supset y = \mathrm{gift} \wedge \mathrm{Broken}(\mathrm{gift}) \wedge \mathrm{Fragile}(\mathrm{gift}))\};$$
$$((\vec{e} \gg \mathrm{dropbox}) * \mathrm{clink})_3 = \{w \mid w \models \Sigma_{\mathrm{dyn}} \wedge \forall y (\mathrm{InBox}(y) \supset y = \mathrm{gift} \wedge (\mathrm{Broken}(y) \equiv \mathrm{Fragile}(y)))\};$$

$$((\vec{e} \gg \text{dropbox}) * \text{clink})_4 = \{w \mid w \models \Sigma_{\text{dyn}} \wedge \forall y\,(\text{InBox}(y) \wedge \text{Fragile}(y) \supset \text{Broken}(y))\}.$$

Since clink has no physical effect, $(\vec{e} \gg \text{dropbox}) \gg \text{clink} = (\vec{e} \gg \text{dropbox}) * \text{clink}$. So $w \models \text{InBox(gift)} \wedge \text{Broken(gift)}$ for all $w \in (\vec{e} \gg \text{dropbox} \gg \text{clink})_1$. Moreover, the worlds do not agree on the denotation of gift, so there is no standard name n such that $w \models (\text{gift} = n)$ for all $w \in (\vec{e} \gg \text{dropbox} \gg \text{clink})_1$. Thus the query, which says that the gift is believed to be in the box and broken but the agent has no clue what the gift is, comes out true.

Q4. $[\text{dropbox}][\text{clink}][\text{unbox}(\#5)]\exists y\,\mathbf{B}(\text{gift} = y \wedge \neg\text{InBox(gift)} \wedge \neg\text{Broken(gift)})$.

We need to make another progression by unbox(#5). Firstly, the state is strongly revised by IF(unbox(#5)), which is equivalent to $\text{InBox}(\#5) \wedge \neg\text{Broken}(\#5)$. The first two levels of the revised state thus contain the IF(unbox(#5))-worlds from $(\vec{e} \gg \text{dropbox} \gg \text{clink})_3$ and $(\vec{e} \gg \text{dropbox} \gg \text{clink})_4$. For space reasons we only consider the first plausibility level, which is

$$((\vec{e} \gg \text{dropbox} \gg \text{clink}) * \text{unbox}(\#5))_1 = \{w \mid w \models \Sigma_{\text{dyn}} \wedge \forall y\,(\text{InBox}(y) \equiv y = \#5) \wedge \text{gift} = \#5 \wedge \neg\text{Broken}(\#5) \wedge \neg\text{Fragile}(\#5)\},$$

and when we then apply the physical effect of unbox(#5), namely make InBox(#5) false, we obtain

$$(\vec{e} \gg \text{dropbox} \gg \text{clink} \gg \text{unbox}(\#5))_1 = \{w \mid w \models \Sigma_{\text{dyn}} \wedge \forall y\,\neg\text{InBox}(y) \wedge \text{gift} = \#5 \wedge \neg\text{InBox}(\#5) \wedge \neg\text{Broken}(\#5) \wedge \neg\text{Fragile}(\#5)\}.$$

The query is thus true, because all worlds at the first plausibility level agree on gift being #5.

It is remarkable that Q4 would not have come out true if clink was a strong-revision action. Then the agent would have rather believed that there were two (or more) items in the box than that the clink was due to something other than an object in the box breaking. That is quite reasonable: making clink a strong-revision action would have meant strong trust in its information; the agent would therefore be reluctant to give up the belief that something inside the box broke.

5.5 Projection by regression

The first solution we offer for the belief projection problem is by regression. Regression rewrites a formula about future situations to a formula about the initial situation. The idea, due to Reiter (Reiter 1991, 2001), is to successively replace subformulas $[t]F(t_1, \ldots, t_k)$ with the right-hand side of F's successor-state axiom γ_F. Intuitively this is sound because the successor-state axioms ensure that action effects are deterministic. As we shall see in this section, we can regress conditional beliefs after actions in a similar way.

Definition 5.5.1 A formula that mentions no $\Box$ or **O** and no fluent predicates other than those from $\mathcal{F} \cup \{\mathrm{IF}\}$ is called *regressable*.

To ease the technical treatment we assume that the formula to be regressed adheres to the following form:

- it is rectified: quantifiers use distinct variables, and none of the variables occurs in the basic action theory;
- the t_i in action terms $A(t_1, \ldots, t_n)$ are standard names or variables.

It is easy to see that any formula can be rewritten to satisfy these constraints. For example, IF(unbox(gift)) is transformed to $\exists x(x = \mathrm{gift} \wedge \mathrm{IF}(\mathrm{unbox}(x)))$. The first restriction is needed because otherwise scopes of variables may collide during regression. The second one will allow us to push action operators inside **B**, which would be inappropriate for action terms like unbox(gift) because the denotation of gift shall be determined by the real world.

Definition 5.5.2 The *regression* of an *objective* regressable formula α after actions r with respect to a basic action theory with dynamic axioms Σ_{dyn} is defined as follows:

R1. $\mathcal{R}[r, R(t_1, \ldots, t_k)] = R(t_1, \ldots, t_k)$ for rigid R;

R2. $\mathcal{R}[r, F(t_1, \ldots, t_k)]$ for fluent $F \in \mathcal{F}$ is defined inductively on r:

- $\mathcal{R}[\langle\rangle, F(t_1, \ldots, t_k)] = F(t_1, \ldots, t_k)$;
- $\mathcal{R}[r \cdot t, F(t_1, \ldots, t_k)] = \mathcal{R}[r, \gamma_F{}^{x_1 \ldots x_k\, a}_{t_1 \ldots t_k\, t}]$;

R3. $\mathcal{R}[r, \mathrm{IF}(t)] = \mathcal{R}[r, \varphi^a_t]$;

R4. $\mathcal{R}[r, (t_1 = t_2)] = (t_1 = t_2)$;

R5. $\mathcal{R}[r, \neg\alpha] = \neg\mathcal{R}[r, \alpha]$;

R6. $\mathcal{R}[r, (\alpha_1 \vee \alpha_2)] = (\mathcal{R}[r, \alpha_1] \vee \mathcal{R}[r, \alpha_2])$;

R7. $\mathcal{R}[r, \exists x\, \alpha] = \exists x \mathcal{R}[r, \alpha]$;

R8. $\mathcal{R}[r, [t]\alpha] = \mathcal{R}[r \cdot t, \alpha]$.

We write $\mathcal{R}[\alpha]$ for $\mathcal{R}[\langle\rangle, \alpha]$.

With that definition, we can already regress objective formulas. The following theorem states that regression soundly takes the dynamics out of the reasoning task.

Theorem 5.5.3 *Let Σ_{dyn} be the dynamic axioms of a basic action theory, ϕ be a fluent sentence, and ψ be an objective regressable sentence. Then $\Sigma_{\text{dyn}} \wedge \phi \models \psi$ iff $\phi \models \mathcal{R}[\psi]$.*

The proof is in Appendix B.2. Induction proofs about regression are quite involved because formulas grow during regression. Our proofs are unconventional in that we use a non-standard length measure (Definition B.2.9). Once it is shown that this measure is well-behaved for inductions, the actual proofs come out quite easily. Besides that, the idea is similar to the regression theorem for $\mathcal{ES}$ sketched by Lakemeyer and Levesque (2011).

The key to extending regression to conditional beliefs is the relationship between beliefs after an action and the conditional beliefs before that action. The following two theorems establish such a relationship for weak- and strong-revision actions, respectively. We will use this correspondence to regress beliefs similarly to how we use successor-state axioms to regress fluent atoms.

Theorem 5.5.4 *Let a be a weak-revision action variable. Then*

$$\begin{aligned}\models \Box[a]\mathbf{B}(\alpha \Rightarrow \beta) \equiv \neg\mathbf{B}(\mathrm{IF}(a) \Rightarrow \neg[a]\alpha) \wedge \mathbf{B}(\mathrm{IF}(a) \wedge [a]\alpha \Rightarrow [a]\beta) \vee{} \\ \mathbf{B}(\mathrm{IF}(a) \Rightarrow \neg[a]\alpha) \wedge \mathbf{B}([a]\alpha \Rightarrow [a]\beta) \vee{} \\ \mathbf{B}(\mathrm{IF}(a) \Rightarrow \textsc{false}).\end{aligned}$$

The proof is in Appendix B.2. Intuitively the disjunction on the right-hand side considers three different cases. Action a triggers a revision, which promotes certain worlds to the first plausibility level. In the first case, at least one of these worlds satisfies α after a, and therefore we need to consider information learned by a in the antecedent. In the second case, none of them satisfies α after a, and therefore the revision is not relevant to the belief. The third case deals with revision by inconsistent information. The formal proof follows that intuition.

Theorem 5.5.5 *Let a be a strong-revision action variable. Then*

$$\begin{aligned}\models \Box[a]\mathbf{B}(\alpha \Rightarrow \beta) \equiv {} & \neg\mathbf{B}(\mathrm{IF}(a) \wedge [a]\alpha \Rightarrow \textsc{false}) \wedge \mathbf{B}(\mathrm{IF}(a) \wedge [a]\alpha \Rightarrow [a]\beta) \vee \\ & \mathbf{B}(\mathrm{IF}(a) \wedge [a]\alpha \Rightarrow \textsc{false}) \wedge \mathbf{B}([a]\alpha \Rightarrow [a]\beta) \vee \\ & \mathbf{B}(\mathrm{IF}(a) \Rightarrow \textsc{false}).\end{aligned}$$

The proof is in Appendix B.2. The three cases on the right-hand side are similar to the ones for weak revision in Theorem 5.5.4. The strong revision caused by a promotes all IF(a)-worlds over all ¬IF(a)-worlds. In case some of the former worlds satisfy α after a, some of them make up the most-plausible α-worlds after a, so the belief must also be conditioned on IF(a). This is covered by the first case. Otherwise, if none of the promoted worlds satisfies α after a, the revision is irrelevant for that particular conditional belief. The third case deals with revision by inconsistent information. The formal argument follows this intuition and proceeds generally similar to the one of Theorem 5.5.4.

Theorems 5.5.4 and 5.5.5 resemble successor-state axioms in that the action a occurs outside of the belief at the left-hand side, but not at the right-hand side of the equivalence. We can use them in a way similar to Rule R2 to push the action inside the scope of **B**. Once that is done, regression proceeds with the antecedent and consequent in **B**.

Definition 5.5.6 The regression of a regressable formula α is defined as in Definition 5.5.2 plus the following rule:

R9. $\mathcal{R}[r, \mathbf{B}(\alpha \Rightarrow \beta)]$ is defined inductively on r:

- $\mathcal{R}[\langle\rangle, \mathbf{B}(\alpha \Rightarrow \beta)] = \mathbf{B}(\mathcal{R}[\alpha] \Rightarrow \mathcal{R}[\beta])$;
- $\mathcal{R}[r \cdot t, \mathbf{B}(\alpha \Rightarrow \beta)] = \mathcal{R}[r, \sigma^a_t]$ where σ is the right-hand side of Theorem 5.5.4 or 5.5.5 depending on the sort of t.

This completes the regression operator. The following theorem states its correctness.

Theorem 5.5.7 *Let $\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}}$ be a basic action theory and let α be a regressable sentence. Then* $\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}}) \models \alpha$ *iff* $\mathbf{O}\Sigma_{\mathrm{bel}} \models \mathcal{R}[\alpha]$.

The proof can be found in B.2. Let us illustrate regression using the gift-giving example.

Example 5.5.8 Consider query Q4 from Example 5.4.2:

$$\begin{aligned}&\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}}) \models \\ &\qquad [\mathrm{dropbox}][\mathrm{clink}][\mathrm{unbox}(\#5)]\exists y\, \mathbf{B}(\mathrm{gift} = y \wedge \neg\mathrm{InBox}(\mathrm{gift}) \wedge \neg\mathrm{Broken}(\mathrm{gift})).\end{aligned}$$

We first regress $[\mathrm{unbox}(^{\#}5)]\exists y\,\mathbf{B}(\mathrm{gift} = y \land \neg\mathrm{InBox}(\mathrm{gift}) \land \neg\mathrm{Broken}(\mathrm{gift}))$, and then show that the regressed sentence is satisfied by $\vec{e} \gg \mathrm{dropbox} \gg \mathrm{clink}$, the progression of the model of $\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}})$, which we determined in Example 5.4.3. We do not regress by clink and dropbox here for space reasons; we handle them in Example 5.8.4 by progression. After rewriting the formula to adhere to the normal form required for regression, the task is to determine

$$\mathcal{R}[[\mathrm{unbox}(^{\#}5)]\exists y\,\mathbf{B}\exists y'(\mathrm{gift} = y' \land \mathrm{gift} = y \land \neg\mathrm{InBox}(y') \land \neg\mathrm{Broken}(y'))].$$

Regression then moves inside the existential and the action $\mathrm{unbox}(^{\#}5)$ and we obtain

$$\exists y\,\mathcal{R}[\mathrm{unbox}(^{\#}5), \mathbf{B}\exists y'(\mathrm{gift} = y' \land \mathrm{gift} = y \land \neg\mathrm{InBox}(y') \land \neg\mathrm{Broken}(y'))].$$

The action $\mathrm{unbox}(^{\#}5)$ is then pushed inside of the belief modalities and we obtain, after minor simplifications,

$$\begin{aligned}
&\exists y\,(\mathcal{R}[\neg\mathbf{B}(\mathrm{IF}(\mathrm{unbox}(^{\#}5)) \Rightarrow \textsc{false}) \land \mathbf{B}(\mathrm{IF}(\mathrm{unbox}(^{\#}5)) \Rightarrow \psi)] \lor \\
&\quad \mathcal{R}[\mathbf{B}(\mathrm{IF}(\mathrm{unbox}(^{\#}5)) \Rightarrow \textsc{false}) \land \mathbf{B}\psi)] \lor \\
&\quad \mathcal{R}[\mathbf{B}(\mathrm{IF}(\mathrm{unbox}(^{\#}5)) \Rightarrow \textsc{false})]) \\
&\text{where } \psi = [\mathrm{unbox}(^{\#}5)]\exists y'(\mathrm{gift} = y' \land \mathrm{gift} = y \land \neg\mathrm{InBox}(y') \land \neg\mathrm{Broken}(y')).
\end{aligned}$$

Now regression proceeds inside the belief modalities with the antecedents and consequents. In particular, regressing ψ substitutes $\mathrm{InBox}(y')$ and $\mathrm{Broken}(y')$ with the right-hand sides of the successor-state axioms:

$$\begin{aligned}
\mathcal{R}[\psi] = \exists y'(&\mathrm{gift} = y' \land \mathrm{gift} = y \land \\
&\neg(\mathrm{InBox}(y') \land \mathrm{unbox}(^{\#}5) \neq \mathrm{unbox}(y')) \land \\
&\neg(\mathrm{Broken}(y') \lor \mathrm{InBox}(y') \land \mathrm{Fragile}(y') \land \mathrm{unbox}(^{\#}5) = \mathrm{dropbox})).
\end{aligned}$$

After some trivial simplifications, the final regressed formula is equivalent to

$$\begin{aligned}
&(\neg\mathbf{B}(\mathrm{InBox}(^{\#}5) \land \neg\mathrm{Broken}(^{\#}5) \Rightarrow \textsc{false}) \land \\
&\quad \exists y\,\mathbf{B}(\mathrm{InBox}(^{\#}5) \land \neg\mathrm{Broken}(^{\#}5) \Rightarrow \mathrm{gift} = {}^{\#}5 \land \mathrm{gift} = y \land \neg\mathrm{Broken}(^{\#}5))) \lor \\
&\mathbf{B}(\mathrm{InBox}(^{\#}5) \land \neg\mathrm{Broken}(^{\#}5) \Rightarrow \textsc{false}).
\end{aligned}$$

Finally we need to prove that $\vec{e} \gg \mathrm{dropbox} \gg \mathrm{clink}$ satisfies this formula. Note that there are $w \in (\vec{e} \gg \mathrm{dropbox} \gg \mathrm{clink})_3$ with $w \models \mathrm{InBox}(^{\#}5) \land \neg\mathrm{Broken}(^{\#}5)$. Consider any such w. Since such worlds do exist, $\neg\mathbf{B}(\text{InBox}(^{\#}5) \wedge \neg\text{Broken}(^{\#}5) \Rightarrow \text{FALSE})$ is true, and therefore we need to prove that $\exists y\,\mathbf{B}(\text{InBox}(^{\#}5) \wedge \neg\text{Broken}(^{\#}5) \Rightarrow \text{gift} = {}^{\#}5 \wedge \text{gift} = y \wedge \neg\text{Broken}(^{\#}5))$ is true as well. We substitute $^{\#}5$ for the existentially quantified y. Since $w \models \text{InBox}(^{\#}5)$ by assumption and nothing but gift is in the box at level $(\vec{e} \gg \text{dropbox} \gg \text{clink})_3$, w also satisfies the consequent, namely $w \models \text{gift} = {}^{\#}5 \wedge (\text{gift} = y)^{y}_{\#5} \wedge \neg\text{Broken}(^{\#}5)$. Thus $\vec{e} \gg \text{dropbox} \gg \text{clink}$ satisfies the regressed formula. We finish the proof of the query in Example 5.8.4 where we deal with dropbox and clink by progression.

This completes the discussion of our first solution of the belief projection problem, and we turn to progression next.

5.6 Forgetting in only-believing

Progressing a knowledge base by some action n means to update that knowledge base according to the effects of n. In other words, we want to *forget* the initial knowledge and let an updated knowledge base take the place of the now obsolete knowledge. This correspondence between progression and forgetting was first observed by Lin and Reiter (1997). Their notion of forgetting is irreversible, and it is therefore not to be confused with belief contraction known from belief revision theory. Unfortunately, even the seemingly simple forgetting-as-erasure is already highly complex: in general, it requires second-order logic (Lin and Reiter 1994, 1997).

The relationship between forgetting and second-order logic is easily seen by an example. Consider the set of worlds $e = \{w \mid w \models (G_1 \wedge G_2)\}$, which expresses that all we know is G_1 and G_2. Forgetting G_2 should lead to additional possible worlds, namely those which still satisfy G_1 but perhaps not G_2. Assuming an appropriate second-order extension of our language, the resulting set of worlds could be represented as $e' = \{w \mid w \models \exists G_2(G_1 \wedge G_2)\}$, where G_2 is now an existentially quantified second-order variable. (As for the semantics of the existential, we would want something like $w \models \exists G_2(G_1 \wedge G_2)$ iff $w' \models (G_1 \wedge G_2)$ for some w' that agrees with w on everything except perhaps G_2.)

The reader may argue that the representation with second-order quantification is overly complex, as we could simply write $e' = \{w \mid w \models G_1\}$. Indeed an atomic proposition can be easily forgotten without second-order logic by replacing all occurrences with TRUE and FALSE and taking the disjunction of the respective sentences (Lin and Reiter 1994). However, things get more complicated when first-order logic comes into play and whole relations, not just atomic facts, need to be forgotten. As a matter of fact, Lin and Reiter (1997) present an example for which forgetting of a relation cannot be defined with first-order logic alone.

In this and the next two sections, we adopt the idea of progression of a basic action theory through syntactic manipulation using existentially second-order variables. It was introduced by Lin and Reiter (1997) in a non-epistemic setting. The first question that arises here is therefore how to extend our language with second-order quantifiers. The subsequent sections then investigate, firstly, how the revision of a knowledge base can be represented, and, finally, how physical and epistemic effects on a knowledge can be accounted for.

For the purposes of forgetting we need to introduce second-order variables in only-believing expressions $\mathbf{O}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$. But what should be their scope? The variables clearly should be quantified within $\mathbf{O}$, yet its scope should encompass all α_i and β_i, so that all occurrences in α_i, β_i refer to the same variable. Adding full support of second-order quantifiers and allowing them to appear between the $\mathbf{O}$ and its arguments $\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ requires a quite cumbersome semantics, though. On the other hand, full second-order logic is not even required for forgetting – existentials between $\mathbf{O}$ and $\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ suffice. As it permits a much simpler semantics, we parameterize the only-believing operator with a finite set of function and predicate symbols, which are taken to be existentially quantified inside $\mathbf{O}$.

Definition 5.6.1 The set of well-formed formulas is the least set formed from the rules from Definition 5.2.1 and

- $\mathbf{O}_{\mathcal{S}}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ is a formula where the α_i and β_i are formulas and $\mathcal{S}$ is a finite set of object function and predicate symbols.

We say α is *$\mathcal{S}$-free* when it mentions no object function or predicate symbol from $\mathcal{S}$.

Due to the relationship between existential quantification and forgetting, we read $\mathbf{O}_{\mathcal{S}}\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ as "before everything about $\mathcal{S}$ is forgotten, the conditionals $\alpha_i \Rightarrow \beta_i$ are all that is believed." We let $\mathcal{S}$ stand for a finite set of object function and predicate symbols for the rest of this chapter.

To characterize the semantics of existential quantification, we use the following relation to say that two worlds agree on everything except perhaps certain symbols.

Definition 5.6.2 We define $w \approx_{\mathcal{S}} w'$ iff

- $w[g(n_1, \ldots, n_k)] = w'[g(n_1, \ldots, n_k)]$ for all object function symbols $g \notin \mathcal{S}$;
- $w[R(n_1, \ldots, n_k)] = w'[R(n_1, \ldots, n_k)]$ for all rigid predicate symbols $R \notin \mathcal{S}$;

- $w[F(n_1,\ldots,n_k),z] = w'[F(n_1,\ldots,n_k),z]$ for all fluent predicate symbols $F \notin \mathcal{S}$ and action sequences z.

For a set of worlds W and an epistemic state $\vec{e}$, we let $W_{\mathcal{S}} = \{w' \mid w \in W \text{ and } w \approx_{\mathcal{S}} w'\}$ and $\vec{e}_{\mathcal{S}} = \langle (e_1)_{\mathcal{S}},\ldots,(e_{\lceil\vec{e}\rceil})_{\mathcal{S}}\rangle$.

Intuitively, $w \approx_{\mathcal{S}} w'$ means that w and w' agree on everything except perhaps $\mathcal{S}$. Notice that $w \approx_{\{\}} w'$ iff $w = w'$. The epistemic state $\vec{e}_{\mathcal{S}}$ is the result of forgetting everything about $\mathcal{S}$ in $\vec{e}$. For example, suppose one believes that R and $(R \equiv R')$ in $\vec{e}$ and $e_1 \neq \{\}$. Then $w[R] = w[R'] = 1$ for all $w \in e_1$. Belief in R is then lost in $\vec{e}_{\{R\}}$, while R' is retained: for each $w \in e_1$, not only $w \in (\vec{e}_{\{R\}})_1$, but there also is a $w' \in (\vec{e}_{\{R\}})_1$ which agrees with w on everything except that $w'[R] = 0$.

Definition 5.6.3 The semantics of the new only-believing operator is defined using standard only-believing:

$\mathcal{ESB}$11. $\vec{e},w \models \mathbf{O}_{\mathcal{S}}\{\alpha_1 \Rightarrow \beta_1,\ldots,\alpha_m \Rightarrow \beta_m\}$ iff
$\vec{e}',w \models \mathbf{O}\{\alpha_1 \Rightarrow \beta_1,\ldots,\alpha_m \Rightarrow \beta_m\}$ and $\vec{e} = \vec{e}'_{\mathcal{S}}$ for some $\vec{e}'$.

Note that extended only-believing subsumes standard only-believing, as their semantics coincide for $\mathcal{S} = \{\}$. It is not surprising that earlier results such as the unique-model property from Theorem 5.3.16 and the regression result Theorem 5.5.7 carry over to the extended operator.

Corollary 5.6.4 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1,\ldots,\phi_m \Rightarrow \psi_m\}$ *be objective.*
Then there is a unique $\vec{e} = \langle e_1,\ldots,e_{m+1}\rangle$ *such that* $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$.

Proof. $\vec{e}' \models \mathbf{O}\Gamma$ is unique by Theorem 5.3.16, so $\vec{e} = \vec{e}'_{\mathcal{S}}$ is unique by Rule $\mathcal{ESB}$11. □

Theorem 5.6.5 *Let* $\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}$ *be a basic action theory with* $\mathcal{S}$*-free* Σ_{dyn} *and let* α *be a regressable sentence. Then* $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models \alpha$ *iff* $\mathbf{O}_{\mathcal{S}}\Sigma_{\text{bel}} \models \mathcal{R}[\alpha]$.

The proof is an easy consequence of the original regression theorem shown in Appendix B.2.

5.7 Revision of only-believing

How does a conditional knowledge base change when new information comes in? An answer to this brings us already half way to knowledge base progression. Because to update a knowledge base when an action occurs, progression needs to take into account both the revision effect and the physical effect of an action. Here we are only concerned with the former. The goal is to find for a set of objective conditionals Γ a revised set $\Gamma * \upsilon$ that matches the semantic revision by υ.

Semantically, performing an action brings along a revision of the epistemic state, which promotes certain worlds over others. In this section we examine how the semantic revision can be matched syntactically. More precisely, we are looking for a set of conditionals $\Gamma * \upsilon$ which is only-believed when Γ was only-believed before revising by υ.

Recall that by Theorem 5.3.14, $\mathbf{B}(\alpha \vee \beta \Rightarrow \neg\beta)$ asserts that α is more plausible than β or both are considered impossible. We use this to define Γ_δ as the set conditionals whose material implication holds in the first sphere consistent with δ. Γ_δ will be helpful to characterize different spheres in order to represent revision by δ.

Definition 5.7.1 $\Gamma_\delta = \{\alpha \Rightarrow \beta \in \Gamma \mid \mathbf{O}\Gamma \models \mathbf{B}(\delta \vee \neg(\alpha \supset \beta) \Rightarrow (\alpha \supset \beta))\}$.

A syntactic representation of weak revision by υ needs to reflect that the most-plausible υ-worlds are promoted to the first level. To this end we use a new dummy predicate R to partition the worlds: the R-worlds represent those which are promoted, and the $\neg R$-worlds represent the beliefs before the revision.

Definition 5.7.2 Let R be a rigid predicate symbol. Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ and υ be objective and $\{R\}$-free. Then the *weak revision* of Γ by υ is

$$\begin{aligned}
\Gamma *_{\mathrm{w}} \upsilon = \{&\textsc{true} && \Rightarrow R\} && \cup \\
\{&\neg(R \supset \upsilon) && \Rightarrow \textsc{false}\} && \cup \\
\{&\neg(R \supset (\phi \supset \psi)) && \Rightarrow \textsc{false} \mid \phi \Rightarrow \psi \in \Gamma_\upsilon\} && \cup \\
\{&(\neg R \wedge \phi) && \Rightarrow \psi \quad\ \mid \phi \Rightarrow \psi \in \Gamma\}.
\end{aligned}$$

We now prove that the revised set of conditionals (after forgetting R) matches semantical weak revision.

Theorem 5.7.3 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *and* υ *be objective and* $\mathcal{S}$*-free. Let* R *be the nullary rigid predicate newly introduced in* $\Gamma *_{\mathrm{w}} \upsilon$.
If $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$, *then* $\vec{e} *_{\mathrm{w}} \upsilon \models \mathbf{O}_{\mathcal{S}\cup\{R\}}\Gamma *_{\mathrm{w}} \upsilon$.

The proof is in Appendix B.3. It proceeds by constructing a state $\vec{e}'$ whose first sphere is $e'_1 = ((\vec{e} *_{\mathrm{w}} \upsilon) \mid R)_1$ and the subsequent spheres are $e'_p = ((\vec{e} *_{\mathrm{w}} \upsilon) \mid \neg R)_p$ for $p > 1$. Then $\vec{e}'$ is a model of $\mathbf{O}\Gamma *_{\mathrm{w}} \upsilon$, and forgetting $\mathcal{S} \cup \{R\}$ obtains the theorem.

Strong revision changes the ranking of the worlds more profoundly than weak revision, and representing this change is hence more complex. Strong revision by υ promotes all υ-worlds over all $\neg\upsilon$-worlds. We therefore duplicate the conditionals from Γ twice using new predicates, and require υ to be true in the first copy. The revised truth values are then set through additional conditionals based on the dummies' truth values.

To ease the presentation, we restrict our consideration of strong revision to static formulas. They are sufficient for our purposes of progressing a basic action theories, since the initial beliefs Σ_{bel} are static as well. Extending the below definition and theorem for non-static conditionals is straightforward.

Definition 5.7.4 Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ and υ be objective and static. Let $\mathcal{S}'$ be the object function and predicate symbols in Γ, and let $\mathcal{S}''$ be just as many object function and rigid predicate symbols of corresponding arity which do not occur in Γ or υ. For any formula β, let β^* be the formula obtained from β by replacing each symbol from $\mathcal{S}'$ with the corresponding symbol from $\mathcal{S}''$. Let $\Delta = \{\phi \Rightarrow \psi \in \Gamma_\upsilon \mid \mathbf{O}\Gamma \not\models \mathbf{K}(\phi \supset \psi)\}$. Then the *strong revision* of Γ by υ is defined as

$$\begin{aligned}
\Gamma *_{\mathrm{s}} \upsilon = {} & \{\phi^* \Rightarrow \psi^* \mid \phi \Rightarrow \psi \in \Gamma_\upsilon\} & \cup \\
& \{(\phi^* \wedge \neg\upsilon) \Rightarrow \psi^* \mid \phi \Rightarrow \psi \in \Gamma_{\neg\upsilon}\} & \cup \\
& \{\textsc{true} \Rightarrow \upsilon\} \cup \{\neg(\phi^* \supset \psi^*) \vee \neg\upsilon \Rightarrow \upsilon \mid \phi \Rightarrow \psi \in \Delta\} & \cup \\
& \{\neg((\upsilon \wedge \neg\upsilon^*) \supset (\phi \supset \psi)) \Rightarrow \textsc{false} \mid \phi \Rightarrow \psi \in \Gamma_\upsilon\} & \cup \\
& \{\neg((\neg\upsilon \wedge \upsilon^*) \supset (\phi \supset \psi)) \Rightarrow \textsc{false} \mid \phi \Rightarrow \psi \in \Gamma_{\neg\upsilon}\} & \cup \\
& \{\neg((\upsilon \equiv \upsilon^*) \supset (\phi \equiv \phi^*) \wedge (\psi \equiv \psi^*)) \Rightarrow \textsc{false} \mid \phi \Rightarrow \psi \in \Gamma\}. &
\end{aligned}$$

The first and the third line account for the promoted υ-worlds in the revised epistemic state. In particular, the third line asserts that there is the right number of levels where υ holds. The material implications in the last three lines set the original predicates according to the values of the dummy predicates.

As with weak revision, the syntactic strong revision $\Gamma *_{\mathrm{s}} \upsilon$ matches its semantic counterpart.

Theorem 5.7.5 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ and υ be objective and static. Let $\mathcal{S}$ be a finite set of object function and predicate symbols, and let υ be $\mathcal{S}$-free. Let $\mathcal{S}''$ be the symbols newly introduced in $\Gamma *_{\mathrm{s}} \upsilon$. If $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$, then $\vec{e} *_{\mathrm{s}} \upsilon \models \mathbf{O}_{\mathcal{S}\cup\mathcal{S}''}\Gamma *_{\mathrm{s}} \upsilon$.*

The proof is structurally similar to the proof of Theorem 5.7.3 and can be found in Appendix B.3.

5.8 Projection by progression

With the preparatory work from the preceding two sections, we are now ready to define the progression of a basic action theory $\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}}$. Given an action standard name n, we first revise the theory by n's information and then apply n's effects on fluents. The revision is captured by $\Sigma_{\text{bel}} * \varphi^a_n$ where φ is from the informed-fluent axiom $\Box\text{IF}(a) \equiv \varphi \in \Sigma_{\text{dyn}}$, and the type of revision corresponds to the subsort of n. (The reason for taking φ^a_n instead of IF(n) is to keep the belief conditionals fluent.) In this section we show how the physical effects of n are handled.

For two sets of predicate symbols $\mathcal{F} = \{F_1, \ldots, F_k\}$ and $\mathcal{R} = \{R_1, \ldots, R_k\}$ of corresponding arity we denote by $\alpha^{\mathcal{F}}_{\mathcal{R}}$ the formula obtained by replacing each F_i with R_i in α.

Definition 5.8.1 Let $\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}$ be a basic action theory over fluents $\mathcal{F} = \{F_1, \ldots, F_k\}$, and let $\mathcal{R} = \{R_1, \ldots, R_k\}$ be rigid predicates of corresponding arity which do not otherwise occur in Σ_{dyn} or $\Sigma_{\text{bel}} * \varphi^a_n$. Let n be a action standard name. Then the *progression* of Σ_{bel} by n is defined as

$$\Sigma_{\text{bel}} \gg n = (\Sigma_{\text{bel}} * \varphi^a_n)^{\mathcal{F}}_{\mathcal{R}} \cup \{\neg\forall x_1 \ldots \forall x_l (F(x_1, \ldots, x_l) \equiv \gamma_F{}^a_n{}^{\mathcal{F}}_{\mathcal{R}}) \Rightarrow \textsc{false} \mid F \in \mathcal{F}\}.$$

The intuition behind the definition is as follows. When n is executed, the beliefs are first revised by the information φ^a_n produced by n, which leads to $\Sigma_{\text{bel}} * \varphi^a_n$. The beliefs $(\Sigma_{\text{bel}} * \varphi^a_n)^{\mathcal{F}}_{\mathcal{R}}$ represent the same conditionals belief as $\Sigma_{\text{bel}} * \varphi^a_n$, except that each F_i is renamed to R_i. Intuitively, the R_i memorize the value of F_i before the physical effect of n. The additional conditionals in $\Sigma_{\text{bel}} \gg n$ initialize each fluent $F(x_1, \ldots, x_l)$ with its value after doing n, that is, $\gamma_F{}^a_n{}^{\mathcal{F}}_{\mathcal{R}}$. Notice that the progression of a basic action theory again is a basic action theory over $\mathcal{F}$, so progression can iterate.

The following two results establish the correctness of progression. The first theorem says that, if all that is believed is a basic action theory, then after doing action n all that is believed is the progressed basic action theory.

Theorem 5.8.2 *Let $\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}$ be a basic action theory. Let $\mathcal{S}'$ be the symbols newly introduced by $\Sigma_{\text{bel}} \gg n$. Then $\models \mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \supset [n]\mathbf{O}_{\mathcal{S}\cup\mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$.*

The second theorem says that the same beliefs are entailed by a basic action theory after doing action n and the progression by n of that basic action theory.

Theorem 5.8.3 *Let $\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}$ be a basic action theory. Let $\mathcal{S}'$ be the symbols newly introduced by $\Sigma_{\text{bel}} \gg n$. Then $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models [n]\alpha$ iff $\mathbf{O}_{\mathcal{S}\cup\mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n) \models \alpha$.*

The proofs of both theorems are in Appendix B.4. They proceed in two steps. First, we show that $\Sigma_{\text{bel}} * \varphi^a_n$ and $(\Sigma_{\text{bel}} * \varphi^a_n)^{\mathcal{F}}_{\mathcal{R}}$ determine the same conditional beliefs modulo the substitution of $\mathcal{F}$ by $\mathcal{R}$, where $\mathcal{R} \subseteq \mathcal{S}'$ is the set of rigid predicates from Definition 5.8.1. Second, we see that after progressing the individual worlds in the model of $\mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} * \varphi^a_n)$ by n, the resulting epistemic state agrees with the model of

$\mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$ on everything except perhaps $\mathcal{R}$. We therefore have that, when using standard only-believing, the semantic progression and the syntactic progression agree on everything except $\mathcal{R}$. Using the extended only-believing operator also the differences in $\mathcal{R}$ vanish.

Example 5.8.4 Let us proceed with query Q4 from Example 5.4.2 which we investigated in Example 5.5.8 already. The query involves the action sequence dropbox, clink, unbox(#5). Since dropbox is a physical action with no epistemic effect, let us take an abbreviation instead of doing it by the definitions: it is easy to see that $\vec{e} \gg \text{dropbox}$ from Example 5.4.4 satisfies $\mathbf{O}(\Sigma_{\text{dyn}}, \Sigma'_{\text{bel}})$ where

$$\begin{aligned}\Sigma'_{\text{bel}} = \{&\textsc{true} \Rightarrow \forall y\, \neg \text{InBox}(y),\\ &\exists y\, \text{InBox}(y) \Rightarrow \forall y\, (\text{InBox}(y) \equiv y = \text{gift}),\\ &\exists y\, \text{InBox}(y) \Rightarrow \forall y\, (\text{InBox}(y) \supset (\text{Broken}(y) \equiv \text{Fragile}(y))),\\ &\neg \forall y\, (\text{InBox}(y) \land \text{Fragile}(y) \supset \text{Broken}(y)) \Rightarrow \textsc{false}\}.\end{aligned}$$

We focus on the progression of $\mathbf{O}(\Sigma_{\text{dyn}}, \Sigma'_{\text{bel}})$ by the weak-revision action clink. According to Definition 5.7.2, the revision $\Sigma'_{\text{bel}} * \varphi^{a}_{\text{clink}}$ contains the conditionals

- $\textsc{true} \Rightarrow R$;
- $\neg(R \supset \varphi^{a}_{\text{clink}}) \Rightarrow \textsc{false}$;
- $\neg(R \supset (\phi \supset \psi)) \Rightarrow \textsc{false}$ for each $\phi \Rightarrow \psi \in \Sigma'_{\text{bel}}$ such that $\mathbf{O}\Sigma'_{\text{bel}} \models \mathbf{B}(\varphi^{a}_{\text{clink}} \lor \neg(\phi \supset \psi) \Rightarrow (\phi \supset \psi))$;
- $\neg R \land \phi \Rightarrow \psi$ for each $\phi \Rightarrow \psi \in \Sigma'_{\text{bel}}$.

This amounts to

$$\begin{aligned}\Sigma'_{\text{bel}} * \varphi^{a}_{\text{clink}} = \{&\textsc{true} \Rightarrow R,\\ &\neg(R \supset \exists y\, (\text{InBox}(y) \land \text{Broken}(y))) \Rightarrow \textsc{false},\\ &\neg(R \supset (\exists y\, \text{InBox}(y) \supset \forall y\, (\text{InBox}(y) \equiv y = \text{gift}))) \Rightarrow \textsc{false},\\ &\neg(R \supset (\exists y\, \text{InBox}(y) \supset \forall y\, (\text{InBox}(y) \supset (\text{Broken}(y) \equiv \text{Fragile}(y))))) \Rightarrow \textsc{false},\\ &\neg(R \supset (\neg \forall y\, (\text{InBox}(y) \land \text{Fragile}(y) \supset \text{Broken}(y)) \supset \textsc{false})) \Rightarrow \textsc{false},\\ &\textsc{true} \land \neg R \Rightarrow \forall y\, \neg \text{InBox}(y),\\ &\neg R \land \exists y\, \text{InBox}(y) \Rightarrow \forall y\, (\text{InBox}(y) \equiv y = \text{gift}),\\ &\neg R \land \exists y\, \text{InBox}(y) \Rightarrow \forall y\, (\text{InBox}(y) \supset (\text{Broken}(y) \equiv \text{Fragile}(y))),\\ &\neg R \land \neg \forall y\, (\text{InBox}(y) \land \text{Fragile}(y) \supset \text{Broken}(y)) \Rightarrow \textsc{false}\}.\end{aligned}$$

In the progression $\Sigma'_{\mathrm{bel}} \gg \mathrm{clink}$ the fluents InBox and Broken are renamed R_{InBox} and R_{Broken}, respectively, and two conditionals are added to set InBox and Broken to its correct value:

$$\begin{aligned}\Sigma'_{\mathrm{bel}} \gg \mathrm{clink} = {}& (\Sigma'_{\mathrm{bel}} * \varphi^{a}_{\mathrm{clink}})^{\mathrm{InBox}\ \mathrm{Broken}}_{R_{\mathrm{InBox}}\ R_{\mathrm{Broken}}} \cup \\ & \{\neg(\mathrm{InBox}(y) \equiv R_{\mathrm{InBox}}(y) \land \mathrm{clink} \neq \mathrm{unbox}(y)) \Rightarrow \textsc{false}, \\ & \neg(\mathrm{Broken}(y) \equiv R_{\mathrm{Broken}}(y) \lor \\ & \qquad R_{\mathrm{InBox}}(y) \land \mathrm{Fragile}(y) \land \mathrm{clink} = \mathrm{dropbox}) \Rightarrow \textsc{false}\}.\end{aligned}$$

Let us consider projection problem Q4 from Example 5.4.2 another time:

$$\begin{aligned}&\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}}) \models \\ &\quad [\mathrm{dropbox}][\mathrm{clink}][\mathrm{unbox}(\#5)]\exists y\, \mathbf{B}(\mathrm{gift} = y \land \neg\mathrm{InBox}(\mathrm{gift}) \land \neg\mathrm{Broken}(\mathrm{gift})).\end{aligned}$$

So far we have determined the progression of Σ_{bel} by dropbox and clink, and in Example 5.5.8 we regressed the query by unbox($\#5$). Together, this projection problem can therefore be recast as the purely static entailment problem

$$\begin{aligned}&\mathbf{O}_{\{R, R_{\mathrm{InBox}}, R_{\mathrm{Broken}}\}}(\Sigma_{\mathrm{dyn}}, \Sigma'_{\mathrm{bel}} \gg \mathrm{clink}) \models \\ &\quad \neg\mathbf{B}(\mathrm{InBox}(\#5) \land \neg\mathrm{Broken}(\#5) \Rightarrow \textsc{false}) \land \\ &\qquad \mathbf{B}(\mathrm{InBox}(\#5) \land \neg\mathrm{Broken}(\#5) \Rightarrow \mathrm{gift} = \#5 \land \neg\mathrm{Broken}(\#5)) \lor \\ &\quad \mathbf{B}(\mathrm{InBox}(\#5) \land \neg\mathrm{Broken}(\#5) \Rightarrow \textsc{false}).\end{aligned}$$

The equivalence of both entailment problems follows from the progression and regression results, Theorems 5.8.3 and 5.6.5. Let us confirm for this example that the (regressed) query indeed is a logical consequence of the (progressed) theory. In Example 5.4.4 we proved that $\vec{e} \gg \mathrm{dropbox} \gg \mathrm{clink}$ satisfies the regressed query. So it suffices to show that $\vec{e} \gg \mathrm{dropbox} \gg \mathrm{clink}$ is the model of $\mathbf{O}_{\{R, R_{\mathrm{InBox}}, R_{\mathrm{Broken}}\}}(\Sigma_{\mathrm{dyn}}, \Sigma'_{\mathrm{bel}} \gg \mathrm{clink})$. Following the procedure from the proof of Lemma 4.5.2, $\vec{e}\,' \models \mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma'_{\mathrm{bel}} \gg \mathrm{clink})$ iff $\vec{e}\,' = \langle e'_1, e'_2, e'_3, e'_4\rangle$ where

$$\begin{aligned}e'_1 = \{w \mid w \models {}& \Sigma_{\mathrm{dyn}} \land R \land \\ & (R \supset \forall y (R_{\mathrm{InBox}}(y) \equiv y = \mathrm{gift}) \land R_{\mathrm{Broken}}(\mathrm{gift}) \land \mathrm{Fragile}(\mathrm{gift})) \land \\ & \forall y (\mathrm{InBox}(y) \equiv R_{\mathrm{InBox}}(y)) \land \forall y (\mathrm{Broken}(y) \equiv R_{\mathrm{Broken}}(y))\};\end{aligned}$$

$$
\begin{aligned}
e_2' = \{w \mid w \models \Sigma_{\text{dyn}} \wedge{} & (R \supset \forall y (R_{\text{InBox}}(y) \equiv y = \text{gift}) \wedge R_{\text{Broken}}(\text{gift}) \wedge \text{Fragile}(\text{gift})) \wedge \\
& (\neg R \supset \forall y \neg R_{\text{InBox}}(y)) \wedge \\
& \forall y (\text{InBox}(y) \equiv R_{\text{InBox}}(y)) \wedge \forall y (\text{Broken}(y) \equiv R_{\text{Broken}}(y))\}; \\
e_3' = \{w \mid w \models \Sigma_{\text{dyn}} \wedge{} & (R \supset \forall y (R_{\text{InBox}}(y) \equiv y = \text{gift}) \wedge R_{\text{Broken}}(\text{gift}) \wedge \text{Fragile}(\text{gift})) \wedge \\
& (\neg R \supset \forall y (R_{\text{InBox}}(y) \supset y = \text{gift} \wedge (R_{\text{Broken}}(y) \equiv \text{Fragile}(y)))) \wedge \\
& \forall y (\text{InBox}(y) \equiv R_{\text{InBox}}(y)) \wedge \forall y (\text{Broken}(y) \equiv R_{\text{Broken}}(y))\}; \\
e_4' = \{w \mid w \models \Sigma_{\text{dyn}} \wedge{} & (R \supset \forall y (R_{\text{InBox}}(y) \equiv y = \text{gift}) \wedge R_{\text{Broken}}(\text{gift}) \wedge \text{Fragile}(\text{gift})) \wedge \\
& (\neg R \supset \forall y (R_{\text{InBox}}(y) \wedge \text{Fragile}(y) \supset R_{\text{Broken}}(y))) \wedge \\
& \forall y (\text{InBox}(y) \equiv R_{\text{InBox}}(y)) \wedge \forall y (\text{Broken}(y) \equiv R_{\text{Broken}}(y))\};
\end{aligned}
$$

and then it is easy to verify that $\vec{e}\,'_{\{R, R_{\text{InBox}}, R_{\text{Broken}}\}} = \vec{e} \gg \text{dropbox} \gg \text{clink}$: the R-worlds in $\vec{e}\,'$ represent the worlds from $(\vec{e} \gg \text{dropbox} \gg \text{clink})_1$, and the $\neg R$-worlds represent the additional worlds in $(\vec{e} \gg \text{dropbox} \gg \text{clink})_p$. Since there is no other model by Corollary 5.6.4, the static entailment equivalent to Q4 shown above indeed holds.

5.9 Representation theorem

So far in this chapter, we have seen how actions can be taken out of the belief projection problem, which leaves us with static reasoning about beliefs as in $\mathcal{BO}$. In Section 4.8 we saw that such belief entailments can be reduced to ordinary non-modal reasoning. It thus appears straightforward to extend the representation theorem for $\mathcal{BO}$ to the static formulas of $\mathcal{ESB}$.

At first sight, this may seem trivial, as static $\mathcal{ESB}$ adds nothing but terms of sort action to $\mathcal{BO}$. The catch is that action standard names are not atomic but formed with an action function symbol A. In general, taking all ground terms as standard names is problematic: if all we knew was $\phi = P(\#1) \wedge \forall x (R(x) \supset R(g(g(x))))$, which intuitively means that P is the set of odd natural numbers, then $\text{RES}[\![P(x), \phi]\!]$ would have to enumerate all odd "numbers:" $\#1$, $g(g(\#1))$, $g(g(g(g(\#1))))$, and so on. Levesque (1984b) concludes that such standard names would be too expressive for the representation theorem.

Our rescue is that action function symbols may only take terms of sort object, which restricts the nesting of standard names at the first level already. Action standard names are hence of the form $A(n_1, \ldots, n_k)$ for object standard names n_i. We hence only need to update the $\text{RES}[\![\psi, \phi]\!]$ operator (Definition 4.8.1) to handle free action variables.

Care needs to be taken to choose the right action standard names to substitute for the free action variable. For example, a relevant action standard name might not occur directly but be "hidden," as in $\exists x(a = A(x) \wedge x = \#1)$ as opposed to $a = A(\#1)$. Clearly this formula is true for certain values of a, but we can find out so only by trying $A(\#1)$ for a. So at least we need to consider all action standard names that can be formed from the action function symbols and object standard names that occur in ϕ or ψ. As another example, consider $\exists x(a = A(x) \wedge x \neq \#1)$. Again this sentence is true for appropriate a, but we can neither find that out by substituting $A(\#1)$ for a nor with some arbitrary other name A'. Instead, we need to test $A(n)$ for some $n \neq \#1$. So the trick is to take the action function symbols from ϕ and ψ and close them also under some names that do not occur in ϕ or ψ, such as $A(\#2)$ in this example.

While the definition of RES below admittedly seems complex at first, it follows this rather simple idea. We also provide an example to illustrate it below.

Definition 5.9.1 Let ϕ be an objective sentence and ψ be an objective formula. Then $\mathrm{RES}[\![\psi, \phi]\!]$ is defined as follows:

- if ψ has no free variables, then

$$\mathrm{RES}[\![\psi, \phi]\!] = \begin{cases} \text{TRUE} & \text{if} \models (\phi \supset \psi); \\ \text{FALSE} & \text{otherwise;} \end{cases}$$

- if y is a free object variable in ψ and
  - $\mathcal{N}$ contains the object standard names occurring in ϕ or ψ,
  - n' is a new object standard name not in $\mathcal{N}$,

  then

$$\mathrm{RES}[\![\psi, \phi]\!] = \bigvee_{n \in \mathcal{N}} \left((y = n) \wedge \mathrm{RES}[\![(\psi^y_n), \phi]\!]\right) \vee \bigwedge_{n \in \mathcal{N}} \left((y \neq n) \wedge \mathrm{RES}[\![(\psi^y_{n'}), \phi]\!]^{n'}_y\right);$$

- if a is a free action variable in ψ and there is no free object variable in ψ,
  - $\mathcal{A}$ contains the action function symbols occurring in ϕ or ψ,
  - K is the maximal arity of the symbols in $\mathcal{A}$,
  - $\mathcal{N}$ contains the object standard names occurring in ϕ and ψ,
  - $\mathcal{N}' = \{n'_1, \ldots, n'_K \mid n'_i \notin \mathcal{N}\}$ contains K new object standard names that neither occur in ϕ nor in ψ,

- $\mathcal{M} = \{A(n_1, \ldots, n_k) \mid A \in \mathcal{A} \text{ and } n_i \in \mathcal{N}\}$ is the set of action standard names formed from the action symbols $\mathcal{A}$ and object standard names $\mathcal{N}$,
- $\mathcal{M}' = \{A(n_1, \ldots, n_k) \mid A \in \mathcal{A} \text{ and } n_i \in \mathcal{N} \cup \mathcal{N}'\} \setminus \mathcal{M}$ is the set of action standard names formed from the action symbols $\mathcal{A}$ that and the action standard names $\mathcal{N} \cup \mathcal{N}'$ but have at least one argument from $\mathcal{N}'$ that does not occur in ϕ or ψ,
- $A' \notin \mathcal{A}$ is a new action constant that neither occurs in ϕ nor in ψ,
- for $n' = A(n_1, \ldots, n_k) \in \mathcal{M}'$ we let $(n')^{n'_1 \ldots n'_K}_{y_1 \ldots y_K}$ denote the result of replacing in $A(n_1, \ldots, n_k)$ every occurrence of an object standard name $n'_i \in \mathcal{N}'$ (which may occur zero or more times in $n_1, \ldots, n_k$) with the variable y_i,

then

$$\begin{aligned}
\mathrm{RES}[\![\psi, \phi]\!] = {} & \bigvee_{n \in \mathcal{M}} \big((a = n) \land \mathrm{RES}[\![(\psi^a_n), \phi]\!]\big) \lor \\
& \bigvee_{n' \in \mathcal{M}'} \exists y_1 \ldots \exists y_K \big((a = (n')^{n'_1 \ldots n'_K}_{y_1 \ldots y_K}) \land \\
& \qquad \bigwedge_{1 \le i \le K, n'' \in \mathcal{N}} (y_i \neq n'') \land \bigwedge_{1 \le i < j \le K} (y_i \neq y_j) \land \mathrm{RES}[\![(\psi^a_{n'}), \phi]\!]^{n'}_a\big) \lor \\
& \bigwedge_{A \in \mathcal{A}} \forall y_1 \ldots \forall y_k \big((a \neq A(y_1, \ldots, y_k)) \land \mathrm{RES}[\![(\psi^a_{A'}), \phi]\!]^{A'}_a\big).
\end{aligned}$$

Before we explain the new case for free action variables, note that the other two cases are the same as in the representation theorem for $\mathcal{BO}$. Like in Definition 4.8.1, the case for a free object variable y tries all names from ϕ and ψ plus a new one, n'. Notice that n' is eventually replaced by y again, so that no new standard name is introduced by $\mathrm{RES}[\![\psi, \phi]\!]$.

Now consider the case when ψ has a free action variable a and there is no free object variable in ψ. First, we remark that the condition that there shall be no free object variable in ψ is merely a tie-breaker so that RES is well-defined; we could equally well require analogously in the case for free object variables that there shall be no free action variable in ψ.

The first and the last of the three large disjuncts in RES for action variables are similar to the case for free object variables. $\mathcal{M}$ contains all action standard names that can be formed from the symbols in ϕ and ψ, it thus corresponds to the set $\mathcal{N}$ in the case for free object variables; all these names are tested explicitly. The new action constant A' corresponds to n' in the case for an objective variable; it is tested and then again replaced by a to avoid a new standard name in the result.

New is the middle disjunct of RES, that is, the disjunction over $n' \in \mathcal{M}'$. It covers the action names which are partially but not fully formed from symbols in ϕ or ψ. Let us consider an example to illustrate how it works. Suppose ϕ or ψ mention a single binary action function symbol $A \in \mathcal{A}$ and one object name $n_1 \in \mathcal{N}$. Then $K = 2$, so $\mathcal{N}' = \{n'_1, n'_2\}$; intuitively these object names represent all names which do not occur in ϕ or ψ. Let us consider the subformula where $n' \in \mathcal{M}'$ has the value $A(n_1, n'_2)$. After doing the substitution, the subformula begins with $\exists y_1 \exists y_2((a = A(n_1, y_2)) \wedge (y_1 \neq n_1) \wedge (y_2 \neq n_1) \wedge (y_1 \neq y_2) \wedge \mathrm{RES}[\![\psi^a_{A(n_1,n'_2)}, \phi]\!]^{A(n_1,n'_2)}_a)$. Imagine $A(n_1, n''_2)$ where $n''_2 \notin \mathcal{N}$ substituted for a: then the formula is equivalent to $\mathrm{RES}[\![\psi^a_{A(n_1,n'_2)}, \phi]\!]^{A(n_1,n'_2)}_{A(n_1,n''_2)}$, that is, $A(n_1, n''_2)$ is represented by $A(n_1, n'_2)$. Intuitively this works because ψ cannot distinguish between $A(n_1, n'_2)$ and $A(n_1, n''_2)$, as $n'_1, n''_2 \notin \mathcal{N}$.

Besides this extended $\mathrm{RES}[\![\psi, \phi]\!]$ operator we also need to reflect forgetting of the extended only-believing in our reduction. To this end, we add second-order quantifiers to the language.

Definition 5.9.2 The set of well-formed formulas is the least set formed from the rules from Definitions 5.2.1 and 5.6.1 and

- $\exists \mathcal{S}\, \alpha$ is a formula, where $\mathcal{S}$ is a finite set of object function or predicate symbols and α is a formula.

Semantically these quantifiers are easily interpreted with $\approx_{\mathcal{S}}$ from Definition 5.6.2.

Definition 5.9.3 The semantics of existential second-order variables is defined as follows:

$\mathcal{ESB}$12. $\vec{e}, w \models \exists \mathcal{S}\, \alpha$ iff for some w', $\vec{e}, w' \models \alpha$ and $w \approx_{\mathcal{S}} w'$.

We remark that this simple semantics of second-order quantifiers does not behave well with quantifying-in. Lakemeyer and Levesque (2009, 2011) extend the semantics with another parameter for variable maps where the extension of second-order variables is memorized. As we use them for objective reasoning only, the simple semantics is sufficient here.

An objective representation in $\mathcal{ESB}$ is defined analogously to objective representation in $\mathcal{BO}$ (Definition 4.8.2).

Definition 5.9.4 Let $\Gamma = \{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$. Let $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$. An *objective representation* $\vec{\gamma}$ of $\mathbf{O}_{\mathcal{S}}\Gamma$ is an infinite sequence of objective sentences γ_p such that $\gamma_p = \{w \mid w \models \exists \mathcal{S} \gamma_p\}$ for all $p \in \mathbb{P}$. We write $\langle \gamma_1, \ldots, \gamma_q \rangle$ if $\gamma_q = \gamma_p$ for all $p \geq q$.

The following lemma corresponds to a similar result for $\mathcal{BO}$ (Lemma 4.8.3).

Lemma 5.9.5 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be objective and* $\vec{\gamma} = \langle \gamma_1, \ldots, \gamma_{m+1} \rangle$ *be defined by* $\gamma_p = \bigwedge_{i:\gamma_0 \models \neg\phi_i, \ldots, \gamma_{p-1} \models \neg\phi_i} (\phi_i \supset \psi_i)$. *Then* $\vec{\gamma}' = \langle \exists \mathcal{S} \gamma_1, \ldots, \exists \mathcal{S} \gamma_{m+1} \rangle$ *is an objective representation of* $\mathbf{O}_\mathcal{S}\Gamma$, *and for every other objective representation* $\vec{\gamma}''$, *for all* $p \in \mathbb{P}$, $\models \gamma'_p \equiv \gamma''_p$.

Proof. As argued for Lemma 4.8.3, $\vec{\gamma}$ is an objective representation of $\mathbf{O}\Gamma$. Let $\vec{e} \models \mathbf{O}\Gamma$ and $\vec{e}' \models \mathbf{O}_\mathcal{S}\Gamma$. For all $p \in \mathbb{P}$ and for all w, $w \models \gamma'_p$ iff $w' \models \gamma_p$ for some $w' \approx_\mathcal{S} w$ iff $w' \in e_p$ for some $w' \approx_\mathcal{S} w$ iff $w \in (\vec{e}_\mathcal{S})_p$ iff $w \in e_p$. □

We can now reduce belief projection problems $\mathbf{O}\Gamma \models [t_1] \ldots [t_k]\mathbf{B}(\alpha \Rightarrow \beta)$ to static, non-modal (even though in general second-order) logic, provided that Γ is objective. The belief entailment obtained after eliminating the actions by means of regression or progression is solved the same way as in $\mathcal{BO}$ (Definition 4.8.4). By Lemma 5.9.5, we only need to consider a single objective representation of $\mathbf{O}_\mathcal{S}\Gamma$ which is moreover finite. Then the very same procedure as in $\mathcal{BO}$ works to eliminate conditional belief operators from the query.

Definition 5.9.6 A formula is *belief-static* if for every non-static subformula $[t]\alpha$ or $\Box\alpha$, α is objective. Let α be a belief-static formula without $\mathbf{O}$ and let $\vec{\gamma} = \langle \gamma_1, \ldots, \gamma_q \rangle$ be objective sentences. Then $\|\alpha\|_{\vec{\gamma}}$ is defined inductively:

- $\|\alpha\|_{\vec{\gamma}} = \alpha$ if α is an objective formula;
- $\|\neg\alpha\|_{\vec{\gamma}} = \neg\|\alpha\|_{\vec{\gamma}}$;
- $\|(\alpha_1 \vee \alpha_2)\|_{\vec{\gamma}} = (\|\alpha_1\|_{\vec{\gamma}} \vee \|\alpha_2\|_{\vec{\gamma}})$;
- $\|\exists x\, \alpha\|_{\vec{\gamma}} = \exists x\, \|\alpha\|_{\vec{\gamma}}$;
- $\|\mathbf{B}(\alpha \Rightarrow \beta)\|_{\vec{\gamma}} = \bigwedge_{p=1}^{q} \big(\big(\bigwedge_{p'=1}^{p-1} \mathrm{RES}[\![\|\neg\alpha\|_{\vec{\gamma}}, \gamma_{p'}]\!] \big) \supset \mathrm{RES}[\![\|(\alpha \supset \beta)\|_{\vec{\gamma}}, \gamma_p]\!] \big)$.

We use $\|\alpha\|_{\mathbf{O}_\mathcal{S}\Gamma}$ as an abbreviation for $\|\alpha\|_{\vec{\gamma}}$ where $\vec{\gamma}$ is the objective representation of $\mathbf{O}_\mathcal{S}\Gamma$ from Lemma 5.9.5.

The following theorem generalizes the representation theorem from $\mathcal{BO}$ (Theorem 4.8.5) to $\mathcal{ESB}$.

Theorem 5.9.7 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be objective and* α *be belief-static without* $\mathbf{O}$. *Then* $\mathbf{O}_\mathcal{S}\Gamma \models \alpha$ *iff* $\models \|\alpha\|_{\mathbf{O}_\mathcal{S}\Gamma}$.

The proof can be found in Appendix B.5. As a special case, we show the representation theorem for $\mathcal{BO}$ there (Theorem 4.8.5).

We could lift the belief-static restriction on α, which prohibits beliefs within the scope of actions, by a regression-like procedure that transforms the α in belief-static form by means of Theorems 5.5.4 and 5.5.5. Unlike regression, this procedure may leave fluents unchanged and is thus does not require on a basic action theory. We skip such an extended $\|\cdot\|$ operator here, and instead merely show that the representation theorem works well with regression.

Corollary 5.9.8 *Let* $\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}$ *be an* $\mathcal{S}$*-free basic action theory and let* α *be a regressable sentence. Then* $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models \alpha$ *iff* $\models \|\mathcal{R}[\alpha]\|_{\mathbf{O}_{\mathcal{S}}\Sigma_{\text{bel}}}$.

As an alternative to the regression-like procedure sketched above, we could update the objective representation to reflect the epistemic effect of actions. We saw however in Section 5.7 how involved this can get with belief revision, and we do not investigate this any further here. It is however easy to see that the representation theorem goes along equally well with progression as with regression.

Corollary 5.9.9 *Let* $\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}$ *be a basic action theory,* $\mathcal{S}'$ *be the symbols newly introduced by* $\Sigma_{\text{bel}} \gg n$*, and let* α *be a belief-static sentence without* $\mathbf{O}$.
Then $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models [n]\alpha$ *iff* $\models \|\alpha\|_{\mathbf{O}_{\mathcal{S}\cup\mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)}$.

These corollaries are also shown in Appendix B.5. We conclude this section with an example that, unlike Example 4.8.6, also involves nested beliefs and quantifying-in.

Example 5.9.10 Let us illustrate the representation theorem with query Q3 from Example 5.4.2. With progression, the problem reduces to the static belief entailment

$$\mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg \text{dropbox} \gg \text{clink}) \models \mathbf{B}(\text{InBox}(\text{gift}) \land \text{Broken}(\text{gift}) \land \neg\exists y\, \mathbf{B}\text{gift} = y).$$

In Example 5.8.4, we already determined the progression by dropbox and clink, that is, $\mathbf{O}_{\{R, R_{\text{InBox}}, R_{\text{Broken}}\}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg \text{dropbox} \gg \text{clink})$, and its model $\vec{e}'_{\{R, R_{\text{InBox}}, R_{\text{Broken}}\}}$. Reading off from $\vec{e}'$ in Example 5.8.4 and simplifying it a little to eliminate the second-order variables (which is easy here) yields the objective representation

$$\begin{aligned}
\gamma_1 &= \Sigma_{\text{dyn}} \land \forall y (\text{InBox}(y) \equiv y = \text{gift}) \land \text{Broken}(\text{gift}) \land \text{Fragile}(\text{gift});\\
\gamma_2 &= \Sigma_{\text{dyn}} \land \forall y (\text{InBox}(y) \supset y = \text{gift} \land \text{Broken}(\text{gift}) \land \text{Fragile}(\text{gift}));\\
\gamma_3 &= \Sigma_{\text{dyn}} \land \forall y (\text{InBox}(y) \supset y = \text{gift} \land (\text{Broken}(y) \equiv \text{Fragile}(y)));\\
\gamma_4 &= \Sigma_{\text{dyn}} \land \forall y (\text{InBox}(y) \land \text{Fragile}(y) \supset \text{Broken}(y)).
\end{aligned}$$

By Theorem 5.9.7, to prove the entailment problem it suffices to check the validity of $\|\mathbf{B}(\text{InBox}(\text{gift}) \land \text{Broken}(\text{gift}) \land \neg\exists y\, \mathbf{B}\text{gift} = y)\|_{\vec{\gamma}}$. The $\|\alpha\|_{\vec{\gamma}}$ operator works its way from the inside to the outside, so we begin with $||\mathbf{B}\text{gift} = y||_{\vec{\gamma}}$. This expands to $\bigwedge_{p=1}^{4}((\bigwedge_{p'=1}^{p-1} \text{RES}[\![||\neg\textsc{true}||_{\vec{\gamma}}, \gamma_{p'}]\!]) \supset \text{RES}[\![||(\textsc{true} \supset \text{gift} = y)||_{\vec{\gamma}}, \gamma_p]\!])$, and since $\not\models \gamma_{p'} \supset \neg\textsc{true}$ for all p', this is equivalent to $\text{RES}[\![\text{gift} = y, \gamma_1]\!]$. Neither γ_1 nor gift $= y$ contain any object standard name, so we only need to plug in one arbitrary standard name for y, say $^{\#}7$. Then $\text{RES}[\![\text{gift} = y, \gamma_1]\!]$ expands to $\text{RES}[\![\text{gift} = {}^{\#}7, \gamma_1]\!]_y^{\#7}$, and since $\not\models \gamma_1 \supset \text{gift} = {}^{\#}7$, this finally expands to is FALSE.

The next step is to determine $||\mathbf{B}(\text{InBox(gift)} \land \text{Broken(gift)} \land \neg\exists y\, \textsc{false})||_{\vec{\gamma}}$. As in the previous step, $\not\models \gamma_{p'} \supset \neg\textsc{true}$ for all p', that is, the first sphere is consistent, the formula boils down to $\text{RES}[\![\text{InBox(gift)} \land \text{Broken(gift)}, \gamma_1]\!]$, which is TRUE because $\models \gamma_1 \supset \text{InBox(gift)} \land \text{Broken(gift)}$. That is, $\models ||\mathbf{B}(\text{InBox(gift)} \land \text{Broken(gift)} \land \neg\exists y\, \mathbf{B}\text{gift} = y)||_{\vec{\gamma}}$ was eventually reduced to $\models \textsc{true}$. The query is thus proved.

5.10 Belief revision postulates

In this section we relate $\mathcal{ESB}$ to the most well known accounts of belief change: the postulates for single revision by Alchourrón, Gärdenfors, and Makinson (1985) and Gärdenfors (1988), often referred to by their initials AGM; the postulates for iterated revision by Darwiche and Pearl (1997); and the alternative proposal by Nayak, Pagnucco, and Peppas (2003). We will see that the AGM postulates are satisfied and a slight weakening of the Darwiche–Pearl postulates hold. For strong revision furthermore all but the first Nayak–Pagnucco–Peppas postulate are satisfied as well. The divergences from these postulate systems only concern the special case of revision by an inconsistent formula: our semantics provides no escape from the inconsistent epistemic state $\langle\{\}\rangle$ once it is reached.

We prove the postulates for *semantic* revision by objective sentences; the corresponding results for *theory* revision follow by the theorems from Section 5.7. Our translation of the postulates to $\mathcal{ESB}$ is similar to the one by Shapiro et al. (2011). Perhaps the only notable translation is that belief expansion is modelled as material implication: $\vec{e} \models \mathbf{B}(\phi \supset \psi)$ represents that ψ is in the belief set after it is expanded with ϕ. For the remainder of this section let ϕ, ψ, υ be objective sentences.

We begin with the translation of the original belief revision postulates by Alchourrón, Gärdenfors, and Makinson (1985) to our formalism.

Definition 5.10.1 The Alchourrón–Makinson–Gärdenfors postulates in $\mathcal{ESB}$ are as follows.

AGM1. If $\vec{e} * \phi \models \mathbf{B}\psi$ and $\vec{e} * \phi \models \mathbf{B}(\psi \supset \upsilon)$, then $\vec{e} * \phi \models \mathbf{B}\upsilon$.

AGM2. $\vec{e} * \phi \models \mathbf{B}\phi$.

AGM3. If $\vec{e} * \phi \models \mathbf{B}\upsilon$, then $\vec{e} \models \mathbf{B}(\phi \supset \upsilon)$.

AGM4. If $\vec{e} \not\models \mathbf{B}\neg\phi$ and $\vec{e} \models \mathbf{B}(\phi \supset \upsilon)$, then $\vec{e} * \phi \models \mathbf{B}\upsilon$.

AGM5. If $\vec{e} \not\models \mathbf{K}\neg\phi$, then $\vec{e} * \phi \not\models \mathbf{B}\textsc{false}$.

AGM6. If $\vec{e} \models \mathbf{K}(\phi \equiv \psi)$, then $\vec{e} * \phi \models \mathbf{B}\upsilon$ iff $\vec{e} * \psi \models \mathbf{B}\upsilon$ for every υ.

AGM7. If $\vec{e} * (\phi \land \psi) \models \mathbf{B}\upsilon$, then $\vec{e} * \phi \models \mathbf{B}(\psi \supset \upsilon)$.

AGM8. If $\vec{e} * \phi \not\models \mathbf{B}\neg\psi$ and $\vec{e} * \phi \models \mathbf{B}(\psi \supset \upsilon)$, then $\vec{e} * (\phi \land \psi) \models \mathbf{B}\upsilon$.

Theorem 5.10.2 *The Alchourrón–Makinson–Gärdenfors postulates are satisfied.*

Proof. We suppose $e_p \neq \{\}$ for some $p \in \mathbb{P}$, for otherwise the postulates hold trivially as $(\vec{e} * \delta)_p = \{\}$ and thus $\vec{e} * \delta \models \mathbf{B}\textsc{false}$. Since the postulates refer only to a single revision and $(\vec{e} *_{\mathrm{w}} \delta)_1 = (\vec{e} *_{\mathrm{s}} \delta)_1$ by Lemma 5.3.7, the proof does not need to distinguish between weak and strong revision.
AGM1. Follows from Property (iv) of Theorem 5.3.15.

AGM2. If there is no ϕ-world, $(\vec{e} * \phi)_p = \{\}$ for all $p \in \mathbb{P}$; else for all $w \in (\vec{e} * \phi)_1 \neq \{\}$, $w \models \phi$. In either case, $\vec{e} * \phi \models \mathbf{B}\phi$.

AGM3. Let $\vec{e} * \phi \models \mathbf{B}\upsilon$. Suppose $\vec{e} \not\models \mathbf{B}(\phi \supset \upsilon)$. Then for some $w \in e_{\lfloor \vec{e} \mid \textsc{true} \rfloor}$, $w \models \phi \land \neg\upsilon$. Then $w \in (\vec{e} * \phi)_1$, and therefore $\vec{e} * \phi \not\models \mathbf{B}\upsilon$, which contradicts the assumption. Thus $\vec{e} \models \mathbf{B}(\phi \supset \upsilon)$.

AGM4. Let $\vec{e} \not\models \mathbf{B}\neg\phi$ and $\vec{e} \models \mathbf{B}(\phi \supset \upsilon)$. Then for some $w \in e_{\lfloor \vec{e} \rfloor}$, $w \models \phi$, and for all $w \in e_{\lfloor \vec{e} \rfloor}$, $w \models \phi \supset \upsilon$. Therefore $(\vec{e} * \phi)_1 \subseteq e_p$, and for all $w \in (\vec{e} * \phi)_1$, $w \models \phi \land \upsilon$, so $\vec{e} * \phi \models \mathbf{B}\upsilon$.

AGM5. Let $\vec{e} \not\models \mathbf{K}\neg\phi$. Then for some $w \in e_{\lfloor \vec{e} \rfloor}$, $w \models \phi$. Hence $(\vec{e} * \phi)_1 \neq \{\}$, so $\vec{e} * \phi \not\models \mathbf{B}\textsc{false}$.

AGM6. Let $\vec{e} \models \mathbf{K}(\phi \equiv \psi)$. Then $w \models \phi$ iff $w \models \psi$ for all $p \in \mathbb{P}$ and $w \in e_p$. Thus $\vec{e} * \phi = \vec{e} * \psi$.

AGM7. If there is no ϕ-world, $\vec{e} * \phi \models \mathbf{B}(\psi \supset \upsilon)$ holds trivially. Otherwise, consider $w \in (\vec{e} * \phi)_1$ with $w \models \psi$. Then $w \in (\vec{e} * (\phi \land \psi))_1$ and by assumption, $w \models \upsilon$. Hence $\vec{e} * \phi \models \mathbf{B}(\psi \supset \upsilon)$.

AGM8. Let $\vec{e} * \phi \not\models \mathbf{B}\neg\psi$ and $\vec{e} * \phi \models \mathbf{B}(\psi \supset \upsilon)$. Then for some $w \in (\vec{e} * \phi)_1$, $w \models \psi$. Therefore $(\vec{e} * (\phi \land \psi))_1 \subseteq (\vec{e} * \phi)_1$, and for all $w \in (\vec{e} * (\phi \land \psi))_1$, $w \models \psi \land (\psi \supset \upsilon)$, and so $w \models \upsilon$. Hence $\vec{e} * (\phi \land \psi) \models \mathbf{B}\upsilon$. ◻

The most popular postulate system iterated revision is due to Darwiche and Pearl (1997). We slightly restrict them here by weakening the second postulate.

Definition 5.10.3 The restricted Darwiche–Pearl postulates in $\mathcal{ESB}$ are as follows.

DP1. If $\vec{e} \models \mathbf{K}(\psi \supset \phi)$, then $(\vec{e} * \phi) * \psi \models \mathbf{B}\upsilon$ iff $\vec{e} * \psi \models \mathbf{B}\upsilon$.

DP2. If $\vec{e} \models \mathbf{K}(\psi \supset \neg\phi)$ and $\vec{e} \models \mathbf{K}\neg\phi \supset \mathbf{K}\neg\psi$, then $(\vec{e} * \phi) * \psi \models \mathbf{B}\upsilon$ iff $\vec{e} * \psi \models \mathbf{B}\upsilon$ for every υ.

DP3. If $\vec{e} * \psi \models \mathbf{B}\phi$, then $(\vec{e} * \phi) * \psi \models \mathbf{B}\phi$.

DP4. If $\vec{e} * \psi \not\models \mathbf{B}\neg\phi$, then $(\vec{e} * \phi) * \psi \not\models \mathbf{B}\neg\phi$.

The restriction of DP2 concerns the special case of revision by an inconsistent formula. Since $\mathcal{ESB}$ provides no escape from the empty epistemic state, DP2 holds in case the first revision is by an inconsistent formula only if the second revision is by an inconsistent formula as well. We hence require $\vec{e} \models \mathbf{K}\neg\phi \supset \mathbf{K}\neg\psi$ in our variant of DP2. We remark that the restricted postulate is still slightly stronger than NPP4 (see below).

Theorem 5.10.4 *The restricted Darwiche–Pearl postulates are satisfied.*

Proof. DP1. Let $\vec{e} \models \mathbf{K}(\psi \supset \phi)$. Then $(\vec{e}\,|\,\psi)_p \subseteq (\vec{e}\,|\,\phi)_p$ for all $p \in \mathbb{P}$ (*). Therefore $w \in ((\vec{e} * \phi) * \psi)_1$ iff $\lfloor \vec{e} * \phi\,|\,\psi \rfloor \neq \infty$ and $w \in (\vec{e} * \phi\,|\,\psi)_{\lfloor \vec{e} * \phi | \psi \rfloor}$ iff (by (*)) $\lfloor \vec{e}\,|\,\psi \rfloor \neq \infty$ and $w \in (\vec{e}\,|\,\psi)_{\lfloor \vec{e} | \psi \rfloor}$ iff $w \in (\vec{e} * \psi)_1$.

DP2. The proof is very similar to DP1. Let $\vec{e} \models \mathbf{K}(\psi \supset \neg\phi)$ and $\vec{e} \models \mathbf{K}\neg\phi \supset \mathbf{K}\neg\psi$. If $\lfloor \vec{e}\,|\,\psi \rfloor = \infty$, then $\lfloor \vec{e} * \phi\,|\,\psi \rfloor = \infty$, and therefore $\vec{e} * \psi = (\vec{e} * \phi) * \psi = \langle\{\}\rangle$, so the postulate holds. Now suppose $\lfloor \vec{e}\,|\,\psi \rfloor \neq \infty$. By the second assumption, $\vec{e} \not\models \mathbf{K}\neg\phi$, so $\lfloor \vec{e}\,|\,\phi \rfloor \neq \infty$. By the first assumption, $w \not\models \psi$ for all $w \in (\vec{e}\,|\,\phi)_p$ for all $p \in \mathbb{P}$. Thus $(\vec{e}\,|\,\phi)_p \cap (\vec{e}\,|\,\psi)_p = \{\}$ (*), so neither weak nor strong revision by ϕ affects the relative order of the ψ-worlds. Hence $w \in ((\vec{e} * \phi) * \psi)_1$ iff $w \in (\vec{e} * \phi\,|\,\psi)_{\lfloor \vec{e} * \phi | \psi \rfloor}$ iff (by (*)) $w \in (\vec{e}\,|\,\psi)_{\lfloor \vec{e} | \psi \rfloor}$ iff $w \in (\vec{e} * \psi)_1$.

DP3. Let $\vec{e} * \psi \models \mathbf{B}\phi$. If $\lfloor \vec{e}\,|\,\phi \rfloor = \infty$, then $\vec{e} * \phi = (\vec{e} * \phi) * \psi = \langle\{\}\rangle$, so the postulate holds. If $\lfloor \vec{e}\,|\,\psi \rfloor = \infty$, then $\lfloor \vec{e} * \phi\,|\,\psi \rfloor = \infty$, and therefore $(\vec{e} * \phi) * \psi = \langle\{\}\rangle$, so the postulate holds. Now suppose $\lfloor \vec{e}\,|\,\phi \rfloor \neq \infty$ and $\lfloor \vec{e}\,|\,\psi \rfloor \neq \infty$. By assumption, $(\vec{e}\,|\,\psi)_{\lfloor \vec{e} | \psi \rfloor} \subseteq (\vec{e}\,|\,\phi)_{\lfloor \vec{e} | \psi \rfloor}$. Therefore the most-plausible ψ-worlds remain most plausible after weak or strong revision by ϕ, so $(\vec{e} * \phi\,|\,\psi)_{\lfloor \vec{e} * \phi | \psi \rfloor} \subseteq (\vec{e} * \phi\,|\,\phi)_{\lfloor \vec{e} * \phi | \psi \rfloor}$. Thus, if $w \in ((\vec{e} * \phi) * \psi)_1$, then $w \in (\vec{e} * \phi\,|\,\psi)_{\lfloor \vec{e} * \phi | \psi \rfloor}$, and $w \in (\vec{e} * \phi\,|\,\phi)_{\lfloor \vec{e} * \phi | \psi \rfloor}$, so $w \models \phi$.

DP4. Let $\vec{e} * \psi \not\models \mathbf{B}\neg\phi$. Then $\lfloor \vec{e} * \psi\,|\,\phi \rfloor \neq \infty$, and so $\lfloor \vec{e}\,|\,\phi \rfloor \neq \infty$ and $\lfloor \vec{e}\,|\,\psi \rfloor \neq \infty$. By assumption, for some $(\vec{e}\,|\,\psi \wedge \phi)_{\lfloor \vec{e} | \psi \rfloor} \neq \{\}$. The ϕ-worlds among the most-plausible ψ-worlds remain most plausible after weak or strong revision by ϕ, so we have $(\vec{e} * \phi \mid \psi \wedge \phi)_{\lfloor \vec{e} * \phi \mid \psi \rfloor} \neq \{\}$. Hence there is some $w \in (\vec{e} * \phi \mid \psi \wedge \phi)_{\lfloor \vec{e} * \phi \mid \psi \rfloor}$, and therefore also $w \in (\vec{e} * \phi \mid \psi)_{\lfloor \vec{e} * \phi \mid \psi \rfloor}$. Thus $w \models \phi$ for some $w \in ((\vec{e} * \phi) * \psi)_1$. □

Finally, let us turn to the alternative proposal for iterated revision by Nayak, Pagnucco, and Peppas (2003).

Definition 5.10.5 The Nayak–Pagnucco–Peppas postulates in $\mathcal{ESB}$ are as follows.

NPP1. If $\vec{e} \models \mathbf{K}\text{FALSE}$, then $\vec{e} * \phi \models \mathbf{O}\{\text{TRUE} \Rightarrow \phi\}$.

NPP2. AGM1–AGM6 hold.

NPP3. If $\vec{e} \not\models \mathbf{K}\neg(\phi \wedge \psi)$ then $(\vec{e} * \phi) * \psi \models \mathbf{B}\upsilon$ iff $\vec{e} * (\phi \wedge \psi) \models \mathbf{B}\upsilon$ for every υ.

NPP4. If $\vec{e} \models \mathbf{K}(\psi \supset \neg\phi)$ and $\vec{e} \not\models \mathbf{K}\neg\phi$, then $(\vec{e} * \phi) * \psi \models \mathbf{B}\upsilon$ iff $\vec{e} * \psi \models \mathbf{B}\upsilon$.

The Nayak–Pagnucco–Peppas postulates hold with two exceptions. For one thing, the absurdity postulate NPP1 does not hold. NPP1 allows to recover from an inconsistent revision: it says that after revising an inconsistent state by ϕ, ϕ shall be all that is believed. In $\mathcal{ESB}$, NPP1 would be counterintuitive because we would lose any indefeasible *knowledge* we might have had already before reaching the inconsistent $\vec{e}$ (such as the dynamic axioms of a basic action theory). Avoiding this would require additional book keeping, which seems like a lot of effort for relatively little gain.

For another, the conjunction postulate NPP3 only holds for strong revision. NPP3 for weak revision is not satisfied because, unlike strong revision, weak revision by ϕ does not preserve the relative ordering among the ϕ worlds. Hence, after revision by ϕ the most-plausible ψ-worlds might actually not satisfy not ϕ.

Theorem 5.10.6 *The Nayak–Pagnucco–Peppas postulates NPP2, NPP3 for strong revision, and NPP4 are satisfied.*

Proof. We only need to prove NPP3, as we have shown the AGM1–AGM6 in Theorem 5.10.2 already and NPP4 is a special case of DP2. Suppose $w \in ((\vec{e} *_s \phi) *_s \psi)_1$. Then $w \in (\vec{e} *_s \phi)_{\lfloor \vec{e} *_s \phi \mid \psi \rfloor}$. By assumption, $\lfloor \vec{e} *_s \phi \mid \psi \rfloor < \lfloor \vec{e} *_s \phi \mid \neg\phi \rfloor$, so $w \models \phi$ and, since the revision by ϕ did not affect the relative ordering of the $(\phi \wedge \psi)$-worlds, also $w \in e_{\lfloor \vec{e} \mid \phi \wedge \psi \rfloor}$. Thus $w \in (\vec{e} *_s \phi \wedge \psi)_1$. Conversely, suppose $w \in (\vec{e} *_s \phi \wedge \psi)_1$. Then $w \in e_{\lfloor \vec{e} \mid \phi \wedge \psi \rfloor}$. Therefore $w \in (\vec{e} *_s \phi)_{\lfloor (\vec{e} *_s \phi) \mid \psi \rfloor}$, and thus $w \in ((\vec{e} *_s \phi) *_s \psi)_1$. □

5.11 Sensing in $\mathcal{ESB}$

$\mathcal{ESB}$ uses informing as a lightweight alternative to classical sensing, since sensing cannot cope well with contradictory information as elaborated in Section 5.1. Informing is (or seems) weaker than sensing, as there is no ground truth to the information in general. Nevertheless, as we shall see now, informing is expressive enough to mimic classical sensing.

The idea is as follows. Suppose A should sense whether or not ϕ holds in the actual world. We can simulate this with a strong-revision action $A(\tilde{y})$, where $\tilde{y}$ takes a binary value to represent whether ϕ holds in the real world (for example, $\tilde{y} = {}^{\#}1$ iff ϕ is true), and $\mathrm{IF}(A(\tilde{y}))$ is defined as ϕ or $\neg\phi$ depending on the value of $\tilde{y}$. Then $A(\tilde{y})$ informs the agent about the real-world value of ϕ, and the revision promotes the worlds that accord with this value and it is thus believed.

To show how it works, let us translate an $\mathcal{ES}$ projection problem to $\mathcal{ESB}$. In $\mathcal{ES}$, the dynamic part of a basic action theory consists of successor-state axioms like in Definition 5.4.1 plus two more axioms: $\Box\mathrm{Poss}(a) \equiv \pi$ for the action precondition and a sensed-fluent axiom $\Box\mathrm{SF}(a) \equiv \varphi$, where π and φ are fluent $\mathcal{ES}$ formulas (Lakemeyer and Levesque 2011). Besides the dynamic axioms, a basic action theory contains a set of fluent sentences about the initial situation. We use Λ to denote an $\mathcal{ES}$ basic action theory; there is one for the real world and one for the agent's knowledge. The projection projection in $\mathcal{ES}$ is to decide an entailment

$$\Lambda_1 \wedge \mathbf{O}\Lambda_2 \models_{\mathcal{ES}} [A_1(\vec{t}_1)][A_2(\vec{t}_2)]\ldots[A_k(\vec{t}_k)]\mathbf{K}\alpha.$$

For simplicity, let us only consider problems where α is fluent.

Doing an action n in $\mathcal{ES}$ tells the agent that $\mathrm{Poss}(n)$ holds and what the real-world value of $\mathrm{SF}(n)$ is. Let $\mathcal{A}$ contain the action function symbols we are interested in; here, the symbols of the t_i suffice. Consider

$$\Box\mathrm{IF}(a) \equiv \bigwedge_{A\in\mathcal{A}} \forall\vec{y}\forall\tilde{y}(a = A(\vec{y},\tilde{y}) \supset \mathrm{Poss}(a) \wedge (\tilde{y} = {}^{\#}1 \equiv \mathrm{SF}(a)));$$
$$\Box\mathrm{Outcome}(a) \equiv \bigwedge_{A\in\mathcal{A}} \forall\vec{y}\forall\tilde{y}(a = A(\vec{y},\tilde{y}) \supset (\tilde{y} = {}^{\#}1 \equiv \mathrm{SF}(a))).$$

The second axiom binds $\tilde{y}$ on the real-world value of SF, which is then used in IF to revise by that value. Let Λ_i^* be like Λ_i except that every action function symbol is retrofitted with one additional dummy argument $\tilde{y}$ whose value shall have no effect. Moreover, Λ_1^* shall contain the above Outcome axiom and Λ_2^* shall contain the IF axiom.

The query in the above entailment problem can then be translated to $\mathcal{ESB}$ by iteratively replacing every subformula of the form $[A_i(\vec{t}_i)]\beta$ with the new formula

$\exists \tilde{y}_i(\mathrm{Outcome}(A_i(\vec{t}_i, \tilde{y}_i)) \wedge [A_i(\vec{t}_i, \tilde{y}_i)]\beta)$, where A_i shall be of sort strong-revision action. That way, we obtain the $\mathcal{ESB}$ entailment

$$\begin{aligned}
&\Lambda_1^* \wedge \mathbf{O}\{\neg\Lambda_2^* \Rightarrow \textsc{false}\} \models \\
&\quad \exists \tilde{y}_1(\mathrm{Outcome}(A_1(\vec{t}_1, \tilde{y}_1)) \wedge \\
&\qquad [A_1(\vec{t}_1, \tilde{y}_1)]\exists \tilde{y}_2(\mathrm{Outcome}(A_2(\vec{t}_2, \tilde{y}_2)) \wedge \\
&\qquad\qquad [A_2(\vec{t}_2, \tilde{y}_2)]\exists \tilde{y}_3(\mathrm{Outcome}(A_3(\vec{t}_3, \tilde{y}_3)) \wedge \ldots \\
&\qquad\qquad\qquad \ldots [A_{k-1}(\vec{t}_{k-1}, \tilde{y}_{k-1})]\exists \tilde{y}_k(\mathrm{Outcome}(A_k(\vec{t}_k, \tilde{y}_k)) \wedge \\
&\qquad\qquad\qquad\qquad [A_k(\vec{t}_k, \tilde{y}_k)]\mathbf{K}\alpha)\ldots)).
\end{aligned}$$

Every action $A_i(\vec{t}_i, \tilde{y}_i)$ revises by the information that $A_i(\vec{t}_i)$ senses in $\mathcal{ES}$. If these sensings are consistent, that is, $\Lambda_1 \wedge \mathbf{O}\Lambda_2 \not\models_{\mathcal{ES}} [A_1(\vec{t}_1)] \ldots [A_k(\vec{t}_k)]\mathbf{K}\textsc{false}$, then after doing the actions the possible worlds in $\mathcal{ES}$ correspond to the most-plausible worlds in $\mathcal{ESB}$ (this follows from the NPP3 postulate from Theorem 5.10.6) – knowledge after sensing in $\mathcal{ES}$ matches belief after informing in $\mathcal{ESB}$ then. If the sensings are inconsistent, $\mathcal{ES}$ ends up in the empty epistemic state, whereas $\mathcal{ESB}$ tries to avoid this by promoting less-plausible worlds to the first sphere.

This modelling is not limited to binary sensing. For instance, a sonar that senses a distance to some obstacle can be represented with an action sonar($\tilde{y}$), where $\tilde{y}$ is the sensed distance. When the basic action theory stipulates IF(sonar($\tilde{y}$)) $\equiv$ distance $= \tilde{y}$, then $\exists \tilde{y}((\mathrm{distance} = \tilde{y}) \wedge [\mathrm{sonar}(\tilde{y})]\mathbf{B}(\mathrm{distance} = \tilde{y}))$ holds, that is, the agent believes the correct distance.

As noted before already, Bacchus, Halpern, and Levesque (1999) model noisy sensing in a similar way: they use two artificial action parameters, one for the correct value and another one for the nominal sensor reading. For example, in sonar($\tilde{y}, y$), $\tilde{y}$ would represent the noisy distance reported by the sensor and y would be the actual distance to the wall. Their framework accounts for the probability of the sensor reporting $\tilde{y}$ when the real value is y.

5.12 Discussion

In this chapter we integrated the logic $\mathcal{BO}$ from Chapter 4 with actions in the spirit of Reiter's situation calculus. Just like $\mathcal{ES}$ can be seen as an offspring of $\mathcal{OL}$ and the epistemic extension of Reiter's situation calculus by Scherl and Levesque (2003), $\mathcal{ESB}$ has its roots in $\mathcal{BO}$ and an epistemic extension of the situation calculus by Shapiro et al. (2011) that supports belief change, albeit only to some extent. $\mathcal{ES}$ and $\mathcal{ESB}$

inherit only-knowing/only-believing and most of their semantics from $\mathcal{OL}$ and $\mathcal{BO}$, respectively, and adopt the idea to integrate actions with knowledge and belief from (Scherl and Levesque 2003) and (Shapiro et al. 2011).

However, $\mathcal{ESB}$ has grown way beyond the proposal by Shapiro et al. (2011). For one thing, (Shapiro et al. 2011) features no concept like only-believing, and conditionals are only intended to determine the agent's initial knowledge. It is hence quite cumbersome to specify a knowledge base. In fact, it seems like in their approach specifying a ranking of the possible situations by hand is often the easier way compared to finding the right set of (possibly negated) conditionals to characterize the initial situation. For another, the formalism by Shapiro et al. follows the classical sensing approach known from $\mathcal{ES}$ and (Scherl and Levesque 2003); they do not use belief revision techniques like we do. Instead, with every sensing they thin out the set of possible situations, and let the remaining most-plausible situations define the current belief. As a consequence, they cannot deal with contradicting information at all; belief that was discarded once cannot be reinstated as it is lost forever in their approach.

Shapiro et al. (2011) argue that in their framework does not go along well with any belief revision scheme because any modification of the plausibility ranking could lead to counterintuitive results in introspective formulas. The anomalies they mentioned do not occur in our logic because here a world's plausibility is a property of that world alone (and of the epistemic state), independent of the currently considered actual world. In contrast, introspection and quantifying-in both do work as expected in our framework.

To replace classical sensing, we devised a concept called informing. Informing is weaker than sensing, but useful to model unreliable sensors and contradictory information (Section 5.1). As argued before, sensing as in (Lakemeyer and Levesque 2011; Scherl and Levesque 2003; Shapiro et al. 2011) would be insufficient to deal with contradictory information. It is interesting that despite its simplicity informing is still expressive enough to capture the relevant part of sensing (Section 5.11).

On the semantic side, $\mathcal{ESB}$ integrates new information with the agent's beliefs by classical revision techniques, namely natural revision and lexicographic revision, which we referred to as weak and strong revision. Such revision can be matched by syntactic manipulation of the conditional knowledge base (Theorems 5.7.3 and 5.7.5). The resulting theory is very complex, though, because we employ second-order logic for the purposes of forgetting as in (Lin and Reiter 1994).

Perhaps the most important problem when reasoning about beliefs in the context of actions is the belief projection problem (Definition 5.4.1), which refers to reducing a dynamic belief entailment to a static one. We developed two solutions in this chapter.

The first one is by regression, where the query is reduced to a formula about the initial situation (Theorem 5.5.7). Extending Reiter's regression operator to conditional beliefs was straightforward thanks to two theorems about the relation between beliefs before and after an action (Theorems 5.5.4 and 5.5.5). These results can be used for belief regression similar to successor-state axioms for fluents. Similar theorems have been used by van Benthem (2007) to reduce beliefs after revision to initial beliefs.

In the conclusion of the previous chapter we already argued that our semantics of conditional belief greatly helped to generalize the representation theorem for conditional beliefs. Similarly, the simplicity of our regression operator is in large part due to our semantics of conditional belief. If we defined $\mathbf{B}(\alpha \Rightarrow \beta)$ as $\mathbf{B}\alpha$ after revision by β, like Boutilier (1993) does, Theorems 5.5.4 and 5.5.5 would have required an intricate rewriting to push the actions inside all belief conditionals.

Our second solution of the belief projection problem is by progression, where the effects of an action are applied to the knowledge base (Theorems 5.8.2 and 5.8.3). We combined our findings on the revision of a conditional knowledge base with Lin–Reiter-style progression for that purpose.

We also generalized $\mathcal{BO}$'s representation theorem to $\mathcal{ESB}$ (Theorem 5.9.7). That way, belief projection problems can be first reduced to static entailments by means of regression or progression, and then reduced to non-modal reasoning. Care needs to be taken to correctly handle free action variables (Definition 5.9.1), but otherwise the extension is straightforward. This is in large parts due to our non-standard definition of action standard names.

We emphasize that our action standard names do not cause, but actually solve a problem here. Most variants of $\mathcal{ES}$ define action standard names atomically analogously to object standard names, and require all possible worlds to agree with the actual world on the interpretation of rigid terms (this constraint is added to the $\simeq$ relation from Definition 3.10.3). It is not clear how this semantic constraint should be considered in the representation theorem for $\mathcal{ES}$. For example, consider $((n = A) \supset \mathbf{K}(n = A))$ for an action standard name n and an action constant A in classical $\mathcal{ES}$ (Lakemeyer and Levesque 2011). The sentence is valid since all reachable possible worlds map A to the same standard name as the actual world. However, $\|((n = A) \supset \mathbf{K}(n = A))\|_{\{\}}$ is just $(n = A) \supset \textsc{false}$ since $\{\} \not\models (n = A)$ in $\mathcal{ES}$. The problem does not arise in our semantics because A itself is a standard name and thus not subject to interpretation. An alternative approach due to Claßen and Lakemeyer (2006) is to syntactically restrict formulas to mention action functions only in equality expressions, which then get a special treatment by the $\|\cdot\|$ operator.

Future work on $\mathcal{ESB}$ especially concerns the questions of theory revision and progression. In this chapter we used second-order logic for both purposes. As for the progression of physical actions, several classes of actions or theories are known which do not require second-order logic (Lin and Reiter 1997; Liu and Lakemeyer 2009; Vassos, Lakemeyer, and Levesque 2008). If similar classes could be found for the knowledge base revision, these could probably be combined for first-order or even computable progression. Schwering, Lakemeyer, and Pagnucco (2015) presented a first-order account of weak (natural) revision in a slight variant of $\mathcal{ESB}$; a similar result for lexicographic revision appears to be much more difficult, though.

In our definition of basic action theories, the dynamic axioms are indefeasible knowledge. Whether this assumption can be relaxed to defeasible axioms while retaining properties like the regression theorem is an open question. Defeasible successor-state axioms could be useful to represent the usual effects of an axiom.

Another avenue of future work are different revision operators, particularly ones based on conditional ordinal functions (Spohn 1988), which is the predominant framework in belief revision today. In particular c-revisions are an interesting candidate for their connection to conditionals (Kern-Isberner 2001).

Many revision operators, however, bear the problem of exponential growth. For example, every lexicographic revision doubles the number of spheres doubles (in the worst case). By the argument from the proof of Lemma 4.5.2, the number of conditionals to represent this must grow exponentially, too. This renders operators like lexicographic revision impractical for long action sequences even if we had a first-order representation of progression. Perhaps these issues could be addressed from the perspective of limited reasoning. The next two chapters introduce effort-based limited semantics to approximate reasoning in $\mathcal{BO}$. Extending this limited logic by a limited notion of revision and progression is an interesting future challenge.

6 Limited Objective Reasoning

First-order logic is undecidable, which means no procedure can determine in general whether a formula is satisfiable or not. The practical utility of first-order logic is hence severely limited. In this chapter we introduce two non-standard, so called *limited* semantics. Limited reasoning means to restrict the inference capabilities in order to achieve decidability (for a specific class of entailment problems, at least). Typically, completeness is sacrificed in favour of soundness and decidability. While we only look at ordinary first-order logic here – that is, no beliefs or actions – this chapter lays the foundation of the limited version of $\mathcal{BO}$ to be presented in the next chapter.

More precisely, we consider formulas of $\mathcal{L}$ without function symbols here; we call the restricted language $\mathcal{L}^{-}$. The first semantics we introduce is sound but incomplete with respect to the general semantics from Definition 3.3.3; the other semantics is complete but unsound. The former is useful to determine sound inferences, while the latter allows for a sound consistency check. We will see that for propositional formulas both semantics can yield the same results as classical logic. Moreover, we devise decision procedures for a certain class of knowledge bases. We also present a normal form useful for limited reasoning.

Our work on limited reasoning is in line with the stream of research by Lakemeyer and Levesque (2002, 2013, 2014, 2016), Liu (2006), and Liu, Lakemeyer, and Levesque (2004) on the topic. The material presented in this and the following chapters is based on (Schwering and Lakemeyer 2016). The limited sound semantics is a slightly restricted version of a proposal by Lakemeyer and Levesque (2014). The long proofs are given in Appendix C.

6.1 Why incomplete and unsound reasoning matter

First-order logic is a very expressive language and hence the tool of choice for many tasks in knowledge representation (Lifschitz, Morgenstern, and Plaisted 2008). On the downside, great expressivity brings great computational demands, and in case of first-order logic even undecidability. Corollary 3.4.2 shows that validity in $\mathcal{L}$ is undecidable, and by extension this holds for $\mathcal{OL}$, $\mathcal{BO}$, $\mathcal{ES}$, and $\mathcal{ESB}$ as well. This is bad news for knowledge representation where we are typically interested in checking whether or not a certain implication is valid or not. It is hence clear that unless we are willing to give up much expressivity, unsound or incomplete reasoning is the best we can do.

On the other hand, first-order validity is semidecidable, that is, the (infinitely many) valid formulas can be enumerated. So we could try to prove a logical implication by enumerating valid formulas for a while and checking if the implication in question is among them; this procedure would be aborted if it does not come up with an affirmative answer in a certain period of time. The timeout specifies the maximum effort the reasoner may spend on trying to prove the implication. The approach is sound but obviously incomplete.

Simple as it may be, a timeout-based approach is unsatisfactory because there is no semantic justification to it; it is hard to understand why a particular implication is proved and another is not. Lakemeyer and Levesque (2013, 2014, 2016), Liu (2006), and Liu, Lakemeyer, and Levesque (2004) essentially adopt the idea of limiting the reasoning effort the system may spend on proving a formula. But instead of measuring the effort in time, they encode it as the number of case splits. Case splits are intuitive from a human perspective. For example, in Example 4.5.5 we considered the cases where the guest is a vegetarian and where the opposite is true, and in either case we showed her to be presumably not Australian. Moreover, case splits are a semantically perspicuous way to capture reasoning effort, independent from implementation details or computing power.

However, no sort of incomplete reasoning alone can suffice to determine conditional beliefs soundly. Evaluating a conditional "if α, then presumably β" requires to determine the plausibility of α first, that is, to find the most-plausible sphere consistent with α. Only then sound inference can be applied to evaluate $(\alpha \supset \beta)$ in this sphere. With limited reasoning, we can of course only approximate the plausibility of α. And to preserve soundness, we need to approximate it from above, not below: if we picked a too-plausible sphere, which is actually inconsistent with α, the material implication $(\alpha \supset \beta)$ and thus also the conditional in question could come out true even though it should not. Hence, soundly evaluating a conditional belief requires both sound consistency checks and sound inference. Consistency checking boils down to *dis*proving a formula, and to do that soundly, we need a complete (but perhaps unsound) semantics. Therefore, we believe, not only sound but also complete limited reasoning needs to be investigated, if only for the sake of conditional beliefs.

6.2 The language $\mathcal{L}^-$

The language we consider in this chapter is a stripped-down version of $\mathcal{L}$ from Definition 3.2.3.

Definition 6.2.1 The symbols of $\mathcal{L}^-$ are the same as for $\mathcal{L}$ (Definition 3.2.1) minus function symbols. The set of *terms* of $\mathcal{L}^-$ is the least set which includes all variables and standard names. The *formulas* of $\mathcal{L}^-$ is the least set such that

- $P(t_1, \ldots, t_k)$ is a formula where P is a predicate symbol and the t_i are terms;
- $(t_1 = t_2)$ is a formula where t_1 and t_2 are terms;
- $\neg\alpha$, $(\alpha \vee \beta)$, and $\exists x\, \alpha$ are formulas where α and β are formulas and x is a variable.

A *literal* is an atom $P(t_1, \ldots, t_k)$, an equality atom $(t_1 = t_2)$, or their negation. The complement $\overline{\ell}$ of a literal ℓ is defined as $\neg a$ if ℓ is an (equality) atom a, and as a if ℓ is a negated (equality) atom $\neg a$. A *clause* is a set of literals $[\ell_1, \ldots, \ell_k]$ (we use square brackets to ease readability). The empty clause is written as $[\,]$. Every non-empty clause $[\ell_1, \ldots, \ell_k]$ corresponds to the disjunction $(\ell_1 \vee \ldots \vee \ell_k)$ (with arbitrary brackets and order).

It is immediate that any formula of $\mathcal{L}^-$ is also a formula of $\mathcal{L}$. Conversely, however, $\mathcal{L}$ formulas that mention function symbols are not part of the language $\mathcal{L}^-$.

6.3 Setups, unit propagation, and subsumption

As in (Liu, Lakemeyer, and Levesque 2004) and its relatives, our semantics from this chapter are based on three concepts: setups, unit propagation, and subsumption.

Definition 6.3.1 A *setup* is a set of ground clauses.

A setup can be thought of as representing the agent's explicit knowledge. In general, it is not closed under logical deduction. This sets setups apart from other ways to express incomplete knowledge, such as sets of possible worlds, and is foundational to the decidability results. Two simple rules of inference are used to draw inferences from a setup: unit propagation and subsumption.

Definition 6.3.2 The *unit propagation* inference rule is the operation of passing from two clauses $[\ell]$ and $[\overline{\ell}] \cup c$ to the clause c.

For example, from [Aussie, Italian] and [¬Italian] we infer by unit propagation that [Aussie]. Unit propagation is related to the classical modus (tollendo) ponens inference

Table 6.1: The turnstile symbols used in this chapter.

$\models$	satisfaction and entailment in $\mathcal{L}$ (Definition 3.3.3)
$\mathrel{|}\!\overset{\circ}{\approx}$	satisfaction in sound limited semantics of $\mathcal{L}^-$ (Definition 6.4.2)
$\mathrel{|}\!\overset{\bullet}{\approx}$	satisfaction in complete limited semantics of $\mathcal{L}^-$ (Definition 6.6.3)

rule, which allows to pass from $(\alpha \vee \beta)$ and $\neg\alpha$ to β for arbitrary formulas α, β. Clearly, unit propagation is weaker than modus ponens, as it only refers to clauses and the premise even needs to be a unit clause.

Setups will be closed under unit propagation to draw easy, obvious inferences. More inferences will be obtained by augmenting setups with new unit clauses to trigger unit propagation. Besides unit propagation, subsumption allows for other immediate inferences.

Definition 6.3.3 The *subsumption* inference rule is the operation of passing from a clause c to a clause $c \cup c'$ for arbitrary c'.

For example, from [Aussie] we can infer by subsumption that [Aussie, Veggie]. Subsumption is a special case of the classical inference rule of disjunction introduction, which allows to infer $(\alpha \vee \beta)$ from α for arbitrary sentences α, β.

We are going to present two limited semantics for $\mathcal{L}^-$, denoted by $\mathrel{|}\!\overset{\circ}{\approx}$ and $\mathrel{|}\!\overset{\bullet}{\approx}$. They are sound and complete, respectively, in the following sense: given a setup s and an $\mathcal{L}^-$ formula ϕ,

- if s satisfies ϕ in the *sound* semantics $\mathrel{|}\!\overset{\circ}{\approx}$, then s classically entails ϕ, which is to say that every world that satisfies all $c \in s$ also satisfies ϕ in $\mathcal{L}$;
- if s classically entails ϕ, then s satisfies ϕ in the *complete* semantics $\mathrel{|}\!\overset{\bullet}{\approx}$.

In particular, this means that if ϕ is valid in $\mathrel{|}\!\overset{\circ}{\approx}$, then it is valid in $\mathcal{L}$, and conversely if ϕ is valid in $\mathcal{L}$, then it is valid in $\mathrel{|}\!\overset{\bullet}{\approx}$. We use $\models$ to refer to the usual, unlimited truth relation and entailment relation of $\mathcal{L}$ as introduced in Definition 3.3.3. In particular, when we write $s \models \phi$ to mean that the set of clauses s logically implies ϕ in the ordinary semantics of $\mathcal{L}^-$, that is, $w \models \phi$ for all w with $w \models c$ for all $c \in s$.

The next definition is fundamental for the semantics to come. In particular, it makes precise the notions of unit propagation and subsumption on a setup. Note the special treatment of equality. The rationale behind that is that full knowledge about equality of standard names is assumed, similarly to how it is treated in $\mathcal{L}$.

Definition 6.3.4 For a setup s, we define the following expressions:

$$\begin{aligned}
s^- &= \{c \in s \mid \text{for all } c' \subsetneq c, c' \notin s\};\\
s^+ &= \{c \mid \text{for some } c' \subseteq c, c' \in s\};\\
\mathsf{EQ} &= \{[(n = n)], [(n \neq n')] \mid \text{distinct names } n, n'\};\\
\mathsf{UP}(s) &= \text{closure of } \mathsf{EQ} \cup s \text{ under unit propagation.}
\end{aligned}$$

To ease readability, we usually write $\mathsf{UP}^-(s)$ for $\mathsf{UP}(s)^-$, and similarly $\mathsf{UP}^+(s)$ for $\mathsf{UP}(s)^+$. The following lemma states that doing unit resolution or adding or removing subsumed clauses preserves equality in $\mathcal{L}$.

Lemma 6.3.5 *For any world w and setup s, the following are equivalent:*

(i) $w \models c$ *for all* $c \in s$;

(ii) $w \models c$ *for all* $c \in s^-$;

(iii) $w \models c$ *for all* $c \in s^+$;

(iv) $w \models c$ *for all* $c \in \mathsf{UP}(s)$.

Proof. The only-if direction of (i) iff (ii) is trivial. Conversely, suppose (ii) and $c \in s$. If $c' \notin s$ for all $c' \subsetneq c$, then $c \in s^-$, and by assumption $w \models c$. If $c' \in s$ for some $c' \subsetneq c$, and without loss of generality $c'' \notin s$ for all $c'' \subsetneq c'$, then $c' \in s^-$, and by assumption $w \models c'$, and by subsumption $w \models c$.

The if direction of (i) iff (iii) is trivial. Conversely, suppose (i) and $c \in s^+$. Then $c \supseteq c'$ for some $c' \in s$. By assumption $w \models c'$, and by subsumption $w \models c$.

The if direction of (i) iff (iv) is trivial. Conversely suppose (i) and $c \in \mathsf{UP}(s)$. We show by induction on the length of the derivation of c that $w \models c$. The base case $c \in \mathsf{EQ} \cup s$ is trivial. For the induction step, let $c \in \mathsf{UP}(s)$ be the resolvent of $c \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s)$. By induction, $w \models c \vee \ell$ and $w \models \overline{\ell}$. Thus $w \models c$. □

6.4 A sound semantics of $\mathcal{L}^-$

We now define $s, k \mathrel{|\!\!\approx} \phi$ for a setup s, a natural number $k \in \{0, 1, 2, \ldots\}$, and an $\mathcal{L}^-$ formula ϕ. Intuitively, k indicates how much *effort* may be put into proving that ϕ is true. The effort is measured in how many times the setup is extended with new unit clauses.

Recall that the semantics ought to be sound, that is,

$$\text{if } s \models \phi, \text{ then } s, k \mathrel{|\!\approx} \phi.$$

To preserve soundness, we may not simply add some unit clause $[\ell]$ to the setup, but we always must also consider the case where $[\overline{\ell}]$ is added. To ease notation, we use the following abbreviation.

Definition 6.4.1 For a setup s and a ground clause ℓ, we write $s \uplus \ell$ for $s \cup \{[\ell]\}$.

Effort k then means that we may split k times on literals, that is, consider $s \uplus \ell$ and $s \uplus \overline{\ell}$. The rationale is that when the literals to be split are chosen smartly, they hopefully set off a cascade of unit propagations that leads to many new clauses. The principle was probably first used by Davis, Logemann, and Loveland (1962) and Davis and Putnam (1960) in their DPLL algorithm to decide validity of propositional formulas.

Definition 6.4.2 The sound truth relation $\mathrel{|\!\approx}$ is defined with respect to a setup s and $k \in \{0, 1, 2, \ldots\}$:

$\mathcal{L}°1.$ $s, k+1 \mathrel{|\!\approx} \phi$ iff $s \uplus \ell, k \mathrel{|\!\approx} \phi$ and $s \uplus \overline{\ell}, k \mathrel{|\!\approx} \phi$ for some ground literal ℓ;

$\mathcal{L}°2.$ if c is a clause:
$s, 0 \mathrel{|\!\approx} c$ iff $c \in \mathsf{UP}^+(s)$;

$\mathcal{L}°3.$ if $(\phi \vee \psi)$ is not a clause:
$s, 0 \mathrel{|\!\approx} (\phi \vee \psi)$ iff $s, 0 \mathrel{|\!\approx} \phi$ or $s, 0 \mathrel{|\!\approx} \psi$;

$\mathcal{L}°4.$ $s, 0 \mathrel{|\!\approx} \neg(\phi \vee \psi)$ iff $s, 0 \mathrel{|\!\approx} \neg\phi$ and $s, 0 \mathrel{|\!\approx} \neg\psi$;

$\mathcal{L}°5.$ $s, 0 \mathrel{|\!\approx} \neg\neg\phi$ iff $s, 0 \mathrel{|\!\approx} \phi$;

$\mathcal{L}°6.$ $s, 0 \mathrel{|\!\approx} \exists x \phi$ iff $s, 0 \mathrel{|\!\approx} \phi^x_n$ for some name n;

$\mathcal{L}°7.$ $s, 0 \mathrel{|\!\approx} \neg\exists x \phi$ iff $s, 0 \mathrel{|\!\approx} \neg\phi^x_n$ for all names n.

Observe that the *split rule*, Rule $\mathcal{L}°1$, branches on the truth value of some ground literal. This semantics is well-defined for $\mathcal{L}^-$ formulas as can be seen easily by induction on the length of ϕ: the cases for an atom, disjunction, or existential are covered by Rules $\mathcal{L}°2$, $\mathcal{L}°3$, $\mathcal{L}°6$, and for $\neg\phi$ the cases for atom, disjunction, negation, or existential are covered by Rules $\mathcal{L}°2$, $\mathcal{L}°4$, $\mathcal{L}°5$, $\mathcal{L}°7$.

Notice that while it is perfectly legal to split equality literals, it gains no new knowledge since $\mathsf{EQ} \subseteq \mathsf{UP}(s)$: on the one hand, $\mathsf{UP}(s \uplus (n = n)) = \mathsf{UP}(s)$ remains unchanged, on the other hand, $\mathsf{UP}(s \uplus (n \neq n)) = \{[]\}^+$ satisfies anything; analogously for $(n \neq n')$.

We illustrate $\approx\!\!\models$ with the kangaroo example introduced in Example 4.1.1 and formalized in Example 4.2.2.

Example 6.4.3 Let $s_\mu = \{[\text{Meat(roo)}], [\neg\text{Meat}(n), \neg\text{Eats}(n), \neg\text{Veggie}] \mid \text{for all names } n\}$ and $s_1 = \{[\neg\text{Aussie}, \neg\text{Italian}], [\neg\text{Aussie}, \text{Eats(roo)}], [\text{Italian}, \text{Veggie}], [\text{Italian}, \text{Aussie}]\} \cup s_\mu$. This setup corresponds to the first sphere e_1 from Example 4.5.5. There we argued that it is inconsistent with Aussie, that is, $s_1 \models \neg\text{Aussie}$. How can be obtain the same result in the limited semantics, that is, $s_1, k \approx\!\!\models \neg\text{Aussie}$?

Split level 0 is clearly not sufficient: we have $[\neg\text{Aussie}] \notin \mathsf{UP}^+(s_1)$, so $s_1, 0 \not\approx\!\!\models \neg\text{Aussie}$. A single split is enough, though: adding [Veggie] to s_1 triggers unit propagation that first yields [¬Meat(roo), ¬Eats(roo)], then [¬Eats(roo)], and then [¬Aussie]; on the other hand, adding [¬Veggie] yields [Italian] and then again [¬Aussie]. Thus, $s_1, 1 \approx\!\!\models \neg\text{Aussie}$.

Therefore, $s_1, k \approx\!\!\models \neg\text{Aussie}$ iff $k \geq 1$. Analogously we can argue that $s_1, k \approx\!\!\models \text{Italian}$ iff $k \geq 1$.

Before we investigate the properties of $\approx\!\!\models$, let us briefly discuss its relation to the limited semantics from (Lakemeyer and Levesque 2014). The most striking difference is that our semantics is phrased as if-and-only-if rules, whereas Lakemeyer and Levesque define it as least relation that satisfies a set of if-then rules. Besides the general appeal of if-and-only-if rules, the advantage of our definition is that it closely corresponds to the complete satisfaction relation presented in Section 6.6. (Defining that complete semantics with if-then rules appears to be cumbersome.)

On the downside, our if-and-only-if rules do not allow to split literals as freely as Lakemeyer and Levesque's semantics, where the formula may be (partially) decomposed first before splitting literals. Our semantics only allows splits in the beginning, which in a sense keeps the semantics deterministic. A drawback of this simplification is that our eventual completeness result in the next section is weaker than Lakemeyer and Levesque's: ours only holds for quantifier-free formulas, theirs also with leading universal quantifiers.

Another minor difference is our treatment of equality literals as part of setups instead of semantic rules. If we followed Lakemeyer and Levesque in adding semantic rules for $(n_1 = n_2)$ and $(n_1 \neq n_2)$, it would not suffice to merely check if n_1 and n_2 are identical or distinct, but another check to see whether the empty clause is in the setup would be necessary, because in this case even classically invalid equality expressions like $(n \neq n)$ should come out true. Lakemeyer and Levesque handle this case with a catch-all rule that says anything is satisfied by a setup that contains the empty clause. We avoided this by making the valid equality literals part of $\mathsf{UP}(s)$ and otherwise treating them like any other literal.

6.5 Soundness and eventual completeness

The following theorem establishes the aforementioned soundness of $\mathrel{|\!\approx}$ with respect to classical logic. Roughly, it holds because the inference rules unit propagation and subsumption are sound in classical logic.

Theorem 6.5.1 *If* $s, k \mathrel{|\!\approx} \phi$*, then* $s \models \phi$*.*

Proof. By induction on k and subinduction on the length of ϕ. (This is a recurring proof scheme for limited reasoning.) For a clause, $s, 0 \mathrel{|\!\approx} c$ iff $c \in \mathsf{UP}^+(s)$ only if $\mathsf{UP}^+(s) \models c$ iff (by Lemma 6.3.5) $s \models c$. For a non-clausal disjunction, $s, 0 \mathrel{|\!\approx} (\phi \vee \psi)$ iff $s, 0 \mathrel{|\!\approx} \phi$ or $s, 0 \mathrel{|\!\approx} \psi$ only if (by subinduction) $s \models \phi$ or $s \models \psi$ only if $s \models (\phi \vee \psi)$. For a negated disjunction, $s, 0 \mathrel{|\!\approx} \neg(\phi \vee \psi)$ iff $s, 0 \mathrel{|\!\approx} \neg\phi$ and $s, 0 \mathrel{|\!\approx} \neg\psi$ only if (by subinduction) $s \models \neg\phi$ and $s \models \neg\psi$ iff $s \models (\neg\phi \wedge \neg\psi)$ iff $s \models \neg(\phi \vee \psi)$. For a double negation, $s, 0 \mathrel{|\!\approx} \neg\neg\phi$ iff $s, 0 \mathrel{|\!\approx} \phi$ only if (by subinduction) $s \models \phi$ iff $s \models \neg\neg\phi$. For an existential, $s, 0 \mathrel{|\!\approx} \exists x\, \phi$ iff $s, 0 \mathrel{|\!\approx} \phi^x_n$ for some n only if (by subinduction) $s \models \phi^x_n$ for some n only if $s \models \exists x\, \phi$. For a negated existential, $s, 0 \mathrel{|\!\approx} \neg\exists x\, \phi$ iff $s, 0 \mathrel{|\!\approx} \neg\phi^x_n$ for all n only if (by subinduction) $s \models \neg\phi^x_n$ for all n iff $s \models \forall x. \neg\phi$ iff $s \models \neg\exists x\, \phi$. This completes the subinduction.

Now suppose the lemma holds for k for the main induction step. Let $w \models c$ for all $c \in s$, and $s, k+1 \mathrel{|\!\approx} \phi$. By the split rule, for some ℓ, $s \uplus \ell, k \mathrel{|\!\approx} \phi$ and $s \uplus \overline{\ell}, k \mathrel{|\!\approx} \phi$. By induction, $s \uplus \ell \models \phi$ and $s \uplus \overline{\ell} \models \phi$. Therefore, since either $w \models \ell$ or $w \models \overline{\ell}$, we have $w \models \phi$. Hence $s \models \phi$. □

Another interesting property is the so-called eventual completeness for propositional formulas, which we prove in Theorem 6.5.5. Intuitively, it says that every classically valid propositional formula can also be proved in $\mathrel{|\!\approx}$ for large enough effort.

Lemma 6.5.2 *Let* $s \subseteq s'$*. Then* $\mathsf{UP}^+(s) \subseteq \mathsf{UP}^+(s')$*.*

Proof. We first show that if $c \in \mathsf{UP}(s)$ then $c \in \mathsf{UP}(s')$ by induction on the length of the derivation of c. If $c \in \mathsf{EQ} \cup s$, then $c \in \mathsf{EQ} \cup s'$. If $c \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s)$, and by induction, $c \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s')$, so $c \in \mathsf{UP}(s')$. Finally, if $c \in \mathsf{UP}^+(s)$, then $c' \in \mathsf{UP}(s)$ for some $c' \subseteq c$. By the above, $c' \in \mathsf{UP}(s')$, and thus $c \in \mathsf{UP}^+(s')$. □

Lemma 6.5.3 *Let* $\mathsf{UP}^+(s) \subseteq \mathsf{UP}^+(s')$*. If* $s, k \mathrel{|\!\approx} \phi$*, then* $s', k \mathrel{|\!\approx} \phi$*.*

Proof. By induction on k and subinduction on the length of ϕ. For a clause, $s, 0 \mathrel{|\!\approx} c$ iff $c \in \mathsf{UP}^+(s)$ only if (by Lemma 6.5.2) $c \in \mathsf{UP}^+(s')$ iff $s', 0 \mathrel{|\!\approx} c$. The subinduction steps are trivial. For the induction step suppose the lemma holds for k and $s, k+1 \mathrel{|\!\approx} \phi$. By the split rule, for some ℓ, $s \uplus \ell, k \mathrel{|\!\approx} \phi$ and $s \uplus \overline{\ell}, k \mathrel{|\!\approx} \phi$. By the main induction, $s' \uplus \ell, k \mathrel{|\!\approx} \phi$ and $s' \uplus \overline{\ell}, k \mathrel{|\!\approx} \phi$. By the split rule, $s', k+1 \mathrel{|\!\approx} \phi$. □

Lemma 6.5.4 *If* $s, k \mathrel{|\!\approx} \phi$*, then* $s, k+1 \mathrel{|\!\approx} \phi$*.*

Proof. Suppose $s, k \mathrel{|\!\overset{\circ}{\approx}} \phi$. By the split rule, $s, k+1 \mathrel{|\!\overset{\circ}{\approx}} \phi$ iff for some ℓ, $s \uplus \ell, k \mathrel{|\!\overset{\circ}{\approx}} \phi$ and $s \uplus \overline{\ell}, k \mathrel{|\!\overset{\circ}{\approx}} \phi$, which hold by Lemmas 6.5.2 and 6.5.3 and by assumption. □

Theorem 6.5.5 *Let s be finite and ϕ be propositional.*
Then for some k' and for all $k \geq k'$, $s \models \phi$ iff $s, k' \mathrel{|\!\overset{\circ}{\approx}} \phi$.

Proof. The if direction follows from Theorem 6.5.1. Conversely, suppose $s \models \phi$. Let w be an arbitrary world. Then either $w \not\models c$ for some $c \in s$, or $w \models \phi$. Let $\ell \in L$ iff $w \models \ell$ and ℓ or $\overline{\ell}$ occurs in s or ϕ. Clearly, L is finite. We write $s \uplus \{\ell_1, \ldots, \ell_j\}$ for $s \uplus \ell_1 \uplus \ldots \uplus \ell_j$. We show that $s \uplus L, 0 \mathrel{|\!\overset{\circ}{\approx}} \phi$. If $w \not\models c$ for some $c \in s$, then $\overline{\ell} \in L$ for all $\ell \in c$, and thus $[] \in \mathsf{UP}^+(s \uplus L)$, so $s \uplus L, 0 \mathrel{|\!\overset{\circ}{\approx}} \phi$ as can be shown by a trivial induction on the length of ϕ. If $w \models c$ for all $c \in s$, we show $s \uplus L, 0 \mathrel{|\!\overset{\circ}{\approx}} \phi$ by induction on the length of ϕ. For a clause, $w \models c'$ iff $w \models \ell$ for some $\ell \in c'$ iff $\ell \in L$ only if $c' \in \mathsf{UP}^+(s \uplus L)$ iff $s \uplus L, 0 \mathrel{|\!\overset{\circ}{\approx}} c'$. The induction steps are trivial. Since w was arbitrary, we have that for all truth assignments L of the atoms from s and ϕ, $s \uplus L \mathrel{|\!\overset{\circ}{\approx}} \phi$. Thus splitting these atoms obtains ϕ, that is, $s, k \mathrel{|\!\overset{\circ}{\approx}} \phi$ where $k = |L|$. By Lemma 6.5.4, $s, k' \mathrel{|\!\overset{\circ}{\approx}} \phi$ for all $k' \geq k$. □

In (Lakemeyer and Levesque 2014), the eventual completeness also holds true for sentences $\forall \vec{x} \phi$ where ϕ is quantifier-free. As discussed in the previous section, the reason is that they can first handle the quantifiers by substituting some $\vec{n}$ for $\vec{x}$, and then split depending on $\vec{n}$.

6.6 A complete semantics of $\mathcal{L}^-$

Now we turn to $s, l \mathrel{|\!\overset{\bullet}{\approx}} \phi$ for a setup s, a natural number $l \in \{0, 1, 2, \ldots\}$, and an $\mathcal{L}^-$ formula ϕ. Recall that this semantics ought to be complete, that is,

$$\text{if } s, l \mathrel{|\!\overset{\bullet}{\approx}} \phi, \text{ then } s \models \phi.$$

In $\mathrel{|\!\overset{\bullet}{\approx}}$, it is often more intuitive to consider the task of *dis*proving that s satisfies ϕ. Accordingly, we often express completeness by the contrapositive:

$$\text{if } s \not\models \phi, \text{ then } s, l \mathrel{\not|\!\overset{\bullet}{\approx}} \phi.$$

Again, l specifies the reasoning effort. In $s, k \mathrel{|\!\overset{\circ}{\approx}} \phi$ one (roughly) shows that for some atoms, ϕ obviously comes out true in s under any truth assignment of these atoms (where "obvious" means after unit propagation and subsumption). By contrast, in $s, l \mathrel{\not|\!\overset{\bullet}{\approx}} \phi$ we need to show that after adding certain atoms to s, the resulting setup obviously disproves ϕ.

In particular, this requires to detect whether the setup *might* be inconsistent. For that, we use a very simple heuristic: whenever the setup mentions some literal both positively and negatively after removing all subsumed clauses, it is deemed possibly-inconsistent. While this is of course not sophisticated, the naivete of this test can be compensated by increasing l, that is, by more reasoning effort.

Definition 6.6.1 For a setup s, a (perhaps non-ground) clause c, and a (perhaps non-ground) literal ℓ, we define the following expressions:

$$\begin{aligned}
\mathsf{XP}(s) &= \mathsf{UP}(\{[\ell] \mid \text{for some } c,\, c \cup [\ell] \in \mathsf{UP}^-(s)\});\\
\mathsf{gnd}(c) &= \{c^{x_1 \ldots x_k}_{n_1 \ldots n_k} \mid n_1, \ldots, n_k \text{ are standard names}\};\\
\mathsf{L}(\ell, s) &= \{[\ell'] \in \mathsf{gnd}([\ell]) \mid [\overline{\ell'}] \notin \mathsf{UP}^+(s)\};\\
s \otimes \ell &= s \cup \mathsf{L}(\ell, s).
\end{aligned}$$

The motivation behind $\mathsf{XP}(s)$ is to facilitate a simple consistency check of s. $\mathsf{XP}(s)$ contains the unit clause of every literal that occurs in $\mathsf{UP}^+(s)$, and closes this set under unit propagation. As a consequence, if $[] \notin \mathsf{XP}(s)$, then s is classically consistent. For example, $[] \in \mathsf{XP}(\{[P, Q], [P, \neg Q]\})$ because of the occurrence of Q and $\neg Q$, but $[] \notin \mathsf{XP}(\{[P], [P, Q], P, \neg Q]\})$ because $[P]$ subsumes the clauses with Q and $\neg Q$.

The operator $\mathsf{gnd}(c)$ grounds a clause by substituting standard names for variables. For example, $\mathsf{gnd}([P(x_1), Q(x_2)]) = \{[P(n_1), Q(n_2)] \mid n_1, n_2 \text{ are standard names}\}$.

The idea behind $\mathsf{L}(\ell, s)$ is to determine all ground instances of ℓ which are not obviously inconsistent with s. That is, $\mathsf{L}(\ell, s)$ contains every instance of ℓ whose negation does not occur as a unit clause in $\mathsf{UP}^+(s)$. For example, $\mathsf{L}(P(x), \{[P(\#1)]\}) = \{[P(n)] \mid n \text{ is a standard name other than } \#1\}$. Therefore, $s \otimes \ell$ in a sense applies a closed-world assumption for all instances of ℓ, except for those whose negation is already known. The rationale is that often a setup may contain infinitely many instances of some predicate, and all of them should be fixed to the same truth value. Observe the differences between $s \otimes \ell$ and $s \uplus \ell'$: the former is also defined for non-ground ℓ and only adds instances of $[\ell]$ that do not lead immediately to the empty clause; the latter adds the ground unit clause $[\ell']$ to the setup no matter what.

Lemma 6.6.2 *If* $[] \notin \mathsf{XP}(s)$, *then for some* w, *for all* $c \in s$, $w \models c$.

Proof. Let $[] \notin \mathsf{XP}(s)$ and let $w \models \ell$ iff $[\ell] \in \mathsf{XP}(s)$, which exists as $[\ell] \notin \mathsf{XP}(s)$ or $[\overline{\ell}] \notin \mathsf{XP}(s)$. By subsumption, $w \models c$ for all $c \in \mathsf{UP}^-(s)$. By Lemma 6.3.5, $w \models c$ for all $c \in s$. □

Definition 6.6.3 The complete truth relation $\models^{\mathrm{e}}$ is defined with respect to a setup s and

$l \in \{0, 1, 2, \ldots\}$:

$\mathcal{L}^{\bullet}1$. $s, l+1 \mathrel{|\!\approx^{\bullet}} \phi$ iff $s \otimes \ell, l \mathrel{|\!\approx^{\bullet}} \phi$ for all literals ℓ;

$\mathcal{L}^{\bullet}2$. if c is a clause:
$s, 0 \mathrel{|\!\approx^{\bullet}} \neg c$ iff $[] \in \mathsf{XP}(s)$ or $c \notin \mathsf{UP}^{+}(s)$;

$\mathcal{L}^{\bullet}3$. $s, 0 \mathrel{|\!\approx^{\bullet}} (\phi \vee \psi)$ iff $s, 0 \mathrel{|\!\approx^{\bullet}} \phi$ or $s, 0 \mathrel{|\!\approx^{\bullet}} \psi$;

$\mathcal{L}^{\bullet}4$. if $(\phi \vee \psi)$ is not a clause:
$s, 0 \mathrel{|\!\approx^{\bullet}} \neg(\phi \vee \psi)$ iff $s, 0 \mathrel{|\!\approx^{\bullet}} \neg\phi$ and $s, 0 \mathrel{|\!\approx^{\bullet}} \neg\psi$;

$\mathcal{L}^{\bullet}5$. $s, 0 \mathrel{|\!\approx^{\bullet}} \neg\neg\phi$ iff $s, 0 \mathrel{|\!\approx^{\bullet}} \phi$;

$\mathcal{L}^{\bullet}6$. $s, 0 \mathrel{|\!\approx^{\bullet}} \exists x \phi$ iff $s, 0 \mathrel{|\!\approx^{\bullet}} \phi^x_n$ for some name n;

$\mathcal{L}^{\bullet}7$. $s, 0 \mathrel{|\!\approx^{\bullet}} \neg\exists x \phi$ iff $s, 0 \mathrel{|\!\approx^{\bullet}} \neg\phi^x_n$ for all names n.

Notice that Rule $\mathcal{L}^{\bullet}1$, the so called *add rule*, allows the literal to be non-ground. This semantics is well-defined for formulas of $\mathcal{L}^{-}$, just like $\mathrel{|\!\approx^{\circ}}$, except that the case for a positive literal ℓ is not as obvious: $s, 0 \mathrel{|\!\approx^{\bullet}} \ell$ iff (by Rule $\mathcal{L}^{\bullet}5$) $s, 0 \mathrel{|\!\approx^{\bullet}} \neg\neg\ell$ iff (by Rule $\mathcal{L}^{\bullet}2$) $s, 0 \mathrel{|\!\approx^{\bullet}} \neg\overline{\ell}$ where $\neg\overline{\ell}$ is taken as negated clause $\neg c$.

Let us illustrate how $\mathrel{|\!\approx^{\bullet}}$ works with the kangaroo example.

Example 6.6.4 Consider s_1 from Example 6.4.3, and let us see whether it is consistent with Italian, that is, $s_1, l \not\mathrel{|\!\approx^{\bullet}} \neg$Italian for certain l. Clearly, $s_1, 0 \mathrel{|\!\approx^{\bullet}} \neg$Italian, since $\mathsf{UP}^{-}(s_1)$ mentions, for example, Italian and $\neg$Italian in clauses, and thus $[] \in \mathsf{XP}(s_1)$. For $l \geq 1$, however, we are allowed to add some literal to s_1 so as to build a countermodel that clearly disproves $\neg$Italian. Indeed, adding, for example, [$\neg$Aussie] does the job: $\mathsf{UP}^{-}(s_1 \otimes \neg\text{Aussie}) = \{[\neg\text{Aussie}], [\text{Italian}]\} \cup \{[\text{Meat(roo)}], [\neg\text{Eats(roo)}, \neg\text{Veggie}], [\neg\text{Meat}(n), \neg\text{Eats}(n), \neg\text{Veggie}] \mid \text{for all names } n \neq \text{roo}\} \cup \mathsf{EQ}$ is obviously consistent, that is, $[] \notin \mathsf{XP}(s_1)$, and moreover $[\text{Italian}] \in \mathsf{UP}^{+}(s_1 \otimes \neg\text{Aussie})$. So by adding $\neg$Aussie, we have shown that s_1 can falsify $\neg$Italian. Thus, $s_1, l \not\mathrel{|\!\approx^{\bullet}} \neg$Italian iff $l \geq 1$.

6.7 Completeness and eventual soundness

To relate $\mathrel{|\!\approx^{\bullet}}$ to classical logic, we show that it is complete and eventually sound. Since $\mathrel{|\!\approx^{\circ}}$ is sound and eventually complete, both limited semantics complement each other.

The intuitive argument for the completeness for $\mathrel{|\!\approx^{\bullet}}$ is this: if $[] \notin \mathsf{XP}(s)$ and $c \in \mathsf{UP}(s)$, then s must be classically consistent and $s \models c$, and therefore $s \not\models \neg c$.

Theorem 6.7.1 *If $s \models \phi$, then $s, l \mathrel{|\!\approx^{\bullet}} \phi$.*

Proof. By induction on l. Suppose $s, 0 \not\mathrel{|\!\approx} \phi$. Then $[] \notin \mathsf{XP}(s)$. By Lemma 6.6.2, there is a w such that $w \models c$ for all $c \in s$ (*). We first show that $w \not\models \phi$ by subinduction on the length of ϕ. For a negated clause, $s, 0 \not\mathrel{|\!\approx} \neg c$ iff $[] \notin \mathsf{XP}(s)$ and $c \in \mathsf{UP}^+(s)$ only if (by (*) and Lemma 6.3.5) $w \models c$ iff $w \not\models \neg c$. For a literal, $s, 0 \not\mathrel{|\!\approx} \ell$ iff $s, 0 \not\mathrel{|\!\approx} \neg\overline{\ell}$ only if (by the same argument as for negated clauses) $w \not\models \neg\overline{\ell}$ iff $w \not\models \ell$. For a disjunction, $s, 0 \not\mathrel{|\!\approx} (\phi \lor \psi)$ iff $s, 0 \not\mathrel{|\!\approx} \phi$ and $s, 0 \not\mathrel{|\!\approx} \psi$ only if (by subinduction) $w \not\models \phi$ and $w \not\models \psi$ iff $w \models (\neg\phi \land \neg\psi)$ iff $w \not\models (\phi \lor \psi)$. For a negated non-clausal disjunction, $s, 0 \not\mathrel{|\!\approx} \neg(\phi \lor \psi)$ iff $s, 0 \not\mathrel{|\!\approx} \neg\phi$ or $s, 0 \not\mathrel{|\!\approx} \neg\psi$ only if (by subinduction) $w \not\models \neg\phi$ or $w \not\models \neg\psi$ iff $w \models (\phi \lor \psi)$ iff $w \not\models \neg(\phi \lor \psi)$. For a double negation, $s, 0 \not\mathrel{|\!\approx} \neg\neg\phi$ iff $s, 0 \not\mathrel{|\!\approx} \phi$ only if (by subinduction) $w \not\models \phi$ iff $w \not\models \neg\neg\phi$. For an existential, $s, 0 \not\mathrel{|\!\approx} \exists x\, \phi$ iff $s, 0 \not\mathrel{|\!\approx} \phi^x_n$ for all n only if (by subinduction) $w \not\models \phi^x_n$ for all n iff $w \models \forall x \neg\phi$ iff $w \not\models \exists x\, \phi$. For a negated existential, $s, 0 \not\mathrel{|\!\approx} \neg\exists x\, \phi$ iff $s, 0 \not\mathrel{|\!\approx} \neg\phi^x_n$ for some n only if (by subinduction) $w \not\models \neg\phi^x_n$ for some n iff $w \models \exists x\, \phi$ iff $w \not\models \neg\exists x\, \phi$. This completes the subinduction.

Now suppose the lemma holds for l for the main induction step. Let $s, l+1 \not\mathrel{|\!\approx} \phi$. Then $s \otimes \ell, l \not\mathrel{|\!\approx} \phi$ for some ℓ. By induction, $s \otimes \ell \not\models \phi$. By monotonicity, $s \not\models \phi$. □

In the propositional case $\not\mathrel{|\!\approx}$ is eventually sound, that is, all invalid inferences can be detected for large enough l.

Lemma 6.7.2 *If $s, l \not\mathrel{|\!\approx} \phi$, then $s, l+1 \not\mathrel{|\!\approx} \phi$.*

Proof. By induction on l and subinduction on the length of ϕ. For a negated clause, $s, 0 \not\mathrel{|\!\approx} \neg c$ iff (since $\mathsf{L}((x \neq x), s) = \{\}$) iff $s \otimes (x \neq x), 0 \not\mathrel{|\!\approx} \neg c$ only if $s, 1 \not\mathrel{|\!\approx} \neg c$. The other cases for the subinduction are trivial. Now suppose the lemma holds for $l-1$ for the main induction step. Let $s, l \not\mathrel{|\!\approx} \phi$. Then $s \otimes \ell, l-1 \not\mathrel{|\!\approx} \phi$ for some ℓ. By induction, $s \otimes \ell, l \not\mathrel{|\!\approx} \phi$. Thus $s, l+1 \not\mathrel{|\!\approx} \phi$. □

Theorem 6.7.3 *Let s be finite and ϕ be propositional.*
Then for some l' and for all $l \geq l'$, $s \models \phi$ iff $s, l' \mathrel{|\!\approx} \phi$.

Proof. The only-if direction follows from Theorem 6.7.1. Conversely, suppose $s \not\models \phi$. Then for some w, $w \models c$ for all $c \in s$, and $w \not\models \phi$. Let $\ell \in L$ iff $w \models \ell$ and ℓ or $\overline{\ell}$ occurs in s or ϕ. Clearly, L is finite. We write $s \uplus \{\ell_1, \ldots, \ell_j\}$ for $s \uplus \ell_1 \uplus \ldots \uplus \ell_j$. Since w exists, $[] \notin \mathsf{UP}(s)$ by Lemma 6.3.5, and so $\mathsf{UP}^-(s \uplus L) = L$, and thus $[] \notin \mathsf{XP}(s \uplus L)$ (*). We show that $s \uplus L, 0 \not\mathrel{|\!\approx} \phi$ by induction on the length of ϕ. For a negated clause, $w \not\models \neg c$ iff $w \models \ell$ for some $\ell \in c$ only if $\ell \in L$ for some $\ell \in c$ only if (by (*)) $[] \notin \mathsf{XP}(s \uplus L)$ and $c \in \mathsf{UP}^+(s \uplus L)$ iff $s \uplus L, 0 \not\mathrel{|\!\approx} \neg c$. The other cases for the induction are trivial. Note that L is finite. Hence, $s \uplus L, 0 \not\mathrel{|\!\approx} \phi$ only if $s, l \not\mathrel{|\!\approx} \phi$ where $l = |L|$. By Lemma 6.7.2, $s, l' \not\mathrel{|\!\approx} \phi$ for all $l' \geq l$. □

6.8 Decision procedures for proper$^+$ knowledge bases

The big advantage of the limited semantics $\mathrel{|\!\overset{\circ}{\approx}}$ and $\mathrel{|\!\overset{\bullet}{\approx}}$ over the full semantics is that, for a specific class of knowledge bases at least, reasoning is decidable. In this section we investigate decision procedures for so-called proper$^+$ knowledge bases. These will be foundational to our decision procedure for limited conditional beliefs in the next chapter.

There are two key ingredients for these results. For one thing, proper$^+$ formulas determine a canonical model, so that entailment reduces to model checking for this single setup. For another, standard names which do not occur in the knowledge base or query cannot be distinguished. Hence only a finite number of them needs to be considered: those from the knowledge base and query, plus one for every quantifier, multiplied with k and l (roughly).

Definition 6.8.1 A sentence π is *proper*$^+$ when it is of the form $\bigwedge_j \forall \vec{x}_j c_j$ for clauses c_j. Then we let $\mathsf{gnd}(\pi) = \bigcup_j \mathsf{gnd}(c_j)$. A setup s is UP^+*-minimal* with $s, k \mathrel{|\!\overset{\circ}{\approx}} \phi$ iff there is no s' with $\mathsf{UP}^+(s') \subsetneq \mathsf{UP}^+(s)$ and $s', k \mathrel{|\!\overset{\circ}{\approx}} \phi$.

It appears plausible that $\mathsf{gnd}(\pi)$ is the canonical model of π. The following Theorem 6.8.4 confirms this intuition. First we need two lemmas, which are also used frequently in the decidability proofs.

Lemma 6.8.2 *Let* $f \in \{\mathsf{UP}^+, \mathsf{UP}^-, \mathsf{XP}\}$.
Then $f(s \cup s') = f(s^- \cup s') = f(s^+ \cup s') = f(\mathsf{UP}(s) \cup s')$.

Proof. We first show the lemma for UP^+.

First consider $\mathsf{UP}^+(s \cup s') = \mathsf{UP}^+(s^- \cup s')$. The $\supseteq$ direction follows from Lemma 6.5.2 since $s \cup s' \supseteq s^- \cup s'$. Conversely, suppose $c \in \mathsf{UP}^+(s \cup s')$. Then $c \supseteq c'$ for some $c' \in \mathsf{UP}(s \cup s')$. Let c' be minimal, that is, $c'' \notin \mathsf{UP}(s \cup s')$ for all $c'' \subsetneq c'$. We show by induction on the length of the derivation of c' that $c' \in \mathsf{UP}(s^- \cup s')$, which implies $c \in \mathsf{UP}^+(s^- \cup s')$. If $c' \in \mathsf{EQ} \cup s \cup s'$, then $c' \in \mathsf{EQ} \cup s^- \cup s'$. Otherwise, $c' \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s \cup s')$, and $c' \cup [\ell], [\overline{\ell}]$ are minimal among the clauses of that derivation length, that is, the derivation of all $c'' \in \mathsf{UP}(s \cup s')$ with $c'' \subsetneq c' \cup [\ell]$ or $c' \subsetneq [\ell]$ is longer than those of $c' \cup [\ell]$ and $[\overline{\ell}]$. Then by induction $c' \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s^- \cup s')$ and thus $c' \in \mathsf{UP}(s^- \cup s)$.

Now consider $\mathsf{UP}^+(s \cup s') = \mathsf{UP}^+(s^+ \cup s')$. The $\subseteq$ direction holds by Lemma 6.5.2 since $s \cup s' \subseteq s^+ \cup s'$. Conversely, suppose $c \in \mathsf{UP}^+(s^+ \cup s')$. Then $c \supseteq c'$ for some $c' \in \mathsf{UP}(s^+ \cup s')$. We show by induction on the length of the derivation of c' that $c'' \in \mathsf{UP}(s \cup s')$ for some $c'' \subseteq c'$, which implies $c \in \mathsf{UP}^+(s \cup s')$. If $c' \in \mathsf{EQ} \cup s^+ \cup s'$, then $c'' \in \mathsf{EQ} \cup s \cup s'$ for some $c'' \subseteq c'$. Otherwise, $c' \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s^+ \cup s')$, and by induction, $c_1'', c_2'' \in \mathsf{UP}(s^+ \cup s')$ for some $c_1'' \subseteq c' \cup [\ell]$ and $c_2'' \subseteq [\overline{\ell}]$. If $\ell \notin c_1''$ or $\overline{\ell} \notin c_2''$, then $c'' \in \mathsf{UP}(s \cup s')$ for some $c'' \subseteq c'$, and otherwise $c_1'' \setminus [\ell] \in \mathsf{UP}(s \cup s')$ and $c_1'' \setminus [\ell] \subseteq c'$.

Now consider $\mathsf{UP}^+(s \cup s') = \mathsf{UP}^+(\mathsf{UP}(s) \cup s')$. The $\subseteq$ direction holds by Lemma 6.5.2 since $s \cup s' \subseteq \mathsf{UP}(s) \cup s'$. Conversely, suppose $c \in \mathsf{UP}^+(\mathsf{UP}(s) \cup s')$. Then $c \supseteq c'$ for some $c' \in \mathsf{UP}(\mathsf{UP}(s) \cup s')$. We show by induction on the length of the derivation of c' that $c' \in \mathsf{UP}(s \cup s')$, which implies $c \in \mathsf{UP}^+(s \cup s')$. If $c' \in s'$, then trivially $c' \in \mathsf{UP}(s \cup s')$. If $c' \in \mathsf{UP}(s)$, then either $c' \in s$, in which case $c' \in \mathsf{UP}(s \cup s')$ is again trivial, or $c' \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s)$, so $c' \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s \cup s')$ by Lemma 6.5.2, and hence $c' \in \mathsf{UP}(s \cup s')$. Otherwise, $c' \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(\mathsf{UP}(s) \cup s')$, and by induction, $c' \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s \cup s')$, so $c' \in \mathsf{UP}(s \cup s')$.

The lemma for UP^- follows from the above since $(s^+)^- = s^-$ for arbitrary s. This in turn immediately gives the lemma for XP. □

Lemma 6.8.3 *Let* $f \in \{\mathsf{UP}, \mathsf{UP}^+, \mathsf{UP}^-\}$.

(i) $s \cup s', k \mathrel{|\!\!\approx} \phi$ *iff* $f(s) \cup s', k \mathrel{|\!\!\approx} \phi$;

(ii) $s \cup s', l \mathrel{|\!\!\approx} \phi$ *iff* $f(s) \cup s', l \mathrel{|\!\!\approx} \phi$.

Proof. By Lemma 6.8.2, which also gives that $\mathsf{L}(\ell, s \cup s') = \mathsf{L}(\ell, f(s) \cup s')$, both claims follow by a simple induction on k or l and subinduction on the length of ϕ. □

Theorem 6.8.4 (Lakemeyer and Levesque 2013) *Let* π *be proper*$^+$.
Then s *is* UP^+*-minimal with* $s, 0 \mathrel{|\!\!\approx} \pi$ *iff* $\mathsf{UP}^+(s) = \mathsf{UP}^+(\mathsf{gnd}(\pi))$.

Proof. For the only-if direction let s be UP^+-minimal with $s, 0 \mathrel{|\!\!\approx} \pi$. Then $s, 0 \mathrel{|\!\!\approx} c$ for every $c \in \mathsf{gnd}(\pi)$, and thus $c \in \mathsf{UP}^+(s)$. Thus $\mathsf{gnd}(\pi) \subseteq \mathsf{UP}^+(s)$. Hence $\mathsf{UP}^+(\mathsf{gnd}(\pi)) \subseteq \mathsf{UP}^+(\mathsf{UP}^+(s))$, and by Lemma 6.8.2, $\mathsf{UP}^+(\mathsf{gnd}(\pi)) \subseteq \mathsf{UP}^+(s)$. Moreover, $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} \pi$ and by assumption s is UP^+-minimal with $s, 0 \mathrel{|\!\!\approx} \pi$, so $\mathsf{UP}^+(s) \subseteq \mathsf{UP}^+(\mathsf{gnd}(\pi))$.

Conversely, let $\mathsf{UP}^+(s) = \mathsf{UP}^+(\mathsf{gnd}(\pi))$. Clearly $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} \pi$, and by applying Lemma 6.8.3 twice, $s, 0 \mathrel{|\!\!\approx} \pi$. Now consider an s' with $\mathsf{UP}^+(s') \subsetneq \mathsf{UP}^+(s)$. Then there is a $c \in \mathsf{gnd}(\pi)$ such that $c \notin \mathsf{UP}^+(s')$. Hence $s', 0 \not\mathrel{|\!\!\approx} c$ and so $s', 0 \not\mathrel{|\!\!\approx} \pi$. Thus s is UP^+-minimal with $s, 0 \mathrel{|\!\!\approx} \pi$. □

Together with Lemma 6.5.3, this theorem reduces the consequences of proper$^+$ knowledge bases to model checking. This property will be fundamental especially in the next chapter in the context of only-believing. For now, let us investigate how to decide which formulas are satisfied by $\mathsf{gnd}(\pi)$.

Definition 6.8.5 We let the width $|\phi|_{\mathrm{w}}$ of ϕ be the maximum of

- the highest arity of any predicate symbol in ϕ other than $=$, and

- the largest number of free variables in any subformula of ϕ.

For any set of names N, we let

$$\begin{aligned}
\mathsf{gnd}_N(c) &= \{c^{x_1 \dots x_k}_{n_1 \dots n_k} \mid n_1, \dots, n_k \in N\};\\
\mathsf{L}_N(\ell, s) &= \{[\ell'] \in \mathsf{gnd}_N([\ell]) \mid [\overline{\ell'}] \notin \mathsf{UP}^+(s)\};\\
s \otimes_N \ell &= s \cup \mathsf{L}_N(\ell, s).
\end{aligned}$$

As we shall see, considering all names that occur in the knowledge base π or the query ϕ plus $(k+1) \cdot \max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\}$ more names is sufficient to decide $\mathsf{gnd}(\pi), k \mathrel{|\!\!\approx} \phi$, and analogously for $\mathsf{gnd}(\pi), l \mathrel{|\!\!\approx} \phi$. We begin with the decision procedure for the sound semantics.

Definition 6.8.6 The decision procedure $\mathsf{S}[N, s, k, \phi] \in \{0, 1\}$ for $s, k \mathrel{|\!\!\approx} \phi$ is defined as follows:

- $\mathsf{S}[N, s, k+1, \phi] = 1$ iff $\mathsf{S}[N, s \uplus \ell, k, \phi] = \mathsf{S}[N, s \uplus \overline{\ell}, k, \phi] = 1$ for some ground ℓ whose symbol occurs in s or ϕ and whose names are from N;
- if c is a clause:
  $\mathsf{S}[N, s, 0, c] = 1$ iff $c \in \mathsf{UP}^+(s)$;
- if $(\phi \vee \psi)$ is not a clause:
  $\mathsf{S}[N, s, 0, (\phi \vee \psi)] = \max\{\mathsf{S}[N, s, 0, \phi], \mathsf{S}[N, s, 0, \psi]\}$;
- $\mathsf{S}[N, s, 0, \neg(\phi \vee \psi)] = \min\{\mathsf{S}[N, s, 0, \neg\phi], \mathsf{S}[N, s, 0, \neg\psi]\}$;
- $\mathsf{S}[N, s, 0, \neg\neg\phi] = \mathsf{S}[N, s, 0, \phi]$;
- $\mathsf{S}[N, s, 0, \exists x\, \phi] = \max\{\mathsf{S}[N, s, 0, \phi^x_n] \mid n \in N\}$;
- $\mathsf{S}[N, s, 0, \neg\exists x\, \phi] = \min\{\mathsf{S}[N, s, 0, \neg\phi^x_n] \mid n \in N\}$.

Theorem 6.8.7 *Let π be proper$^+$ and let N contain the names from π and ϕ plus $k \cdot v + v$ names for $v \geq |\pi|_\mathrm{w}$ and $v \geq |\phi|_\mathrm{w}$. Then* $\mathsf{gnd}(\pi), k \mathrel{|\!\!\approx} \phi$ *iff* $\mathsf{S}[N, \mathsf{gnd}_N(\pi), k, \phi] = 1$.

The proof is given in Appendix C.1. Two key results need to be established. Firstly, as sketched above, quantification and grounding can be restricted to the finite set of names N. Secondly, relevant split literals are only those whose symbol occurs in π or ϕ and whose names are from N. The proof is then by induction on k and subinduction on the length of ϕ.

Analysing the complexity of the decision procedure obtains that it grows exponentially in the effort k and in the widths $|\pi|_\mathrm{w}$ and $|\phi|_\mathrm{w}$.

Theorem 6.8.8 *Let π be proper$^+$. Then* $\mathsf{gnd}(\pi), k \mathrel{|\!\approx} \phi$ *can be decided in time* $O((|\pi| + k)^{k+1} \cdot |\phi|^{k+1} \cdot (\max\{|\pi|_w, |\phi|_w\} \cdot (|\pi| + |\phi| + k + 1))^{(|\pi|_w+|\phi|_w)\cdot(k+1)} \cdot 2^k)$.

The proof is given in Appendix C.1. The third factor in the complexity bound represents the blowup due to names that need to be substituted for variables. For one thing, these names lead to larger setups, and for another to a large set of possibly relevant split literals. In the propositional case, this complexity disappears and we obtain the following corollary.

Corollary 6.8.9 *Let π be proper$^+$ and propositional, and let ϕ also be propositional. Then* $\mathsf{gnd}(\pi), k \mathrel{|\!\approx} \phi$ *can be decided in time* $O((|\pi| + k)^{k+1} \cdot |\phi|^{k+1} \cdot 2^k)$.

Very similar bounds on the complexity were established for an ancestor of our semantics in (Liu 2006; Liu, Lakemeyer, and Levesque 2004). In their formalism, only clauses occurring already in the setup are split.

Next, we turn to the complete semantics for which we provide analogous results.

Definition 6.8.10 The decision procedure $\mathsf{C}[N, s, l, \phi] \in \{0, 1\}$ for $s, l \mathrel{|\!\approx} \phi$ is defined as follows:

- $\mathsf{C}[N, s, l + 1, \phi] = 1$ iff $\mathsf{C}[N, s \otimes_N \ell, l, \phi] = 1$ for all (perhaps non-ground) ℓ whose symbol occurs in s or ϕ and whose names are from N;
- if ℓ is a positive literal:
  $\mathsf{C}[N, s, 0, \ell] = \mathsf{C}[N, s, 0, \neg\overline{\ell}]$;
- if c is a clause:
  $\mathsf{C}[N, s, 0, \neg c] = 1$ iff $[] \in \mathsf{XP}(s)$ or $c \notin \mathsf{UP}^+(s)$;
- $\mathsf{C}[N, s, 0, (\phi \vee \psi)] = \max\{\mathsf{C}[N, s, 0, \phi], \mathsf{C}[N, s, 0, \psi]\}$;
- if $(\phi \vee \psi)$ is not a clause:
  $\mathsf{C}[N, s, 0, \neg(\phi \vee \psi)] = \min\{\mathsf{C}[N, s, 0, \neg\phi], \mathsf{C}[N, s, 0, \neg\psi]\}$;
- if $\neg\phi$ is not a clause:
  $\mathsf{C}[N, s, 0, \neg\neg\phi] = \mathsf{C}[N, s, 0, \phi]$;
- $\mathsf{C}[N, s, 0, \exists x\, \phi] = \max\{\mathsf{C}[N, s, 0, \phi^x_n] \mid n \in N\}$;
- $\mathsf{C}[N, s, 0, \neg\exists x\, \phi] = \min\{\mathsf{C}[N, s, 0, \neg\phi^x_n] \mid n \in N\}$.

Theorem 6.8.11 *Let π be proper$^+$ and let N contain the names from π and ϕ plus $l \cdot v + v$ names for $v \geq |\pi|_w$ and $v \geq |\phi|_w$. Then* $\mathsf{gnd}(\pi), l \mathrel{|\!\approx} \phi$ *iff* $\mathsf{C}[N, \mathsf{gnd}_N(\pi), l, \phi] = 1$.

The proof is also in Appendix C.1. Analogously to Theorem 6.8.7, only a finite number predicate symbols and standard names is relevant.

The complexity of deciding $\mathsf{gnd}(\pi), l \mathrel{|\!\!\approx^{\bullet}} \phi$ is essentially the same as for $\mathsf{gnd}(\pi), k \mathrel{|\!\!\approx^{\circ}} \phi$.

Theorem 6.8.12 *Let* π *be proper*$^+$. *Then* $\mathsf{gnd}(\pi), l \mathrel{|\!\!\approx^{\bullet}} \phi$ *can be decided in time* $O((|\pi| + l)^{l+1} \cdot |\phi|^{l+1} \cdot (\max\{|\pi|_{\mathrm{w}}, |\phi|_{\mathrm{w}}\} \cdot (|\pi| + |\phi| + l + 2))^{(\max\{|\pi|_{\mathrm{w}},|\phi|_{\mathrm{w}}\}+|\phi|_{\mathrm{w}})\cdot(l+1)})$.

The proof can be found in Appendix C.1. There are two differences to the complexity of deciding $\mathsf{gnd}(\pi), k \mathrel{|\!\!\approx^{\circ}} \phi$ from Theorem 6.8.8: the factor 2^k disappears since there is no splitting in $\mathrel{|\!\!\approx^{\bullet}}$, but the third factor grows since the space of relevant (non-ground) literals by which the setup is augmented is larger. In the propositional case, this third factor disappears.

Corollary 6.8.13 *Let* π *be proper*$^+$ *and propositional, and let* ϕ *also be propositional. Then* $\mathsf{gnd}(\pi), l \mathrel{|\!\!\approx^{\bullet}} \phi$ *can be decided in time* $O((|\pi| + l)^{l+1} \cdot |\phi|^{l+1})$.

6.9 A normal form

In many cases, some syntactic manipulation can help (dis)prove formulas in $\mathrel{|\!\!\approx^{\circ}}$ and $\mathrel{|\!\!\approx^{\bullet}}$. For example, clearly $\{[P, Q]\}, 0 \mathrel{|\!\!\approx^{\circ}} (P \vee Q)$ holds by subsumption, but $\{[P, Q]\}, 0 \not\mathrel{|\!\!\approx^{\circ}} (P \vee \neg\neg Q)$. Generally, $\mathrel{|\!\!\approx^{\circ}}$ prefers clauses, and hence a straightforward way to obtain more inferences is to eliminate double negations and pull quantifiers out of clauses. Analogously, $\mathrel{|\!\!\approx^{\bullet}}$ prefers negated clauses. We hence propose a normal form that is similar to prenex negation negation normal form, but does not push negations inside clauses or pull quantifiers out of non-clauses.

Definition 6.9.1 We let $\mathsf{NF}[\phi]$ denote the result of eliminating all double negations in ϕ and pulling all quantifiers out of disjunctions. More precisely, first ϕ is rewritten in a preprocessing step so that every variables is bound by only one quantifier. For ϕ of this form, $\mathsf{NF}[\phi]$ is defined as follows, where ℓ ranges over literals and c over clauses:

- $\mathsf{NF}[\ell] = \ell$;

- $$\mathsf{NF}[(\phi_1 \vee \phi_2)] = \begin{cases} \S\vec{x}_1\,\S\vec{x}_2\,(c_1 \vee c_2) & \text{if } \S\vec{x}_i \text{ is a word over } \{\neg, \exists x_{i1}, \exists x_{i2}, \ldots\} \\ & \text{with an even number of } \neg\text{, and} \\ & \quad - \ \mathsf{NF}[\phi_i] = \S\vec{x}_i\, c_i\text{, or} \\ & \quad - \ \mathsf{NF}[\phi_i] = \S'\vec{x}_i\, a_i \text{ where} \\ & \quad\quad \S\vec{x}_i = \S'\vec{x}_i\,\neg \text{ and } c_i = [\neg a_i] \\ & \quad\quad \text{for some atom } a_i; \\ (\mathsf{NF}[\phi_1] \vee \mathsf{NF}[\phi_2]) & \text{otherwise;} \end{cases}$$

- $\mathsf{NF}[\neg(\phi_1 \lor \phi_2)] = \neg\mathsf{NF}[(\phi_1 \lor \phi_2)]$;
- $\mathsf{NF}[\neg\neg\phi] = \mathsf{NF}[\phi]$;
- $\mathsf{NF}[\exists x\, \phi] = \exists x\, \mathsf{NF}[\phi]$;
- $\mathsf{NF}[\neg\exists x\, \phi] = \neg\exists x\, \mathsf{NF}[\phi]$.

The intuition behind the first case for $(\phi_1 \lor \phi_2)$ is to pull out quantifiers out of clauses. We need to make sure the quantifiers involve an even number of negations. Otherwise we would negate the other disjunct. An exception is when an odd number of quantifiers is followed by an atom; then we can simply append a negation to the sequence of quantifiers to make it even, and negate the atom, which still is clause.

For example, $\mathsf{NF}[(\exists x\, P(x) \supset \forall x'\, Q(x'))]$ stands for $\mathsf{NF}[(\neg\exists x\, P(x) \lor \neg\exists x' \neg Q(x'))]$ after expanding the abbreviations $\supset$ and $\forall$. The disjuncts are already in normal form, that is, $\mathsf{NF}[\neg\exists x\, P(x)] = \neg\exists x\, P(x)$ and $\mathsf{NF}[\neg\exists x' \neg Q(x')] = \neg\exists x' \neg Q(x')$. The prefix $\neg\exists x' \neg$ can be pulled out of the disjunction as is because it mentions an even number of negations. As for $\neg\exists x$, NF intuitively introduces a double negation as in $\neg\exists x \neg\neg P(x)$, and pulls out $\neg\exists x \neg$ but leaves $\neg P(x)$ inside the disjunction. Therefore, the result is $\neg\exists x \neg\neg\exists x' \neg(\neg P(x) \lor Q(x'))$, or using the abbreviation for universal quantifiers, $\forall x \forall x' (\neg P(x) \lor Q(x'))$.

As shown by the following theorems, this normal form preservers classical equivalence, and leads to "better" results in both limited semantics, that is, more proofs in $\mathrel{|\!\approx^{\circ}}$ and more disproofs in $\mathrel{|\!\approx^{\bullet}}$.

Theorem 6.9.2 $\models \phi \equiv \mathsf{NF}[\phi]$.

Proof. By simple induction on the length of ϕ. □

Theorem 6.9.3

(i) If $s, k \mathrel{|\!\approx^{\circ}} \phi$, *then* $s, k \mathrel{|\!\approx^{\circ}} \mathsf{NF}[\phi]$.

(ii) If $s, l \not\mathrel{|\!\approx^{\bullet}} \phi$, *then* $s, l \not\mathrel{|\!\approx^{\bullet}} \mathsf{NF}[\phi]$.

The proof involves multiple long inductions, including double subinductions. We give it in Appendix C.2.

6.10 Discussion

In this chapter we introduced two limited semantics for the fragment of $\mathcal{L}$ without functions, called $\mathcal{L}^-$. They complement each other in that one of them is sound and the other is complete. The former allows to draw sound inferences, and the latter facilitates sound consistency check.

In contrast to the other semantics in this thesis, the limited ones are not based on (sets of) worlds as semantic primitives. Instead, they employ sets of ground clauses, called setups, to represent explicit belief, which is not closed under logical consequence. Reasoning is done by augmenting the setup with new unit clauses and doing unit propagation.

Our approach is in line with the work on limited reasoning by Lakemeyer and Levesque (2002, 2013, 2014, 2016), Liu (2006), and Liu, Lakemeyer, and Levesque (2004). In particular, the sound semantics $\mid\!\approx^{\circ}$ is largely based on (Lakemeyer and Levesque 2014). The complete semantics $\mid\!\approx^{\bullet}$ is designed in the same spirit in order to complement $\mid\!\approx^{\circ}$.

The fundamental results from this section are soundness and completeness proofs for $\mid\!\approx^{\circ}$ and $\mid\!\approx^{\bullet}$, respectively (Theorems 6.5.1 and 6.7.1), and the decision procedures for proper$^+$ knowledge bases (Theorems 6.8.7 and 6.8.11). In the propositional case and for fixed effort parameters, these procedures are tractable (Corollaries 6.8.9 and 6.8.13). We also showed that the semantics are eventually complete and eventually sound for propositional formulas (Theorems 6.5.5 and 6.7.3), which means that for large enough effort parameters they agree with classical propositional logic.

Recently, Lakemeyer and Levesque (2016) extended the idea from (Lakemeyer and Levesque 2014) to accommodate functions. This is a considerable gain in expressivity for proper$^+$ knowledge bases: while predicates can be easily expressed with functions, the converse does not hold in proper$^+$ knowledge bases because it needs an existential quantifier to say that there always is a return value. Moreover, Skolemization can be used to express existentials. It hence seems worthwhile to investigate if our complete can be extended to the case of functions in a similar way to (Lakemeyer and Levesque 2016).

7 Limited Conditional Belief

Conditional belief in $\mathcal{BO}$ is closed under logical consequence. This reasoning power bears problems from both a practical and a philosophical standpoint. For one thing, the first-order features of the language make reasoning undecidable, and even the propositional fragment is intractable. The practical applicability of $\mathcal{BO}$ is hence limited, at best. For another, bearing human reasoning in mind, closing belief under logical consequence, which for example includes all tautologies, is just unrealistic. Philosophers refer to this as the problem of logical omniscience, typically associated with possible-worlds semantics (Hintikka 1975). In this chapter we develop a logic of limited conditional belief, called $\mathcal{BOL}$, to address both issues.

From a knowledge representation perspective, a particularly interesting class of problems in $\mathcal{BO}$ are belief entailments of the form $\mathbf{O}\Gamma \models \mathbf{B}(\phi \Rightarrow \psi)$. Such an entailment problem represents the task of querying the conditional knowledge base Γ whether the conditional $\phi \Rightarrow \psi$ is believed. In the limited variant of this problem, the operators $\mathbf{O}$ and $\mathbf{B}$ are decorated with parameters to specify the maximum allowed reasoning effort. At this point, the limited semantics from Chapter 6 comes into play.

Belief entailments in $\mathcal{BOL}$ are sound with respect to $\mathcal{BO}$ for a certain class of conditional knowledge bases, namely, a generalization of proper$^+$ to conditionals. Moreover, at the cost of completeness, we can give a decision procedure for limited entailment problems. We hence provide the foundation to develop a reasoning service for conditional beliefs in this chapter.

$\mathcal{BOL}$ stands in the tradition of the logics of limited knowledge by Liu, Lakemeyer, and Levesque (Lakemeyer and Levesque 2002, 2013, 2014, 2016; Liu 2006; Liu, Lakemeyer, and Levesque 2004). Like the objective limited semantics from Chapter 6, this chapter's content is based on (Schwering and Lakemeyer 2016). Most of the proofs for this chapter are quite technical; we defer them until Appendix D for that reason.

7.1 Approximating plausibilities and spheres

The semantics of conditional belief is more elaborate than that of indefeasible knowledge. Namely, conditional belief relies on an additional notion of plausibility, represented in $\mathcal{BO}$ by the arrangement of possible worlds in spheres. As it turns out, it is essential for limited reasoning to approximate plausibilities.

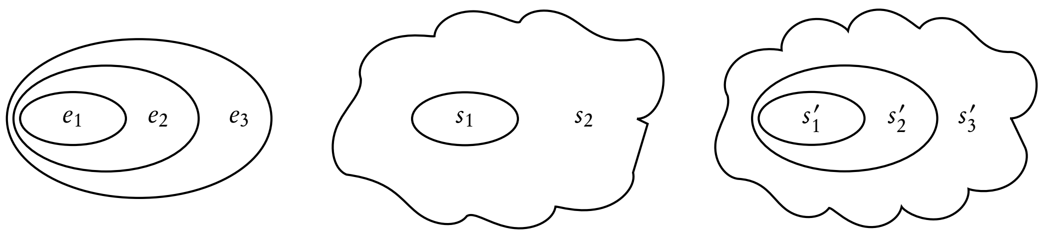
To begin with, consider the conditional belief $\mathbf{B}(\phi \Rightarrow \psi)$. Given an epistemic state as depicted in Figure 7.1a, the belief holds when the most-plausible sphere consistent with ϕ satisfies $(\phi \supset \psi)$. As locating the right sphere is already an undecidable problem, limited conditional belief cannot work with exact plausibilities.

Hence the best we can do is approximate the plausibility of ϕ (in a decidable way). Here, it is crucial not to underestimate, for then we would select a too-narrow sphere and obtain false beliefs as a result. As an example, suppose e_2 is the most-plausible sphere consistent with ϕ. If the approximation underestimates and selects e_1, then the belief always comes out true (since all worlds in e_1 satisfy $\neg\phi$ and hence also $(\phi \supset \psi)$). By contrast, approximating from *above* is sound: we might select the sphere e_3, and if all worlds from e_3 satisfy $(\phi \supset \psi)$, then it also holds in $e_2 \subseteq e_3$.

Plausibilities are also fundamental for only-believing $\mathbf{O}\{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$. In the objective case only-believing uniquely determines an epistemic state, but generating this model is still far from trivial, as illustrated by Example 4.5.5. Namely, to determine the pth sphere of the corresponding epistemic state we need to decide for which i the plausibility of ϕ_i is $\geq p$. Approximating plausibilities from above does not help with that task. For instance, if we overapproximate the plausibility of ϕ_i, only-believing would assert $(\phi_i \supset \psi_i)$ in spheres where it should not necessarily hold, which in turn could distort the plausibilities of other ϕ_j and thus eventually lead to a skewed system of spheres. The same problem arises with approximations from below.

For only-believing we hence need both, approximations from below *and* above: when the approximations from below and above *agree* on the plausibility of ϕ_i being $\geq p$ or $< p$, then we know whether $(\phi_i \supset \psi_i)$ should be asserted in the pth sphere or not. As long as both bounds agree for every i, we can faithfully represent the pth sphere in the approximation.

When the bounds are inconsistent, though, it is not clear what the pth sphere looks like. The idea is then to skip to the last sphere. In the unlimited case, the last sphere is represented by the maximal set of $(\phi_i \supset \psi_i)$ which taken together are inconsistent with the ϕ_i, that is, all have plausibility ∞. This set is approximated from below using the plausibility approximations from below.



(a) An epistemic state. (b) An approximation. (c) A better approximation.

Figure 7.1: An epistemic state and two approximations. Each ellipse represents the scenarios considered possible in that sphere. The most-plausible spheres e_1, s_1, s'_1 correspond to each other; and so do e_2, s'_2. The outermost spheres s_2 and s'_3, by contrast, are merely approximations of e_3.

Two such approximative systems of spheres are depicted in Figure 7.1: in Figure 7.1b the bounds are inconsistent already for the second sphere; Figure 7.1c faithfully represents the first two spheres, but is pessimistic about the outermost ones, that is, considers too many scenarios. It is important that the last sphere is not optimistic, for otherwise it might satisfy formulas that the last sphere of the $\vec{e}$ does not.

$\mathcal{BOL}$ employs the limited semantics from Chapter 6 and uses setups instead of sets of worlds to represent spheres semantically. Plausibilities are approximated from below and above with the sound semantics $\mathrel{|\!\!\approx}^{\circ}$ and the complete semantics $\mathrel{|\!\!\approx}^{\bullet}$, respectively. The limited belief operators **B** and **O** are decorated with sub- and superscripts $k, l \in \{0, 1, 2, \ldots\}$ to specify the reasoning efforts for $\mathrel{|\!\!\approx}^{\circ}$ and $\mathrel{|\!\!\approx}^{\bullet}$.

7.2 The language $\mathcal{BOL}$

$\mathcal{BOL}$ makes a few restrictions compared to the full logic $\mathcal{BO}$. For one thing, to simplify the technical treatment it does not allow predicates to occur outside of belief operators, and no nested beliefs. For another, like $\mathcal{L}^-$, $\mathcal{BOL}$ allows no function symbols.

Definition 7.2.1 The symbols of $\mathcal{BOL}$ are the same as for $\mathcal{L}^-$ (Definition 6.2.1) plus curly braces, $\Rightarrow$, **B**, **O**, and natural numbers $0, 1, 2, \ldots$, written schematically as k, l. The *terms* are the same as in $\mathcal{L}^-$ (Definition 6.2.1). The set of *formulas* is the least set such that

- $(t_1 = t_2)$ is a formula where t_1 and t_2 are terms;
- $\neg\alpha$, $(\alpha \vee \beta)$, and $\exists x\, \alpha$ are formulas where α and β are formulas and x is a variable;

- $\mathbf{B}^l_k(\phi_1 \Rightarrow \psi_1)$ and $\mathbf{O}^l_k\{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ are formulas if ϕ_i, ψ_i are formulas of $\mathcal{L}^-$ and $k, l \in \{0, 1, 2, \ldots\}$.

It is easy to see that every $\mathcal{BOL}$ formula after stripping the effort parameters from $\mathbf{B}$ and $\mathbf{O}$ is a formula of $\mathcal{BO}$ (Definition 4.2.1). The converse does not hold, because in $\mathcal{BOL}$ we only consider formulas of $\mathcal{BO}$ which are subjective and contain no nested beliefs and no function symbols. Nevertheless $\mathcal{BOL}$ is a meaningful fragment of $\mathcal{BO}$. In particular, it allows us to express problems like "does $\mathbf{O}^l_k\Gamma$ entail $\mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$?"

7.3 The semantics of $\mathcal{BOL}$

As mentioned before, in $\mathcal{BOL}$ setups take the place of sets of worlds in $\mathcal{BO}$. Recall that a setup is a set of (ground) clauses. Intuitively, a setup represents explicit knowledge, which is not closed under logical consequence. Nevertheless every setup corresponds to the set of compatible worlds, namely those worlds which satisfy all of its clauses.

Mind the antithetical behaviour of sets of worlds and setups when new information is added: the set of possible world shrinks, the setup grows. In light of the correspondence between setups and sets of worlds, this behaviour is not surprising: a bigger setup has fewer compatible worlds.

Definition 7.3.1 A *limited epistemic state* $\vec{s}$ is an infinite sequence of setups (Definition 6.3.1) s_p, $p \in \mathbb{P}$, that

- is concentric, that is, $\mathsf{UP}^+(s_p) \supseteq \mathsf{UP}^+(s_{p+1})$ for all $p \in \mathbb{P}$;
- converges, that is, $s_q = s_p$ for some $q \in \mathbb{P}$ and all $p \geq q$.

We use $\langle s_1, \ldots, s_q \rangle$ as a short notation for $\vec{s}$ when it converges at level q or earlier.

Note that the concentricity constraint in this definition uses the reverse subset relation as opposed to Definition 4.3.1. The reason is that, as described above, smaller setups represent less information just like bigger sets of worlds do.

Definition 7.3.2 *Plausibility approximations* from below and above are defined by

$$\lfloor \vec{s}, k \overset{\circ}{|} \phi \rfloor = \min\{p \mid p = \infty \text{ or } s_p, k \overset{\circ}{\not\models} \neg\phi\};$$
$$\lfloor \vec{s}, l \overset{\bullet}{|} \phi \rfloor = \min\{p \mid p = \infty \text{ or } s_p, l \overset{\bullet}{\not\models} \neg\phi\}.$$

We say $\vec{s}$ is ${}^l_k$*-bound-consistent* at p with respect to $\phi_1, \ldots, \phi_m$ iff for all i, $\lfloor \vec{s}, k \overset{\circ}{|} \phi_i \rfloor \geq p$ iff $\lfloor \vec{s}, l \overset{\bullet}{|} \phi_i \rfloor \geq p$. We omit "with respect to $\phi_1, \ldots, \phi_m$" when it is clear from context.

Table 7.1: The turnstile symbols used in this chapter.

$\models$	satisfaction and entailment in $\mathcal{BO}$ (Definition 4.3.2)
$\mid\!\approx$	satisfaction and entailment in $\mathcal{BOL}$ (Definition 7.3.4)
$\mid\!\overset{\circ}{\approx}$	satisfaction in sound semantics of $\mathcal{L}^-$ (Definition 6.4.2)
$\mid\!\overset{\bullet}{\approx}$	satisfaction in complete semantics of $\mathcal{L}^-$ (Definition 6.6.3)

We define the ^l_k*-approximation* with respect to $\phi_1, \ldots, \phi_m$ of $\vec{s} = \langle s_1, \ldots, s_q \rangle$ as

$$\vec{s}\,|^l_k = \begin{cases} \langle s_1, \ldots, s_p, s_q \rangle & \text{if } \vec{s} \text{ is } {}^l_k\text{-bound-consistent at } 1, \ldots, p \text{ but not at } p+1; \\ \vec{s} & \text{otherwise.} \end{cases}$$

The approximation from below ($\lfloor \vec{s}, k \mathbin{\overset{\circ}{\dagger}} \phi \rfloor$) and above ($\lfloor \vec{s}, l \mathbin{\overset{\bullet}{\dagger}} \phi \rfloor$) use the sound ($\mid\!\overset{\circ}{\approx}$) and complete ($\mid\!\overset{\bullet}{\approx}$) semantics from Chapter 6. For the intuition behind ^l_k-bound-consistency and $\vec{s}\,|^l_k$ recall our sketch of how a system of spheres could be approximated: as long as the approximation from below and above agree on whether the plausibilities of ϕ_i are $\geq p$ or $< p$, the pth sphere can genuinely represent the pth sphere of the reference system of spheres in $\mathcal{BO}$; once we are beyond this point, the approximative system of spheres needs to skip to the last sphere. Assuming $\vec{s}$ was constructed in a fashion analogous to Rule $\mathcal{BO}7$, $\vec{s}\,|^l_k$ captures the idea of taking all spheres for which ^l_k-bound-consistency is given but thereafter skipping to the end.

The following lemma tells us that the plausibilities are well-behaved: increasing the effort does not worsen their quality. In particular, it means that ^l_k-bound-consistency always implies $^{l'}_{k'}$-bound-consistency for larger effort $k' \geq k$, $l' \geq l$.

Lemma 7.3.3 *Let ϕ be a formula of $\mathcal{L}^-$.*
Then $\lfloor \vec{s}, k \mathbin{\overset{\circ}{\dagger}} \phi \rfloor \leq \lfloor \vec{s}, k+1 \mathbin{\overset{\circ}{\dagger}} \phi \rfloor \leq \lfloor \vec{s}, l+1 \mathbin{\overset{\bullet}{\dagger}} \phi \rfloor \leq \lfloor \vec{s}, l \mathbin{\overset{\bullet}{\dagger}} \phi \rfloor$.

Proof. For the first inequality, suppose $s_p, k+1 \not\mid\!\overset{\circ}{\approx} \neg\phi$. By Lemma 6.5.4, $s_p, k \not\mid\!\overset{\circ}{\approx} \neg\phi$. For the second inequality, suppose $s_p, l+1 \not\mid\!\overset{\bullet}{\approx} \neg\phi$. By Theorem 6.7.1, $s_p \not\models \neg\phi$. By Theorem 6.5.1, $s, k+1 \not\mid\!\overset{\circ}{\approx} \neg\phi$. For the third inequality, suppose $s_p, l \not\mid\!\overset{\bullet}{\approx} \neg\phi$. By Lemma 6.7.2, $s_p, l+1 \not\mid\!\overset{\bullet}{\approx} \neg\phi$. □

Now we can define the semantics of $\mathcal{BOL}$. To distinguish it from the semantics of unlimited $\mathcal{BO}$, we denote the satisfaction relation by $\mid\!\approx$. Recall that a setup that satisfies a certain formula is UP^+-minimal iff there is no smaller (modulo UP^+) model of that formula (Definition 6.8.1).

Definition 7.3.4 The truth relation $\mid\!\approx$ of $\mathcal{BOL}$ is defined with respect to a limited epistemic state $\vec{s}$:

$\mathcal{BOL}1$. $\vec{s} \mathrel{|\!\approx} (n_1 = n_2)$ iff n_1 and n_2 are identical names;

$\mathcal{BOL}2$. $\vec{s} \mathrel{|\!\approx} (\alpha \vee \beta)$ iff $\vec{s} \mathrel{|\!\approx} \alpha$ or $\vec{s} \mathrel{|\!\approx} \beta$;

$\mathcal{BOL}3$. $\vec{s} \mathrel{|\!\approx} \neg\alpha$ iff $\vec{s} \mathrel{|\!\not\approx} \alpha$;

$\mathcal{BOL}4$. $\vec{s} \mathrel{|\!\approx} \exists x\, \alpha$ iff $\vec{s} \mathrel{|\!\approx} \alpha^x_n$ for some name n;

$\mathcal{BOL}5$. $\vec{s} \mathrel{|\!\approx} \mathbf{B}^l_k(\phi \Rightarrow \psi)$ iff for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{s}, l \dagger \phi \rfloor$, then $s_p, k \mathrel{|\!\overset{\circ}{\approx}} (\phi \supset \psi)$;

$\mathcal{BOL}6$. $\vec{s} \mathrel{|\!\approx} \mathbf{O}^l_k\{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ iff
for all $p \in \mathbb{P}$, s'_p is UP^+-minimal with $s'_p, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s}', k \dagger \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)]$, and
$\vec{s}'\,|^l_k = \vec{s}$ for some $\vec{s}'$.

Rule $\mathcal{BOL}5$ approximates the plausibility of ϕ from *above*, which prevents us from selecting a too-plausible spheres that is actually inconsistent with ϕ, and then applies sound inference. That way, $\mathbf{B}^l_k(\phi \Rightarrow \psi)$ is a conservative variant of $\mathcal{BO}$'s conditional belief operator $\mathbf{B}(\phi \Rightarrow \psi)$.

The spirit behind Rule $\mathcal{BOL}6$ is the same. The intuition is to build up the system of spheres as long as the lower and upper bound of all plausibilities are consistent. Once they are not, it is unclear how the next sphere should look like, so we skip to the last one. That last sphere is determined by conditionals which (mutually) contradict their premises, so there is no scenario where any of them could be true. The parameters k and l determine how much effort is put into checking the plausibility bound-consistency. Note that there may be conditionals in Γ with unsatisfiable antecedents which do not occur in the last sphere: we only take those conditionals whose antecedents can be proved unsatisfiable by sound reasoning; otherwise the outermost sphere could be too strong. Figure 7.1 illustrates such approximations: when $\vec{e} \models \mathbf{O}\Gamma$ and $\vec{s} \mathrel{|\!\approx} \mathbf{O}^l_k\Gamma$, then the first spheres of $\vec{s}$ correspond to the respective spheres of $\vec{e}$, but the last sphere of $\vec{s}$ is possibly weaker than the last sphere of $\vec{e}$, and between these some spheres of $\vec{e}$ may have no counterpart in $\vec{s}$ for they were skipped due to ${}^l_k$-bound-inconsistency.

Notice that Rule $\mathcal{BOL}6$ converts the formula to NF from Definition 6.9.1. The rationale is that $\mathrel{|\!\overset{\circ}{\approx}}$ is particularly good at dealing with clauses, and often simple symbol pushing can increase the number of clauses in $\bigwedge_i(\phi_i \supset \psi_i)$. For example, $(\exists x P(x) \supset \forall x' Q(x'))$ clearly is not a clause, but the NF of this formula is $\forall x \forall x'(\neg P(x) \vee Q(x'))$.

7.4 Soundness for proper$^+$ knowledge bases

From a knowledge representation perspective, belief implications $\mathbf{O}^l_k\Gamma \approx\models \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$ are perhaps the most important class of reasoning problems in $\mathcal{BOL}$. In this section we prove a couple of useful properties of belief implications for proper$^+$ knowledge bases. Most notably, this includes a soundness result with respect to the unlimited $\mathcal{BO}$.

Proper$^+$ sentences are required to be in clausal form and mention no existentials (Lakemeyer and Levesque 2002). First, we generalize our original Definition 6.8.1 to belief conditionals.

Definition 7.4.1 A set of conditionals $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_1 \Rightarrow \psi_m\}$ is *proper*$^+$ when $\mathsf{NF}[\bigwedge_i(\phi_i \supset \psi_i)]$ is proper$^+$ in the sense of Definition 6.8.1.

Applying NF (Definition 6.9.1) to $\bigwedge_i(\phi_i \supset \psi_i)$ significantly expands the class of proper$^+$ knowledge bases. For example, $((P \wedge Q) \supset R)$ abbreviates $\neg\neg(\neg P \vee \neg Q) \vee R$, which clearly is not proper$^+$. By contrast, $\mathsf{NF}[\neg\neg(\neg P \vee \neg Q) \vee R]$ eliminates the double negation and yields the clause $((P \vee Q) \vee R)$. That is, $\{(P \wedge Q) \Rightarrow R\}$ would not be proper$^+$ it it wasn't for NF.

We now examine proper$^+$ knowledge bases. To begin with, the unique-model property from $\mathcal{BO}$ carries over to $\mathcal{BOL}$ for proper$^+$ knowledge bases.

Theorem 7.4.2 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be proper*$^+$*. Then there is an* $\vec{s} = \langle s_1, \ldots, s_{m+1}\rangle$ *such that* $\vec{s} \models \mathbf{O}^l_k\Gamma$*, and for all* $\vec{s}\,' \approx\models \mathbf{O}^l_k\Gamma$, $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$ *for all* p.

The argument is similar to the one for the original unique-model property for $\mathcal{BO}$; additionally, it needs Theorem 6.8.4. We give the full proof in Appendix D.1.

It is also not surprising that the effort in belief entailments is monotonic, that is, beliefs are retained when we increase the reasoning effort k or l.

Theorem 7.4.3 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be proper*$^+$*.*
If $\mathbf{O}^l_k\Gamma \approx\models \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$*, then* $\mathbf{O}^{\tilde{l}}_{\tilde{k}}\Gamma \approx\models \mathbf{B}^{\tilde{l}'}_{\tilde{k}'}(\phi \Rightarrow \psi)$ *for all* $\tilde{k} \geq k, \tilde{l} \geq l, \tilde{k}' \geq k', \tilde{l}' \geq l$.

The proof can be found in Appendix D.2. The following theorem states the main result of this section, namely that belief implications in $\mathcal{BOL}$ for proper$^+$ knowledge bases are sound with respect to $\mathcal{BO}$.

Theorem 7.4.4 *Let* $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ *be proper*$^+$*.*
If $\mathbf{O}^l_k\Gamma \approx\models \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$*, then* $\mathbf{O}\Gamma \models \mathbf{B}(\phi \Rightarrow \psi)$*.*

We prove this correctness result in Appendix D.3. The intuition is as follows. By Theorem 7.4.2, we only need to consider a single model of $s \approx\models \mathbf{O}^l_k\Gamma$. All spheres of $\vec{s}$ except the last one faithfully match the corresponding spheres of $\vec{e} \models \mathbf{O}\Gamma$, and the final sphere of $\vec{s}$ is weaker than the last sphere of $\vec{e}$, as depicted in Figure 7.1. Hence, everything that can be inferred from a sphere of $\vec{s}$ by sound inference can also be inferred from the same corresponding sphere of $\vec{e}$. Since $\mathrel{|\!\overset{\circ}{\approx}}$ is sound, the claim follows.

Before we turn to decidability of $\mathcal{BOL}$, let us illustrate its workings with the conditionals from Example 4.2.2.

Example 7.4.5 Note that Γ from Example 4.2.2 is proper$^+$. Let $k \geq 1$ and $l \geq 1$, and let s_1 and s_μ be as in Example 6.4.3. Then s_1 is the first sphere of $\vec{s} \mathrel{|\!\approx} \mathbf{O}^l_k\Gamma$. To determine the next sphere, we first need to see whether $\vec{s}$ is $\frac{k}{l}$-bound-consistent at 2, that is, $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \phi \rfloor \geq 2$ iff $\lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \phi \rfloor \geq 2$ for all $\phi \Rightarrow \psi \in \Gamma$. We can reuse our results from Examples 6.4.3 and 6.6.4. For example, we have shown in Example 6.6.4 that $s_1, l \mathrel{|\!\!\not\overset{\bullet}{\approx}} \neg$Italian, so we have $\lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \text{Italian} \rfloor = 1$. Similarly, in Example 6.4.3 we have shown $s_1, k \mathrel{|\!\overset{\circ}{\approx}} \neg$Aussie, so $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \text{Aussie} \rfloor \geq 2$. That way and with Lemma 7.3.3, we obtain

- $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \text{Italian} \rfloor = 1$ and $\lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \text{Italian} \rfloor = 1$;
- $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \text{Aussie} \rfloor \geq 2$ and $\lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \text{Aussie} \rfloor \geq 2$;
- $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \textsc{true} \rfloor = 1$ and $\lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \textsc{true} \rfloor = 1$;
- $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \neg\text{Italian} \rfloor \geq 2$ and $\lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \neg\text{Italian} \rfloor \geq 2$.

The plausibilities of the last two conditionals in Example 4.2.2 are omitted, as they are vacuously ∞. Hence, $\vec{s}$ is $\frac{k}{l}$-bound-consistent at 2. The conditionals with plausibility ≥ 2 determine the second sphere, so we obtain

$$\mathsf{UP}^+(s_2) = \mathsf{UP}^+(\{[\neg\text{Aussie}, \neg\text{Italian}], [\neg\text{Aussie}, \text{Eats(roo)}], [\text{Italian}, \text{Aussie}]\} \cup s_\mu).$$

It is easy to see that $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \text{Aussie} \rfloor = \lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \neg\text{Italian} \rfloor = 2$. Moreover $\lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \text{Aussie} \rfloor = \lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \neg\text{Italian} \rfloor = 2$ can be shown by adding Aussie to the setup. So for the final sphere s_3 we have

$$\mathsf{UP}^+(s_3) = \mathsf{UP}^+(s_\mu).$$

Having the model of $\mathbf{O}^l_k\Gamma$, we can easily prove the entailment $\mathbf{O}^l_k\Gamma \mathrel{|\!\approx} \mathbf{B}^{l'}_{k'}(\neg\text{Italian} \Rightarrow \neg\text{Veggie})$ for $k' \geq 1$, $l' \geq 1$: since $\lfloor \vec{s}, l' \mathrel{\overset{\bullet}{|}} \neg\text{Italian} \rfloor = 2$, we only need to show $s_2, k' \mathrel{|\!\overset{\circ}{\approx}}$ Italian $\vee \neg$Veggie, which is easy by splitting Italian.

Note that for $k = 0$ or $l = 0$, the model of $\mathbf{O}^l_k\Gamma$ would have consisted of s_1 followed immediately by s_μ, because of $\frac{l}{k}$-bound-inconsistency at 2. In this case, no k' or l' would have been large enough to show $\mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$.

7.5 Decision procedure for proper$^+$ knowledge bases

In Chapter 6 we saw that reasoning in $\mid\!\approx^{s}$ and $\mid\!\approx^{c}$ is decidable and in the propositional case even tractable for so-called proper$^+$ knowledge bases. As it turns out, these results carry over to belief entailments in $\mathcal{BOL}$ provided that the knowledge base is proper$^+$.

The idea behind the decision procedure is rather straightforward. By Theorem 7.4.2, the number of relevant setups is bounded. Hence Rules $\mathcal{BOL}$5 and $\mathcal{BOL}$6 can be replicated with the decision procedures for $\mid\!\approx^{s}$ and $\mid\!\approx^{c}$ from the previous chapter (Definitions 6.8.6 and 6.8.10). The following procedure generates the system of spheres that matches the one from Rule $\mathcal{BOL}$6.

Definition 7.5.1 Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be proper$^+$.
Then $\mathsf{MOD}[N, k, l, \Gamma] = \vec{s}$ is generated by:

- let $\vec{s}' = \langle s'_1, \ldots, s'_{m+1}\rangle$ such that $s'_p = \mathsf{gnd}_N(\mathsf{NF}[\bigwedge_{i:p=1 \text{ or } \mathsf{S}[N,s'_{p-1},k,\neg\phi_i]=1}(\phi_i \supset \psi_i)])$;
- let $p^\star = \max\{p \in \{1, \ldots, m\} \mid$ for some i,
$$\max\{\mathsf{S}[N, s'_{p'}, k, \neg\phi_i] \mid p' < p\} = \max\{\mathsf{C}[N, s'_{p'}, l, \neg\phi_i] \mid p' < p\}\};$$
- let $\vec{s} = \langle s'_1, \ldots, s'_{p^\star}, s'_{m+1}\rangle$.

The sequences of setups from MOD correspond to $\vec{s}'$ and $\vec{s}$ in Rule $\mathcal{BOL}$6, except that here they are grounded only with the names from N. The value of $p^\star$ represents the last sphere at which $\vec{s}'$ is $\frac{l}{k}$-bound-consistent (except in the special case $p^\star = m$, in which case $\vec{s}' = \vec{s}$). Note that the role of $\mid\!\approx^{s}$ and $\mid\!\approx^{c}$ in the bound-consistency checks is assumed by the respective decision procedures S and C (Definitions 6.8.6 and 6.8.10). Moreover recall that the first $m + 1$ setups suffice by Theorem 7.4.2.

Based on this model, we define the decision procedure for belief implications of the form $\mathbf{O}^l_k\Gamma \mid\!\approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$.

Definition 7.5.2 Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be proper$^+$.
Then $\mathsf{BEL}[N, k, l, k', l', \Gamma, \phi, \psi] = b \in \{0, 1\}$ is generated by

- let $\langle s_1, \ldots, s_j\rangle = \mathsf{MOD}[N, k, l, \Gamma]$;
- let $p^* = \min\{p \mid \mathsf{C}[N, l', s_p, \neg\phi] = 0 \text{ or } p = j\}$;
- let $b = \mathsf{S}[N, k', s_{p^*}, (\phi \supset \psi)]$.

The procedure $\mathsf{BEL}[N, k, l, k', l', \Gamma, \phi, \psi]$ is intended to determine an entailment $\mathbf{O}^l_k\Gamma \mid\!\approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$. As before, grounding happens only with the names from N. If N is chosen appropriately, then BEL is a correct decision procedure.

Definition 7.5.3 Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$. We write $|\Gamma|_\mathrm{w}$ for $|\mathsf{NF}[\bigwedge_i(\phi_i \supset \psi_i)]|_\mathrm{w}$ and $\|\Gamma\|$ for $|\bigwedge_i(\phi_i \supset \psi_i)|$.

Theorem 7.5.4 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be proper$^+$. Let N contain the names from Γ, ϕ, ψ plus $\max\{k, l, k', l'\} \cdot v + v$ names for $v \geq |\Gamma|_\mathrm{w}$ and $v \geq |\phi|_\mathrm{w}$ and $v \geq |\psi|_\mathrm{w}$. Then $\mathbf{O}^l_k \Gamma \approx\models \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$ iff $\mathsf{BEL}[N, k, l, k', l', \Gamma, \phi, \psi] = 1$.*

The proof is given in Appendix D.4. It proceeds by first showing that $\vec{s} \approx\models \mathbf{O}^l_k \Gamma$ matches $\mathsf{MOD}[N, k, l, \Gamma]$, except that it is only finitely grounded. Then it is easy to show that BEL decides the entailment.

Based on the complexities of S and C, we obtain the following complexity bound for BEL, which is exponential in the effort parameters and the width $|\Gamma|_\mathrm{w}$ and $|\phi \supset \psi|_\mathrm{w}$.

Theorem 7.5.5 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be proper$^+$. Suppose $\|\Gamma\| \geq |(\phi \supset \psi)|$. Let $j = \max\{k, l\}$ and $j' = \{k', l'\}$ and $i = \max\{j, j'\}$. Then $\mathbf{O}^l_k \Gamma \approx\models \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$ can be determined in*

$$\begin{aligned}&O(m^2 \cdot (\|\Gamma\| + j)^{2\cdot(j+1)} \cdot ((|\Gamma|_\mathrm{w} + |(\phi \supset \psi)|_\mathrm{w} + 1) \cdot (\|\Gamma\| + i + 1))^{2\cdot|\Gamma|_\mathrm{w}\cdot(j+1)} \cdot 2^k +\\ &\quad m \cdot (\|\Gamma\| + j')^{j'+1} \cdot |(\phi \supset \psi)|^{j'+1} \cdot\\ &\qquad ((|\Gamma|_\mathrm{w} + |(\phi \supset \psi)|_\mathrm{w}) \cdot (\|\Gamma\| + i + 2))^{(\max\{|\Gamma|_\mathrm{w}, |\Gamma|_\mathrm{w}\} + |(\phi\supset\psi)|_\mathrm{w})\cdot(j'+2)}).\end{aligned}$$

The first summand is due to the computation of $\mathsf{MOD}[N, k, l, \Gamma]$, which is, among others, quadratic in the number of conditionals. In many practical scenarios one would compute the model therefore only once and reuse it. The proof of the complexity bound is given in Appendix D.4. For the propositional case, the complexity is much easier to read.

Corollary 7.5.6 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be proper$^+$ and propositional, and let ϕ and ψ also be propositional. Suppose $\|\Gamma\| \geq |(\phi \supset \psi)|$. Let $j = \max\{k, l\}$ and $j' = \{k', l'\}$ and $i = \max\{j, j'\}$. Then $\mathbf{O}^l_k \Gamma \approx\models \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$ can be determined in*

$$O(m^2 \cdot (\|\Gamma\| + j)^{2\cdot(j+1)} + m \cdot (\|\Gamma\| + j')^{j+1} \cdot |(\phi \supset \psi)|^{j'+1}).$$

7.6 Discussion

In this chapter we introduced the logic of limited conditional belief $\mathcal{BOL}$. It is modelled after $\mathcal{BO}$, but uses the setup-based semantics from Chapter 6 instead of possible worlds to avoid logical omniscience. $\mathcal{BOL}$ generalizes earlier work by Liu, Lakemeyer, and Levesque to the case of conditional belief (Lakemeyer and Levesque 2002, 2013, 2014, 2016; Liu 2006; Liu, Lakemeyer, and Levesque 2004); the closest relative is arguably (Lakemeyer and Levesque 2014).

The main challenge of limited conditional belief is to approximate plausibilities appropriately. This requires not only sound but also complete first-order reasoning in order to soundly determine which sphere is certainly consistent with a conditional's premise. Consequently the limited belief operator $\mathbf{B}_k^l$ features two effort parameters: l determines the effort put into finding an appropriate sphere, and k specifies how much effort is spend on reasoning in that specific sphere.

As for limited only-believing, we employ under- and overapproximations of plausibilities to build up the system of spheres. The better these approximations, the more faithful the resulting system of spheres is to the corresponding epistemic state in $\mathcal{BO}$. For that reason, the only-believing operator $\mathbf{O}_k^l$ has effort parameters k and l for the lower and upper plausibility bounds. This is in contrast to Liu, Lakemeyer, and Levesque's work on limited knowledge, where no relevant reasoning is necessary to associate a setup with only-knowing.

For proper$^+$ knowledge bases – that is, conditionals that can be easily transformed to clausal form without existentially quantified variables – $\mathcal{BOL}$ has several attractive properties. In particular, limited belief entailments are sound with respect to their unlimited counterparts in $\mathcal{BO}$ (Theorem 7.4.4). And limited belief entailments are decidable: we presented a decision procedure (Theorem 7.5.4) and showed that in the propositional case it is even tractable for fixed effort (Corollary 7.5.6).

Next, we want to complete an ongoing implementation effort of the decision procedures presented in this chapter. We plan to evaluate the system in restricted domains first, such as games like Battleship or the German card game Skat. In the long term, we hope to deploy a limited reasoner on a robot to support the high-level control system.

A very interesting open issue on the more theoretical side is a limited variant of revision. We imagine effort-parameterized revision operators that approximate the genuine system of spheres, perhaps similarly to how limited only-believing approximates the classical only-believing. Such an approximation would be particularly useful for operators like lexicographic revision, which in their unlimited version bring along exponential growth of the system of spheres.

Limited revision operators would also help to integrate actions into $\mathcal{BOL}$. As for the physical effects of actions, we could draw on the proposal by Lakemeyer and Levesque (2014), which provides action operators like the ones in $\mathcal{ES}$ and $\mathcal{ESB}$. The semantics of their language is also based on setups and unit propagation, where literals are augmented with a sequence of actions. For example, from the unit clauses [InBox(n)]

and [Fragile(n)] together with [¬InBox(n), ¬Fragile(n), [dropbox]Broken(n)] they can infer [[dropbox]Broken(n)] simply by unit propagation. We expect this extension will carry over to $\mathcal{BOL}$ easily. It is remarkable, though, that such semantic representation of actions seems to bring about no performance gain over regression; at least our preliminary experiments with prototypical implementations suggest so. We believe the reason is the growing space of relevant split literals.

Be that as it may, a semantic account of limited actions would be interesting alone to study a limited form progression, analogously to limited revision. Here, knowledge and belief about the effect's of actions should depend on the effort as well. Similar to the envisioned limited revision, the actual effects would only be approximated. A limited form of progression appears to be an interesting question of investigation in such a model.

Whether the action model from (Lakemeyer and Levesque 2014) is the best choice, or whether action effects should be limited by effort parameter similar to the envisioned limited revision, is an open question.

Perhaps the next improvement of $\mathcal{BOL}$ is to accommodate functions. Functions are highly attractive because, for one thing, they allow more intuitive modelling than predicates in many cases, and predicates cannot imitate functions in proper$^+$ knowledge bases for the lack of existentials. Moreover, functions can be used to represent existentials in the knowledge base by means of Skolemization. As we can build on (Lakemeyer and Levesque 2016) in this issue, the main remaining task is to generalize the complete semantics $\models^{\text{\textcircled{s}}}$ from Chapter 6 appropriately.

We moreover presume the work on introspection by Lakemeyer and Levesque (2013) in limited reasoning carries over to $\mathcal{BOL}$ easily. For many applications, multi-agent limited reasoning is relevant as well.

8 Conclusion

This chapter concludes the thesis. In the first part we recapitulate the questions under consideration and summarize the answers we proposed. The second part suggests possible directions of future work.

8.1 Summary

This thesis investigated conditional belief from a knowledge-representation perspective. In particular, we studied the following three questions.

1. How can we capture the meaning of a conditional knowledge base in a semantically perspicuous way?
2. How do conditional beliefs change in the face of physical actions and new information?
3. How can reasoning about conditional beliefs be kept decidable and, sometimes, tractable?

We addressed the first question by extending Levesque's logic of only-knowing $\mathcal{OL}$ to embrace conditional beliefs. We generalized Levesque's only-knowing in order to capture conditional knowledge bases in a reasonable way. Numerous properties and results are shared among $\mathcal{OL}$ and our logic – most notably perhaps the unique-model property and the representation theorem. This confirms that our logic captures the spirit of $\mathcal{OL}$ very well while adding to its expressivity. Further, a close relation to Pearl's System Z indicates that our semantics stands to reason.

To consider the effect of actions on conditional beliefs we amalgamated our logic of only-believing with situation calculus-style actions. A notion of informing allows new information to flow into the system, which is accounted for by means of classical belief revision. The main subject of our investigation was the belief projection problem, which is to decide what is believed after a sequence of actions. Two fundamental approaches to tackle this problem are known: regression, which performs backward reasoning by rolling back the actions in the query, and progression, which reasons forward by updating the knowledge base. We showed how both techniques can be applied to conditional belief. By these results, reasoning about belief in dynamic systems can be reduced to the static case.

Finally, we dealt with the question of decidability in our framework. We devised a logic of limited conditional belief, where belief is parameterized with the effort that should be spend on proving it. Our approach is based on both a sound and a complete approximation of ordinary first-order semantics, which together allow us to approximate the notions of conditional belief and only-believing in a reasonable fashion. For a specific class of knowledge bases, this notion of limited belief is sound with respect to its unlimited archetype. Moreover, at the cost of completeness, limited reasoning is decidable. In the propositional case, the corresponding decision procedures are even tractable (for fixed effort).

8.2 Future Work

The possible directions of future work are manifold, and many open questions were discussed in the conclusions of the previous chapters. Here, we give only a brief overview of the major open challenges.

A very interesting – perhaps the most interesting – open question is how practical belief revision operators could look like. As mentioned before, operators like lexicographic revision bring along exponential growth, which is not feasible with hundreds or thousands of revisions as they will appear in a practical scenario. Limited reasoning seems to be a promising way to attack this problem. The rough idea is that the revised belief structure would be only approximated, and the differences among the least plausible scenarios would be forgotten. A model of actions where effects are known or believed only as far as the reasoning effort permits, and a corresponding notion of limited progression are related questions.

Aside from the potential concept of limited revision operators, we plan to investigate if and how our results on the projection problem carry over to other classical revision operators. Particularly the question of first-order-definable progression for such complex revision operators is open.

Finally, a couple of open questions concern the basic concept of only-believing. For one thing, the relationships to only-knowing and System Z yields numerous new relatives, and examining these relationships could be interesting. Especially investigating an amalgamation of only-believing with probabilities, in the propositional fragment or the first-order setting, appears worthwhile. From a more theoretical view, an axiom system for the logic of only-believing is desirable to get a second perspective on the logic. However, as remarked before, such an axiom system can be sound and complete only for the propositional fragment without giving up recursiveness.

A Long Proofs for $\mathcal{BO}$

A.1 Proof of the $\mathcal{OL}$ embedding theorem

Here we prove Theorem 4.6.4, the embedding of $\mathcal{OL}$ in $\mathcal{BO}$.

Lemma A.1.1 *$\vec{e} \models \mathbf{O}\{\neg\alpha \Rightarrow \textsc{false}\}$ iff for all $p \in \mathbb{P}$, $w \in e_p$ iff $\vec{e}, w \models \alpha$.*

Proof. For the only-if direction, let $\vec{e} \models \mathbf{O}\{\neg\alpha \Rightarrow \textsc{false}\}$. By Rule $\mathcal{BO}7$, for all $p \in \mathbb{P}$, $w \in e_p$ iff $\vec{e}, w \models (\neg\alpha \supset \textsc{false})$ or $\lfloor \vec{e} \mid \neg\alpha \rfloor < p$, which simplifies to $\vec{e}, w \models \alpha$ or $\lfloor \vec{e} \mid \neg\alpha \rfloor < p$ (*). We show by induction on p that $\lfloor \vec{e} \mid \neg\alpha \rfloor \geq p$ for all $p \in \mathbb{P}$, which immediately gives us the right-hand side of the lemma. The base case holds trivially. For the induction step, suppose $\lfloor \vec{e} \mid \neg\alpha \rfloor \geq p$. Then $\vec{e}, w \models \alpha$ for all $w \in e_p$ by (*), and thus $\lfloor \vec{e} \mid \neg\alpha \rfloor > p$, that is, $\lfloor \vec{e} \mid \neg\alpha \rfloor \geq p + 1$.

Conversely, let $w \in e_p$ iff $\vec{e}, w \models \alpha$ for all $p \in \mathbb{P}$. Then $\lfloor \vec{e} \mid \neg\alpha \rfloor = \infty$, and by Rule $\mathcal{BO}7$, $\vec{e} \models \mathbf{O}\{\neg\alpha \Rightarrow \textsc{false}\}$. □

For this remainder of this section, let α denote an arbitrary sentence of $\mathcal{OL}$. Recall from Definition 4.6.3 that $\sharp$ maps $\mathbf{K}\alpha$ to $\mathbf{K}\alpha^\sharp$ and $\mathbf{O}\alpha$ to $\mathbf{O}\{\neg\alpha^\sharp \Rightarrow \textsc{false}\}$.

Lemma A.1.2 *For every e and w, $e, w \models_{\mathcal{OL}} \alpha$ iff $\langle e \rangle, w \models \alpha^\sharp$.*

Proof. By induction on the length of α. Let $\vec{e} = \langle e \rangle$. We only consider the induction steps for $\mathbf{K}\alpha$ and $\mathbf{O}\alpha$ here; the other cases are trivial.

For $\mathbf{K}\alpha$, $e \models_{\mathcal{OL}} \mathbf{K}\alpha$ iff $e, w \models_{\mathcal{OL}} \alpha$ for all $w \in e$ iff (by induction) $\vec{e}, w \models \alpha^\sharp$ for all $w \in e$ iff (by construction of $\vec{e}$) $\vec{e}, w \models \alpha^\sharp$ for all $w \in e_p$ and $p \in \mathbb{P}$ iff (by Theorem 4.4.2) $\vec{e} \models \mathbf{K}\alpha^\sharp$.

For $\mathbf{O}\alpha$, $e \models_{\mathcal{OL}} \mathbf{O}\alpha$ iff $e = \{w \mid e, w \models_{\mathcal{OL}} \alpha\}$ iff (by induction) $e = \{w \mid \vec{e}, w \models \alpha^\sharp\}$ iff (by construction of $\vec{e}$) $e_p = \{w \mid \vec{e}, w \models \alpha^\sharp\}$ for all $p \in \mathbb{P}$ iff (by Lemma A.1.1) $\vec{e} \models \mathbf{O}\{\neg\alpha^\sharp \Rightarrow \textsc{false}\}$. □

This lemma showed how to construct a $\mathcal{BO}$ model from an $\mathcal{OL}$ model. Constructing an $\mathcal{OL}$ model from a $\mathcal{BO}$ model is more tricky. On the one hand, the $\mathcal{OL}$ model shall represent all worlds that occur in the $\mathcal{BO}$ model regardless of which spheres that world occurs in. On the other hand, if any sphere "misses" a world, the $\mathcal{OL}$ model shall miss one, too.

Definition A.1.3 For a primitive atom ρ, we define w_i^ρ such that $w_i^\rho[\rho] = i$ and $w_i^\rho[\tau] = w[\tau]$ for all primitive terms and atoms τ distinct from ρ.

Definition A.1.4 For arbitrary $\vec{e}$ we define $\tilde{e}^\rho = \hat{e} \cup \{w_0^\rho \mid w \in \check{e} \text{ and } w_1^\rho \notin \hat{e}\}$ where $\hat{e} = \bigcap_{p \in \mathbb{P}} e_p$ and $\check{e} = \bigcup_{p \in \mathbb{P}} e_p$.

Lemma A.1.5 *For every $\vec{e}$ and w, $\vec{e}, w \models \alpha^\sharp$ iff $\tilde{e}^\rho, w \models_{\mathcal{OL}}$, where ρ is a primitive atom whose symbol does not occur in α and $\tilde{e} = \tilde{e}^\rho$.*

Proof. By induction on the length of α. Observe that $\vec{e}, w_0^\rho \models \beta$ iff $\vec{e}, w_1^\rho \models \beta$ for any β that does not mention the symbol of ρ (*), as can be shown by a simple induction on the length of α. We only consider the induction steps for $\mathbf{K}\alpha$ and $\mathbf{O}\alpha$ here; the other cases are trivial.

For $\mathbf{K}\alpha$, $\tilde{e} \models_{\mathcal{OL}} \mathbf{K}\alpha$ iff $\tilde{e}, w \models_{\mathcal{OL}} \alpha$ for all $w \in \tilde{e}$ iff (by induction) $\vec{e}, w \models \alpha^\sharp$ for all $w \in \tilde{e}$ iff (by construction of $\tilde{e}$ and (*)) $\vec{e}, w \models \alpha^\sharp$ for all $w \in e_p$ and $p \in \mathbb{P}$ iff (by Theorem 4.4.2) $\vec{e} \models \mathbf{K}\alpha^\sharp$.

For $\mathbf{O}\alpha$ we first make the following observations. Suppose $\tilde{e} = \{w \mid \vec{e}, w \models \gamma\}$ for some γ that does not mention the symbol of ρ. First suppose there is some w such that $w \in \check{e}$ and $w_1^\rho \notin \hat{e}$. Then $w_0^\rho \in \tilde{e}$ but $w_1^\rho \notin \tilde{e}$. Contradiction to (*). Hence $\{w_0^\rho \mid w \in \check{e} \text{ and } w_1^\rho \notin \hat{e}\} = \{\}$, so we have $\hat{e} = \tilde{e}$. Next suppose there is some w such that $w \in \check{e}$ and $w_0^\rho \notin \hat{e}$. So $w_0^\rho \notin \hat{e} = \tilde{e}$, and by (*), $w_1^\rho \notin \tilde{e} = \hat{e}$. Then by construction, $w_0^\rho \in \tilde{e} = \hat{e}$. Contradiction. Hence $\check{e} \subseteq \hat{e}$. So we have $\tilde{e} = \hat{e} = \check{e}$ (**). Now for the induction step, $\tilde{e} \models_{\mathcal{OL}} \mathbf{O}\alpha$ iff $\tilde{e} = \{w \mid \tilde{e}, w \models_{\mathcal{OL}} \alpha\}$ iff (by induction) $\tilde{e} = \{w \mid \vec{e}, w \models \alpha^\sharp\}$ iff (by (**)) $\tilde{e} = \{w \mid \vec{e}, w \models \alpha^\sharp\}$ and $\tilde{e} = \hat{e} = \check{e}$ iff $\hat{e} = \{w \mid \vec{e}, w \models \alpha^\sharp\}$ and $e_p = e_{p'}$ for all $p, p' \in \mathbb{P}$ iff $e_p = \{w \mid \vec{e}, w \models \alpha^\sharp\}$ for all $p \in \mathbb{P}$ iff (by Lemma A.1.1) $\vec{e} \models \mathbf{O}\{\neg\alpha^\sharp \Rightarrow \textsc{false}\}$. ◻

Theorem 4.6.4 $\models_{\mathcal{OL}} \alpha$ *iff* $\models \alpha^\sharp$.

Proof. For the only-if direction, suppose $\models_{\mathcal{OL}} \alpha$ but $\vec{e}, w \not\models \alpha^\sharp$. Then $\vec{e}, w \models \neg\alpha^\sharp$, and by Lemma A.1.5, $\neg\alpha$ is satisfiable in $\mathcal{OL}$. Contradiction. Conversely, suppose $\models \alpha^\sharp$ but $e, w \not\models_{\mathcal{OL}} \alpha$. Then $e, w \models_{\mathcal{OL}} \neg\alpha$, and by Lemma A.1.2, $\neg\alpha^\sharp$ is satisfiable in $\mathcal{BO}$. Contradiction. ◻

A.2 Proof of the Z-ordering theorem

Here we prove Theorem 4.7.4, which states the correspondence between only-believing and Z-ordering. For this purpose let ϕ and ψ denote objective and propositional formulas as in Section 4.7. Moreover, let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be objective and propositional, and let $\vec{e} \models \mathbf{O}\Gamma$, which exists and is unique by Theorem 4.5.3.

Lemma A.2.1 *The Z-ranking is well-defined iff Γ is consistent.*

Proof. According to Pearl (1990), Γ is consistent iff $\bigcup_i \Gamma_i = \Gamma$. Thus, Γ is inconsistent iff there is some $\phi \Rightarrow \psi \notin \Gamma_i$ for all i iff $Z(\phi \Rightarrow \psi)$ is undefined for some $\phi \Rightarrow \psi \in \Gamma$. □

Lemma A.2.2 *Let $\vec{e} \models \mathbf{O}\Gamma$ and Γ be consistent.*
Then $\lfloor \vec{e} \mid \phi \rfloor = Z(\phi \Rightarrow \psi) + 1$ for every $\phi \Rightarrow \psi \in \Gamma$.

Proof. By Lemma A.2.1, Z is well-defined. We show by induction on p that $\lfloor \vec{e} \mid \phi \rfloor \leq p$ iff $Z(\phi \Rightarrow \psi) \leq p - 1$ for all $p \in \mathbb{P}$ and $\phi \Rightarrow \psi \in \Gamma$. The lemma is then an easy consequence: from $\lfloor \vec{e} \mid \phi \rfloor \leq \lfloor \vec{e} \mid \phi \rfloor$ we obtain $Z(\phi \Rightarrow \psi) \leq \lfloor \vec{e} \mid \phi \rfloor - 1$, and similarly from $Z(\phi \Rightarrow \psi) \leq Z(\phi \Rightarrow \psi) + 1 - 1$ we obtain $\lfloor \vec{e} \mid \phi \rfloor \leq Z(\phi \Rightarrow \psi) + 1$; hence the lemma holds.

For the base case, $\lfloor \vec{e} \mid \phi \rfloor \leq 1$ iff $\lfloor \vec{e} \mid \phi \rfloor = 1$ iff $w \models \phi$ for some $w \in e_1$ iff (by Rule $\mathcal{BO}7$) $w \models \phi \wedge \bigwedge_{\phi_i \Rightarrow \psi_i \in \Gamma} (\phi_i \supset \psi_i)$ for some w iff (by Lemma 4.7.2) Γ tolerates $\phi \Rightarrow \psi$ iff (by Definition 4.7.3) $\phi \Rightarrow \psi \in \Gamma_0$ iff $Z(\phi \Rightarrow \psi) = 0$ iff $Z(\phi \Rightarrow \psi) \leq 0$.

For the induction step suppose that $\lfloor \vec{e} \mid \phi \rfloor < p$ iff $Z(\phi \Rightarrow \psi) < p - 1$ for all $p \in \mathbb{P}$ and $\phi \Rightarrow \psi \in \Gamma$. First observe that $\lfloor \vec{e} \mid \phi_i \rfloor \geq p$ iff (by induction) $Z(\phi_i \Rightarrow \psi_i) \geq p - 1$ iff (by Definition 4.7.3) $\phi_i \Rightarrow \psi_i \notin \Gamma_0 \cup \ldots \cup \Gamma_{p-2}$ iff $\phi_i \Rightarrow \psi_i \in \Gamma \setminus (\Gamma_0 \cup \ldots \cup \Gamma_{p-2})$ (*). Then $\lfloor \vec{e} \mid \phi \rfloor \leq p$ iff $w \models \phi$ for some $w \in e_p$ iff (by Rule $\mathcal{BO}7$) $w \models \phi \wedge \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)$ for some w iff (by (*)) $w \models \phi \wedge \bigwedge_{\phi_i \Rightarrow \psi_i \in \Gamma \setminus (\Gamma_0 \cup \ldots \Gamma_{p-2})} (\phi_i \supset \psi_i)$ for some w iff (by Lemma 4.7.2) $\Gamma \setminus (\Gamma_0 \cup \ldots \Gamma_{p-2})$ tolerates $\phi \Rightarrow \psi$ iff (by Definition 4.7.3) $\phi \Rightarrow \psi \in \Gamma_0 \cup \ldots \cup \Gamma_{p-1}$ iff $Z(\phi \Rightarrow \psi) \leq p - 1$. □

Lemma A.2.3 *Let $\vec{e} \models \mathbf{O}\Gamma$. Then Γ is inconsistent iff $\lfloor \vec{e} \mid \phi \rfloor = \infty$ for some $\phi \Rightarrow \psi \in \Gamma$.*

Proof. For the only-if direction suppose Γ is inconsistent. Then some non-empty $\Gamma' \subseteq \Gamma$ does not tolerate any $\phi \Rightarrow \psi \in \Gamma'$. We show by induction on p that $\lfloor \vec{e} \mid \phi \rfloor > p$ for all $\phi \Rightarrow \psi \in \Gamma'$. For the base case consider $p = 1$. By Lemma 4.7.2, for every $\phi \Rightarrow \psi \in \Gamma'$, $w \not\models \phi \wedge \bigwedge_{\phi' \Rightarrow \psi' \in \Gamma'} (\phi' \supset \psi')$ for all w, so by Rule $\mathcal{BO}7$, $w \not\models \phi$ for all $w \in e_1$, and hence $\lfloor \vec{e} \mid \phi \rfloor > 1$. For the induction step consider $p > 1$. By induction, $\lfloor \vec{e} \mid \phi \rfloor \geq p$ for all $\phi \Rightarrow \psi \in \Gamma'$. Thus and by Lemma 4.7.2, for every $\phi \Rightarrow \psi \in \Gamma'$, $w \not\models \phi \wedge \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)$ for all w, so by Rule $\mathcal{BO}7$, $w \not\models \phi$ for all $w \in e_p$, and hence $\lfloor \vec{e} \mid \phi \rfloor > p$. Hence the only-if direction holds, that is, $\lfloor \vec{e} \mid \phi \rfloor = \infty$ for all $\phi \Rightarrow \psi \in \Gamma'$.

Conversely, suppose Γ is consistent. Then $Z(\phi \Rightarrow \psi) \in \{0, 1, 2, \ldots\}$ for all $\phi \Rightarrow \psi \in \Gamma$ by Definition 4.7.3, and thus $\lfloor \vec{e} \mid \phi \rfloor \neq \infty$ by Lemma A.2.2. □

Lemma A.2.4 *Let $\vec{e} \models \mathbf{O}\Gamma$ and Γ be consistent. Then* $\min\{p \mid w \in e_p\} = Z(w) + 1$.

Proof. For every world w, p is minimal such that $w \in e_p$ iff (by Rule $\mathcal{BO}7$) p is minimal such that $w \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)$ iff (by Lemma A.2.2) p is minimal such that $w \models \bigwedge_{i:Z(\phi_i \Rightarrow \psi_i) \geq p-1} (\phi_i \supset \psi_i)$ iff (by Definition 4.7.3) $Z(w) = p - 1$. □

Lemma A.2.5 *Let $\vec{e} \models \mathbf{O}\Gamma$, Γ be consistent, and ϕ satisfiable. Then* $\lfloor \vec{e} \mid \phi \rfloor = Z(\phi) + 1$.

Proof. Since ϕ is satisfiable, there is some w with $w \models \phi$, and by Lemma A.2.4, $\min\{p \mid w \in e_p\} = Z(w) + 1$. Thus, since $Z(w)$ cannot take the value ∞, $\lfloor \vec{e} \mid \phi \rfloor \neq \infty$. Hence, $\lfloor \vec{e} \mid \phi \rfloor = \min\{p \mid w \models \phi \text{ and } w \in e_p\} = \min\{\min\{p \mid w \in e_p\} \mid w \models \phi\} =$ (by Lemma A.2.4) $\min\{Z(w) + 1 \mid w \models \phi\} =$ (by Definition 4.7.3) $Z(\phi) + 1$. □

Theorem 4.7.4

(i) Γ *is inconsistent iff* $\lfloor \vec{e} \mid \phi \rfloor = \infty$ *for some* $\phi \Rightarrow \psi \in \Gamma$;

(ii) *if* Γ *is consistent, then* $\lfloor \vec{e} \mid \phi \rfloor = Z(\phi \Rightarrow \psi) + 1$ *for every* $\phi \Rightarrow \psi \in \Gamma$;

(iii) *if* Γ *is consistent, then* $\min\{p \mid w \in e_p\} = Z(w) + 1$;

(iv) *if* Γ *is consistent and* ϕ *is satisfiable, then* $\lfloor \vec{e} \mid \phi \rfloor = Z(\phi) + 1$.

Proof. Follows from Lemmas A.2.3, A.2.2, A.2.4, A.2.5, respectively. □

B Long Proofs for $\mathcal{ESB}$

B.1 Proof of the $\mathcal{BO}$ embedding theorem

Here we prove Theorem 5.3.9, the embedding of $\mathcal{BO}$ in $\mathcal{ESB}$. For this section, let α denote an arbitrary sentence of $\mathcal{BO}$, and let $\Gamma = \{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_m \Rightarrow \beta_m\}$ denote a set of $\mathcal{BO}$ conditionals.

Definition B.1.1 For a $\mathcal{BO}$ world w and an $\mathcal{ESB}$ world w', let $w \sim w'$ iff w and w' agree in the initial situation.

Lemma B.1.2 *Let w and $\vec{e}$ be of $\mathcal{BO}$. Then $\vec{e}, w \models_{\mathcal{BO}} \alpha$ iff $\vec{e}^*, w^* \models \alpha$, where w^* is arbitrary with $w \sim w^*$, and $\vec{e}^*$ is such that for every $p \in \mathbb{P}$, $e^*_p = \{w' \mid w \sim w'$ for some $w \in e_p\}$.*

Proof. By induction on the length of α, where we take the length of $\mathbf{B}(\alpha \Rightarrow \beta)$ as the length of $(\alpha \supset \beta)$ plus 1, and the length of $\mathbf{O}\Gamma$ as the length of $\bigwedge_i(\alpha_i \supset \psi_i)$ plus 1. We only consider the induction steps for $\mathbf{B}(\alpha \Rightarrow \beta)$ and $\mathbf{O}\Gamma$ here; the other cases are trivial.

First consider $\mathbf{B}(\alpha \Rightarrow \beta)$ and suppose the lemma holds for formulas shorter than $\mathbf{B}(\alpha \Rightarrow \beta)$. First notice that for every $p \in \mathbb{P}$, $\lfloor \vec{e} \,|\, \alpha \rfloor \leq p$ iff $\vec{e}, w' \models_{\mathcal{BO}} \alpha$ for some $w' \in e_p$ iff (by induction) $\vec{e}^*, w'^* \models \alpha$ for some $w' \in e_p$ and for arbitrary w'^* with $w' \sim w'^*$ iff $\vec{e}^*, w' \models \alpha$ for some $w' \in e^*_p$ iff $\lfloor \vec{e} \,|\, \alpha \rfloor \leq p$. Hence $\lfloor \vec{e} \,|\, \alpha \rfloor = \lfloor \vec{e}^* \,|\, \alpha \rfloor$ (*). Then $\vec{e}, w \models_{\mathcal{BO}} \mathbf{B}(\alpha \Rightarrow \beta)$ iff for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \,|\, \alpha \rfloor$ and $w' \in e_p$, then $\vec{e}, w \models_{\mathcal{BO}} (\alpha \supset \beta)$ iff (by induction) for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \,|\, \alpha \rfloor$ and $w' \in e_p$, then $\vec{e}^*, w'^* \models (\alpha \supset \beta)$ for arbitrary w'^* with $w' \sim w'^*$ iff (by (*)) for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e}^* \,|\, \alpha \rfloor$ and $w' \in e^*_p$, then $\vec{e}^*, w' \models (\alpha \supset \beta)$ iff $\vec{e}^*, w^* \models \mathbf{B}(\alpha \Rightarrow \beta)$.

Now consider $\mathbf{O}\Gamma$ and suppose the lemma holds for formulas shorter than $\mathbf{O}\Gamma$. By the same argument as above, $\lfloor \vec{e} \,|\, \alpha_i \rfloor = \lfloor \vec{e}^* \,|\, \alpha_i \rfloor$ (*). Then $\vec{e}, w \models_{\mathcal{BO}} \mathbf{O}\Gamma$ iff $e_p = \{w' \mid \vec{e}, w' \models_{\mathcal{BO}} \bigwedge_{i:\lfloor \vec{e} \,|\, \alpha_i \rfloor \geq p}(\alpha_i \supset \beta_i)\}$ for all $p \in \mathbb{P}$ iff (by induction) $e_p = \{w' \mid \vec{e}^*, w'^* \models \bigwedge_{i:\lfloor \vec{e} \,|\, \alpha_i \rfloor \geq p}(\alpha_i \supset \beta_i)$ for arbitrary w'^* with $w' \sim w'^*\}$ for all $p \in \mathbb{P}$ iff (by (*)) $e^*_p = \{w' \mid \vec{e}^*, w' \models \bigwedge_{i:\lfloor \vec{e}^* \,|\, \alpha_i \rfloor \geq p}(\alpha_i \supset \beta_i)\}$ for all $p \in \mathbb{P}$ iff $\vec{e}^*, w^* \models \mathbf{O}\Gamma$. □

Constructing a $\mathcal{BO}$ model from an $\mathcal{ESB}$ model is more involved because we need to avoid a pitfall similar to the problem when translating a $\mathcal{BO}$ model to an $\mathcal{OL}$ model in Appendix A.1. Given an $\mathcal{ESB}$ model such that $e_p \subsetneq \{w \mid \vec{e}, w \models \alpha\}$ for some $p \in \mathbb{P}$, when we simply "cut" the future situations from every individual world in e_p we might end up with a $\mathcal{BO}$ model $e^* = \{w^* \mid e^*, w^* \models_{\mathcal{BO}} \alpha\}$. The intuitive reason is that e_p might contain all possible truth assignments for the initial situation, but not for every future situation.

Definition B.1.3 We say a set of $\mathcal{ESB}$ worlds W is *complete* when for all $\mathcal{ESB}$ worlds w', w'', if $w' \in W$ and there is a $\mathcal{BO}$ world w such that $w \sim w'$ and $w \sim w''$, then $w'' \in W$.

Intuitively a set of $\mathcal{ESB}$ worlds is complete when for every $w' \in W$ it also contains all other worlds that agree with w' on the initial situation. The notation w_i^ρ in the following lemma was introduced in Definition A.1.3.

Lemma B.1.4 *Let w and $\vec{e}$ be of $\mathcal{ESB}$. Then $\vec{e}, w \models \alpha$ iff $\vec{e}^*, w^* \models_{\mathcal{BO}} \alpha$, where w^* is such that $w^* \sim w$, and $\vec{e}^*$ is such that for every $p \in \mathbb{P}$, if e_p is complete, $e_p^* = \{w \mid w \sim w'$ for some $w' \in e_p\}$, and otherwise $e_p^* = \{w_0^\rho \mid w \sim w'$ for some $w' \in e_p\}$.*

Proof. By induction on the length of α, where we take the length of $\mathbf{B}(\alpha \Rightarrow \beta)$ as the length of $(\alpha \supset \beta)$ plus 1, and the length of $\mathbf{O}\Gamma$ as the length of $\bigwedge_i(\alpha_i \supset \psi_i)$ plus 1. Notice that w^* is uniquely determined by w; we hence use $*$ as function that maps arbitrary $\mathcal{ESB}$ worlds w' to w'^* such that $w'^* \sim w'$. We only consider the induction steps for $\mathbf{B}(\alpha \Rightarrow \beta)$ and $\mathbf{O}\Gamma$ here; the other cases are trivial.

First consider $\mathbf{B}(\alpha \Rightarrow \beta)$ and suppose the lemma holds for formulas shorter than $\mathbf{B}(\alpha \Rightarrow \beta)$. First notice that for every $p \in \mathbb{P}$, $\lfloor \vec{e} \mid \alpha \rfloor \leq p$ iff $\vec{e}, w' \models \alpha$ for some $w' \in e_p$ iff (by induction) $\vec{e}^*, w'^* \models_{\mathcal{BO}} \alpha$ for some $w' \in e_p$ iff (since the symbol of ρ does not occur in α) $\vec{e}^*, w' \models_{\mathcal{BO}} \alpha$ for some $w' \in e_p^*$ iff $\lfloor \vec{e}^* \mid \alpha \rfloor \leq p$. Hence $\lfloor \vec{e} \mid \alpha \rfloor = \lfloor \vec{e}^* \mid \alpha \rfloor$ (*). Then $\vec{e}, w \models \mathbf{B}(\alpha \Rightarrow \beta)$ iff for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \alpha \rfloor$ and $w' \in e_p$, then $\vec{e}, w \models (\alpha \supset \beta)$ iff (by induction) for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e} \mid \alpha \rfloor$ and $w' \in e_p$, then $\vec{e}^*, w'^* \models_{\mathcal{BO}} (\alpha \supset \beta)$ iff (since the symbol of ρ does not occur in $\alpha \supset \beta$ and by (*)) for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{e}^* \mid \alpha \rfloor$ and $w' \in e_p^*$, then $\vec{e}^*, w' \models_{\mathcal{BO}} (\alpha \supset \beta)$ iff $\vec{e}^*, w^* \models_{\mathcal{BO}} \mathbf{B}(\alpha \Rightarrow \beta)$.

Now consider $\mathbf{O}\Gamma$ and suppose the lemma holds for formulas shorter than $\mathbf{O}\Gamma$. By the same argument as above, $\lfloor \vec{e} \mid \alpha_i \rfloor = \lfloor \vec{e}^* \mid \alpha_i \rfloor$ (*). We observe the following (**) for any sentence γ of $\mathcal{BO}$ that is shorter than $\mathbf{O}\Gamma$ and does not mention the symbol of ρ.

- If $w' \in e_p$ iff $\vec{e}^*, w'^* \models_{\mathcal{BO}} \gamma$, then e_p is complete and so $w' \in e_p^*$ iff $\vec{e}^*, w' \models_{\mathcal{BO}} \gamma$.
- If $w' \in e_p$ but $\vec{e}^*, w'^* \not\models_{\mathcal{BO}} \gamma$, then also $(w'^*)_i^\rho \in e_p^*$ for either $i = 0$ (when $w'^* = (w'^*)_0^\rho$ or when e_p is incomplete) or $i = 1$ (when $w'^* = (w'^*)_1^\rho$ and e_p is complete), but $\vec{e}^*, (w'^*)_i^\rho \not\models_{\mathcal{BO}} \gamma$.
- If $w' \notin e_p$ but $\vec{e}^*, w'^* \models_{\mathcal{BO}} \gamma$, then (whether e_p is complete or not) $(w'^*)_1^\rho \notin e_p^*$

although $\vec{e}^*, (w'^*)_1^\rho \models_{\mathcal{BO}} \gamma$.

Now for the actual induction step, $\vec{e}, w \models \mathbf{O}\Gamma$ iff $e_p = \{w' \mid \vec{e}, w' \models \bigwedge_{i:\lfloor \vec{e} \mid \alpha_i \rfloor \geq p}(\alpha_i \supset \beta_i)\}$ for all $p \in \mathbb{P}$ iff (by induction) $e_p = \{w' \mid \vec{e}^*, w'^* \models_{\mathcal{BO}} \bigwedge_{i:\lfloor \vec{e} \mid \alpha_i \rfloor \geq p}(\alpha_i \supset \beta_i)\}$ for all $p \in \mathbb{P}$ iff (by (*)) $e_p = \{w' \mid \vec{e}^*, w'^* \models_{\mathcal{BO}} \bigwedge_{i:\lfloor \vec{e}^* \mid \alpha_i \rfloor \geq p}(\alpha_i \supset \beta_i)\}$ for all $p \in \mathbb{P}$ iff (by (**)) $e_p^* = \{w' \mid \vec{e}^*, w' \models_{\mathcal{BO}} \bigwedge_{i:\lfloor \vec{e}^* \mid \alpha_i \rfloor \geq p}(\alpha_i \supset \beta_i)\}$ for all $p \in \mathbb{P}$ iff $\vec{e}^*, w^* \models_{\mathcal{BO}} \mathbf{O}\Gamma$. □

Theorem 5.3.9 *$\models_{\mathcal{BO}} \alpha$ iff $\models \alpha$.*

Proof. For the only-if direction suppose $\models_{\mathcal{BO}} \alpha$ and let $\vec{e}$ and w be an arbitrary $\mathcal{ESB}$ epistemic state and world, respectively. By Lemma B.1.4, $\vec{e}, w \models \alpha$ iff $\vec{e}^*, w^* \models_{\mathcal{BO}} \alpha$, where $\vec{e}^*$ and w^* are as in Lemma B.1.4. Hence by assumption, $\vec{e}, w \models \alpha$.

Conversely suppose $\models \alpha$ and let $\vec{e}$ and w be an arbitrary $\mathcal{BO}$ epistemic state and world, respectively. By Lemma B.1.2, $\vec{e}, w \models_{\mathcal{BO}} \alpha$ iff $e^*, w^* \models \alpha$, where $\vec{e}^*$ and w^* are as in Lemma B.1.2. Hence by assumption, $\vec{e}, w \models_{\mathcal{BO}} \alpha$. □

B.2 Proof of the regression theorems

Here we show the regression results. To begin with, we prove Theorems 5.5.4 and 5.5.5, which relate beliefs after an action to the beliefs before an action similar to successor-state axioms. After that, we prove the actual regression results, Theorems 5.5.3, 5.5.7, and 5.6.5.

Lemma B.2.1 *Let n be a weak-revision action standard name and $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor \neq \infty$.*

(i) If $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor = \lfloor \vec{e} \mid \mathrm{IF}(n) \wedge [n]\alpha \rfloor$, then $\lfloor \vec{e} \gg n \mid \alpha \rfloor = 1$.

(ii) If $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor \neq \lfloor \vec{e} \mid \mathrm{IF}(n) \wedge [n]\alpha \rfloor$, then $\lfloor \vec{e} \gg n \mid \alpha \rfloor = \lfloor \vec{e} \mid [n]\alpha \rfloor + 1$.

Proof. (i) By assumption $\vec{e}, w \models \mathrm{IF}(n) \wedge [n]\alpha$ for some $w \in e_{\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor}$. Thus $\vec{e}, w \models [n]\alpha$ for some $w \in (\vec{e} *_{\mathrm{w}} \mathrm{IF}(n))_1$, and so by Rule $\mathcal{ESB}$7, $\vec{e} \gg n, w \models \alpha$ for some $w \in (\vec{e} \gg n)_1$. Therefore $\lfloor \vec{e} \gg n \mid \alpha \rfloor = 1$.

(ii) By assumption, $\vec{e}, w \not\models \mathrm{IF}(n) \wedge [n]\alpha$ for all $w \in e_{\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor}$. Thus $\vec{e}, w \not\models [n]\alpha$ for all $w \in (\vec{e} *_{\mathrm{w}} \mathrm{IF}(n))_1$ (*). Thus $\vec{e} \gg n, w \not\models \alpha$ for all $w \in (\vec{e} \gg n)_1$, and hence $\lfloor \vec{e} \gg n \mid \alpha \rfloor > 1$. Now let $p \in \mathbb{P}$. First suppose $p < \lfloor \vec{e} \mid [n]\alpha \rfloor$. Then $\vec{e}, w \not\models [n]\alpha$ for all $w \in e_p$. Hence and by (*), $\vec{e}, w \not\models [n]\alpha$ for all $w \in (\vec{e} *_{\mathrm{w}} \mathrm{IF}(n))_{p+1}$. Thus $\vec{e} \gg n, w \not\models \alpha$ for all $w \in (\vec{e} \gg n)_{p+1}$, and therefore $p + 1 < \lfloor \vec{e} \gg n \mid \alpha \rfloor$. Now suppose $p \geq \lfloor \vec{e} \mid [n]\alpha \rfloor$. Then $\vec{e}, w \models [n]\alpha$ for some $w \in e_p \subseteq (\vec{e} *_{\mathrm{w}} \mathrm{IF}(n))_{p+1}$. Thus $\vec{e} \gg n, w \models \alpha$ for some $w \in (\vec{e} \gg n)_{p+1}$, and hence $p + 1 \geq \lfloor \vec{e} \gg n \mid \alpha \rfloor$. □

Lemma B.2.2 *Let n be a strong-revision action standard name and $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor \neq \infty$.*

(i) If $\lfloor \vec{e} \mid \mathrm{IF}(n) \land [n]\alpha \rfloor \neq \infty$*, then* $\lfloor \vec{e} \gg n \mid \alpha \rfloor = \lfloor \vec{e} \mid \mathrm{IF}(n) \land [n]\alpha \rfloor - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor + 1$.

(ii) If $\lfloor \vec{e} \mid \mathrm{IF}(n) \land [n]\alpha \rfloor = \infty$ *and* $\lfloor \vec{e} \mid \neg\mathrm{IF}(n) \rfloor \neq \infty$*, then* $\lfloor \vec{e} \gg n \mid \alpha \rfloor = \lfloor \vec{e} \mid [n]\alpha \rfloor + \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor - \lfloor \vec{e} \mid \neg\mathrm{IF}(n) \rfloor + 2$.

Proof. (i) Suppose $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor \leq p < \lfloor \vec{e} \mid \mathrm{IF}(n) \land [n]\alpha \rfloor$. Then $\vec{e}, w \not\models \mathrm{IF}(n) \land [n]\alpha$ for all $w \in e_p$. Thus $\vec{e}, w \not\models [n]\alpha$ for all $w \in (\vec{e} *_{\mathrm{s}} \mathrm{IF}(n))_{p - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor + 1}$. By Rule $\mathcal{ESB}7$, $\vec{e} \gg n, w \not\models \alpha$ for all $w \in (\vec{e} \gg n)_{p - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor + 1}$. Thus $p - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor + 1 < \lfloor \vec{e} \gg n \mid \alpha \rfloor$. Analogously $p \geq \lfloor \vec{e} \mid \mathrm{IF}(n) \land [n]\alpha \rfloor$ implies $p - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor + 1 \geq \lfloor \vec{e} \gg n \mid \alpha \rfloor$.

(ii) Suppose $\lfloor \vec{e} \mid \neg\mathrm{IF}(n) \rfloor \leq p < \lfloor \vec{e} \mid [n]\alpha \rfloor$. Then $\vec{e}, w \not\models [n]\alpha$ for all $w \in e_p$. So $\vec{e}, w \not\models [n]\alpha$ for all $w \in (\vec{e} *_{\mathrm{s}} \mathrm{IF}(n))_{p^*}$ where $p^* = p + \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor + 1 - \lfloor \vec{e} \mid \neg\mathrm{IF}(n) \rfloor + 1$, because by the same argument as in (i), $\vec{e} \gg n, w \not\models \alpha$ for all $w \in (\vec{e} \gg n)_{p'}$ and $p' \leq \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor + 1$. By Rule $\mathcal{ESB}7$, $\vec{e} \gg n, w \not\models \alpha$ for all $w \in (\vec{e} \gg n)_{p^*}$. Thus $p^* < \lfloor \vec{e} \gg n \mid \alpha \rfloor$. Analogously $p \geq \lfloor \vec{e} \mid [n]\alpha \rfloor$ implies $p^* \geq \lfloor \vec{e} \gg n \mid \alpha \rfloor$. □

Theorem 5.5.4 *Let a be a weak-revision action variable. Then*

$$\begin{aligned} \models \Box[a]\mathbf{B}(\alpha \Rightarrow \beta) \equiv{} & \neg\mathbf{B}(\mathrm{IF}(a) \Rightarrow \neg[a]\alpha) \land \mathbf{B}(\mathrm{IF}(a) \land [a]\alpha \Rightarrow [a]\beta) \lor \\ & \mathbf{B}(\mathrm{IF}(a) \Rightarrow \neg[a]\alpha) \land \mathbf{B}([a]\alpha \Rightarrow [a]\beta) \lor \\ & \mathbf{B}(\mathrm{IF}(a) \Rightarrow \textsc{false}). \end{aligned}$$

Proof. We prove that the equivalence holds in any epistemic state $\vec{e}$ for any weak-revision action n substituted for a. We distinguish three cases. The first case supposes $\vec{e} \not\models \mathbf{B}(\mathrm{IF}(n) \Rightarrow \neg[n]\alpha)$. The second one supposes the opposite plus $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor \neq \infty$. The third case supposes $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor = \infty$. For each case we show the equivalence. Since the cases are exhaustive, the theorem follows.

First suppose $\vec{e} \not\models \mathbf{B}(\mathrm{IF}(n) \Rightarrow \neg[a]\alpha)$. Then also $\vec{e} \not\models \mathbf{B}(\mathrm{IF}(n) \Rightarrow \textsc{false})$. Hence the equivalence to be shown reduces to $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta) \equiv \mathbf{B}(\mathrm{IF}(n) \land [n]\alpha \Rightarrow [n]\beta)$. Notice that by assumption $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor = \lfloor \vec{e} \mid \mathrm{IF}(n) \land [n]\alpha \rfloor \neq \infty$ (*), and by Lemma B.2.1 $\lfloor \vec{e} \gg n \mid \alpha \rfloor \neq \infty$ (**). Now we prove the equivalence: $\vec{e} \models \mathbf{B}(\mathrm{IF}(n) \land [n]\alpha \Rightarrow [n]\beta)$ iff (by Theorem 5.3.12 and (*)) $\vec{e}, w \models \mathrm{IF}(n) \land [n]\alpha \supset [n]\beta$ for all $w \in e_{\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor}$ iff $\vec{e}, w \models [n]\alpha \supset [n]\beta$ for all $w \in (\vec{e} *_{\mathrm{w}} \mathrm{IF}(n))_1$ iff (by (**)) $\vec{e} \gg n, w \models \alpha \supset \beta$ for all $w \in (\vec{e} \gg n)_{\lfloor \vec{e} \gg n \mid \alpha \rfloor}$ iff (by Theorem 5.3.12 and (**)) $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta)$.

Now suppose $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor \neq \infty$ and $\vec{e} \models \mathbf{B}(\mathrm{IF}(n) \Rightarrow \neg[n]\alpha)$. Then $\vec{e} \not\models \mathbf{B}(\mathrm{IF}(n) \Rightarrow \textsc{false})$. Similar to the previous case, the remaining equivalence is $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta) \equiv \mathbf{B}([n]\alpha \Rightarrow [n]\beta)$. Notice that by assumption, $\vec{e}, w \models \mathrm{IF}(n) \supset \neg[n]\alpha$ for all $w \in e_{\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor}$, so $\vec{e}, w \not\models [n]\alpha$ for all $w \in (\vec{e} *_{\mathrm{w}} \mathrm{IF}(n))_1$ (*). Now we prove the equivalence: $\vec{e} \models \mathbf{B}([n]\alpha \Rightarrow [n]\beta)$ iff $\vec{e}, w \models [n]\alpha \supset [n]\beta$ for all $w \in e_p$ for all $p \in \mathbb{P}$ with

$p \leq \lfloor \vec{e} \mid [n]\alpha \rfloor$ iff (by (*)) $\vec{e}, w \models [n]\alpha \supset [n]\beta$ for all $w \in (\vec{e} *_{\mathrm{w}} \mathrm{IF}(n))_p$ for all $p \in \mathbb{P}$ with $p \leq \lfloor \vec{e} \mid [n]\alpha \rfloor + 1$ iff (by Lemma B.2.1) $\vec{e} \gg n, w \models \alpha \supset \beta$ for all $w \in (\vec{e} \gg n)_p$ for all $p \in \mathbb{P}$ with $p \leq \lfloor \vec{e} \gg n \mid \alpha \rfloor$ iff $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta)$.

Finally suppose $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor = \infty$. Then $\vec{e}, w \not\models \mathrm{IF}(n)$ for all $p \in \mathbb{P}$ and $w \in e_p$, and so $\vec{e} \models \mathbf{B}(\mathrm{IF}(n) \Rightarrow \textsc{false})$. Since $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor = \infty$, $(\vec{e} \gg n)_p = \{\}$ for all $p \in \mathbb{P}$, and so $\vec{e} \gg n, w \models \alpha \supset \beta$ for all $w \in (\vec{e} \gg n)_p$. Thus $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta)$. $\square$

Theorem 5.5.5 *Let a be a strong-revision action variable. Then*

$$\begin{aligned}\models \square[a]\mathbf{B}(\alpha \Rightarrow \beta) \equiv\ & \neg\mathbf{B}(\mathrm{IF}(a) \wedge [a]\alpha \Rightarrow \textsc{false}) \wedge \mathbf{B}(\mathrm{IF}(a) \wedge [a]\alpha \Rightarrow [a]\beta) \vee \\ & \mathbf{B}(\mathrm{IF}(a) \wedge [a]\alpha \Rightarrow \textsc{false}) \wedge \mathbf{B}([a]\alpha \Rightarrow [a]\beta) \vee \\ & \mathbf{B}(\mathrm{IF}(a) \Rightarrow \textsc{false}).\end{aligned}$$

Proof. We prove that the equivalence holds in any epistemic state $\vec{e}$ for any strong-revision action n substituted for a. We distinguish three cases. The first case supposes $\vec{e} \not\models \mathbf{B}(\mathrm{IF}(n) \wedge \neg[n]\alpha \Rightarrow \textsc{false})$. The second one supposes the opposite plus $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor \neq \infty$. The third case supposes $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor = \infty$. For each case we show the equivalence. Since the cases are exhaustive, the theorem follows.

First suppose $\vec{e} \not\models \mathbf{B}(\mathrm{IF}(n) \wedge [n]\alpha \Rightarrow \textsc{false})$. Then also $\vec{e} \not\models \mathbf{B}(\mathrm{IF}(n) \Rightarrow \textsc{false})$. Hence the equivalence to be proved reduces to $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta) \equiv \mathbf{B}(\mathrm{IF}(n) \wedge [n]\alpha \Rightarrow [n]\beta)$. Notice that by assumption $\lfloor \vec{e} \mid \mathrm{IF}(n) \wedge [n]\alpha \rfloor \neq \infty$ (*), and by Lemma B.2.1 $\lfloor \vec{e} \gg n \mid \alpha \rfloor \neq \infty$ (**). Now we can prove the equivalence: $\vec{e} \models \mathbf{B}(\mathrm{IF}(n) \wedge [n]\alpha \Rightarrow [n]\beta)$ iff (by Theorem 5.3.12 and (*)) $\vec{e}, w \models \mathrm{IF}(n) \wedge [n]\alpha \supset [n]\beta$ for all $w \in e_{\lfloor \vec{e} \mid \mathrm{IF}(n) \wedge [n]\alpha \rfloor}$ iff $\vec{e}, w \models [n]\alpha \supset [n]\beta$ for all $w \in (\vec{e} *_{\mathrm{s}} \mathrm{IF}(n))_{\lfloor \vec{e} \mid \mathrm{IF}(n) \wedge [n]\alpha \rfloor - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor + 1}$ iff (by Lemma B.2.2) $\vec{e} \gg n, w \models \alpha \supset \beta$ for all $w \in (\vec{e} \gg n)_{\lfloor \vec{e} \gg n \mid \alpha \rfloor}$ iff (by Theorem 5.3.12 and (**)) $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta)$.

Now suppose $\vec{e} \models \mathbf{B}(\mathrm{IF}(n) \wedge [n]\alpha \Rightarrow \textsc{false})$ and $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor \neq \infty$. Then $\vec{e} \not\models \mathbf{B}(\mathrm{IF}(n) \Rightarrow \textsc{false})$. Hence the equivalence left to be shown is $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta) \equiv \mathbf{B}([n]\alpha \Rightarrow [n]\beta)$. Notice that by assumption, $\vec{e}, w \not\models \mathrm{IF}(n) \wedge [n]\alpha$ for all $w \in e_p$ and $p \in \mathbb{P}$ (*). Thus also $\vec{e} \gg n, w \not\models \alpha$ for all $w \in (\vec{e} \gg n)_p$ for all $p \in \mathbb{P}$ with $p \leq \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor + 1$ (**). Now we prove the equivalence. If $\lfloor \vec{e} \mid \neg\mathrm{IF}(n) \rfloor = \infty$, then there are no $\neg\mathrm{IF}(n)$-worlds in $\vec{e}$, so $\vec{e} \models \mathbf{B}([n]\alpha \Rightarrow [n]\beta)$ holds by (*) and $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta)$ holds by (**). Otherwise the equivalence is shown as follows: $\vec{e} \models \mathbf{B}([n]\alpha \Rightarrow [n]\beta)$ iff $\vec{e}, w \models [n]\alpha \supset [n]\beta$ for all $w \in e_p$ for all $p \in \mathbb{P}$ with $p \leq \lfloor \vec{e} \mid [n]\alpha \rfloor$ iff (by (**)) $\vec{e}, w \models [n]\alpha \supset [n]\beta$ for all $w \in (\vec{e} *_{\mathrm{s}} \mathrm{IF}(n))_p$ for all $p \in \mathbb{P}$ with $p \leq \lfloor \vec{e} \mid [n]\alpha \rfloor + \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor - \lfloor \vec{e} \mid \neg\mathrm{IF}(n) \rfloor + 2$ iff (by Lemma B.2.2) $\vec{e} \gg n, w \models \alpha \supset \beta$ for all $w \in (\vec{e} \gg n)_p$ for all $p \in \mathbb{P}$ with $p \leq \lfloor \vec{e} \gg n \mid \alpha \rfloor$ iff $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta)$.

Finally suppose $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor = \infty$. Then $\vec{e}, w \not\models \mathrm{IF}(n)$ for all $p \in \mathbb{P}$ and $w \in e_p$, and so $\vec{e} \models \mathbf{B}(\mathrm{IF}(n) \Rightarrow \textsc{false})$. Since $\lfloor \vec{e} \mid \mathrm{IF}(n) \rfloor = \infty$, $(\vec{e} \gg n)_p = \{\}$ for all $p \in \mathbb{P}$, and so $\vec{e} \gg n, w \models \alpha \supset \beta$ for all $w \in (\vec{e} \gg n)_p$. Thus $\vec{e} \models [n]\mathbf{B}(\alpha \Rightarrow \beta)$. □

Next, we turn to the actual regression theorems. We begin with the regression results from Section 5.5, Theorems 5.5.3 and 5.5.7. Then we generalize Theorem 5.5.7 for the extended only-believing operator to show Theorem 5.6.5.

Proving Theorems 5.5.3 and 5.5.7 follows a scheme similar to the knowledge regression proof in (Lakemeyer and Levesque 2011). Namely, we show that every world and epistemic state can be converted to one that adheres to the dynamic axioms Σ_{dyn} without changing its initial truth values. In Lemma B.2.12 we show that an epistemic state and a world satisfy a regressed sentence iff their Σ_{dyn}-compliant counterparts satisfy the non-regressed sentence. The regression theorem is then an easy consequence.

For the rest of this section, let $\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}}$ be a basic action theory over fluents $\mathcal{F}$. Recall that Σ_{dyn} contains the successor state axioms $\square[a]F(x_1, \ldots, x_k) \equiv \gamma_F$ for $F \in \mathcal{F}$, and the informed fluent axiom $\square\mathrm{IF}(a) \equiv \varphi$.

Definition B.2.3 For a world w, $w_{\Sigma_{\mathrm{dyn}}}$ is a world such that $w_{\Sigma_{\mathrm{dyn}}} \approx_{\mathcal{F} \cup \{\mathrm{IF}\}} w$ and

- $w_{\Sigma_{\mathrm{dyn}}}[F(n_1, \ldots, n_k), \langle\rangle] = w[F(n_1, \ldots, n_k), \langle\rangle]$ for all $F \in \mathcal{F}$;
- $w_{\Sigma_{\mathrm{dyn}}}[F(n_1, \ldots, n_k), z \cdot n] = 1$ iff $w_{\Sigma_{\mathrm{dyn}}} \gg z \models \gamma_F{}^{x_1 \ldots x_k\, a}_{n_1 \ldots n_k\, n}$ for all $F \in \mathcal{F}$, action sequences z, and actions n;
- $w_{\Sigma_{\mathrm{dyn}}}[\mathrm{IF}(n), z] = 1$ iff $w_{\Sigma_{\mathrm{dyn}}} \gg z \models \varphi^a_n$ for all action sequences z.

For a set of worlds W and an epistemic state $\vec{e}$, we let $W_{\Sigma_{\mathrm{dyn}}} = \{w_{\Sigma_{\mathrm{dyn}}} \mid w \in W\}$ and $\vec{e}_{\Sigma_{\mathrm{dyn}}} = \langle (e_1)_{\Sigma_{\mathrm{dyn}}}, \ldots, (e_{\lceil \vec{e} \rceil})_{\Sigma_{\mathrm{dyn}}} \rangle$.

Lemma B.2.4 $w_{\Sigma_{\mathrm{dyn}}}$ *is uniquely defined.*

Proof. Intuitively, once all values except for IF are fixed after z, the truth of γ_F and φ after z is uniquely determined as they are fluent formulas, and thus by definition also the value of F after $z \cdot n$ and of IF after z are uniquely determined. The formal proof is by straightforward induction on z and subinduction on the length of γ_F and φ. □

Lemma B.2.5 $w_{\Sigma_{\mathrm{dyn}}} \models \Sigma_{\mathrm{dyn}}$.

Proof. By definition, $w_{\Sigma_{\mathrm{dyn}}}[F(n_1, \ldots, n_k), z \cdot n] = 1$ iff $w_{\Sigma_{\mathrm{dyn}}} \gg z \models \gamma_F{}^{x_1 \ldots x_k\, a}_{n_1 \ldots n_k\, n}$, so $w_{\Sigma_{\mathrm{dyn}}} \models \square[a]F(x_1, \ldots, x_k) \equiv \gamma_F$ for all $F \in \mathcal{F}$. Analogously, $w_{\Sigma_{\mathrm{dyn}}} \models \square\mathrm{IF}(a) \equiv \varphi$. Hence $w_{\Sigma_{\mathrm{dyn}}} \models \Sigma_{\mathrm{dyn}}$. □

Lemma B.2.6 *If* $w \models \Sigma_{\mathrm{dyn}}$, *then* $w_{\Sigma_{\mathrm{dyn}}} = w$.

Proof. Suppose $w \models \Sigma_{\text{dyn}}$. Then $w \models \Box[a]F(x_1, \ldots, x_n) \equiv \gamma_F$ and $w \models \Box\text{IF}(a) \equiv \varphi$. Thus, w satisfies the conditions from Definition B.2.3: $w[F(n_1, \ldots, n_k), z \cdot n] = 1$ iff $(w \gg z \cdot n)[F(n_1, \ldots, n_k), \langle\rangle] = 1$ iff $w \gg z \models \gamma_F{}^{x_1 \ldots x_k\, a}_{n_1 \ldots n_k\, n}$ for all $F \in \mathcal{F}$; analogously, $w[\text{IF}(n), z] = 1$ iff $(w \gg z)[\text{IF}(n), \langle\rangle] = 1$ iff $w \gg z \models \varphi^a_n$. Since $w_{\Sigma_{\text{dyn}}}$ is unique by Lemma B.2.4, $w_{\Sigma_{\text{dyn}}} = w$. □

Lemma B.2.7 *Let ϕ be a fluent sentence. Then $w \models \phi$ iff $w_{\Sigma_{\text{dyn}}} \models \phi$.*

Proof. By an easy induction on the length of ϕ since $w, w_{\Sigma_{\text{dyn}}}$ agree on all initial values except perhaps for IF. □

Lemma B.2.8 *If $\vec{e} \models \mathbf{O}\Sigma_{\text{bel}}$, then $\vec{e}_{\Sigma_{\text{dyn}}} \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$.*

Proof. Let $\Sigma_{\text{bel}} = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ and $\vec{e} \models \mathbf{O}\Sigma_{\text{bel}}$. We show that $\vec{e}_{\Sigma_{\text{dyn}}} \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$. Note that by Lemma B.2.7, $\lfloor \vec{e} \mid \phi_i \rfloor = \lfloor \vec{e}_{\Sigma_{\text{dyn}}} \mid \phi_i \rfloor$ (*). Suppose $w \in (\vec{e}_{\Sigma_{\text{dyn}}})_p$. Then there is some $w' \in e_p$ such that $w'_{\Sigma_{\text{dyn}}} = w$, and $w' \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)$ iff (by Lemmas B.2.5 and B.2.7 and (*)) $w \models \Sigma_{\text{dyn}} \wedge \bigwedge_{i:\lfloor \vec{e}_{\Sigma_{\text{dyn}}} \mid \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)$. Conversely, suppose $w \models \Sigma_{\text{dyn}} \wedge \bigwedge_{i:\lfloor \vec{e}_{\Sigma_{\text{dyn}}} \mid \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)$. Then $w \in e_p$ by Rule $\mathcal{ESB}$10 and (*). By Lemma B.2.6, $w = w_{\Sigma_{\text{dyn}}} \in (\vec{e}_{\Sigma_{\text{dyn}}})_p$. □

For induction proofs about regression we introduce the following non-standard measure. Intuitively, $\|\alpha\|$ measures the length of the regressed formula $\mathcal{R}[\alpha]$ plus how many of "calls" to the regression operator it takes to determine $\mathcal{R}[\alpha]$ (not counting Rule R8).

Definition B.2.9 Let α be a regressable formula and $k \geq 0$. We define the measure $\|\alpha\|$ with respect to a basic action theory with dynamic axioms Σ_{dyn} as

- $\|[t_1] \ldots [t_k]R(t'_1, \ldots, t'_l)\| = 1$ for rigid R;
- $\|[t_1] \ldots [t_k]F(t'_1, \ldots, t'_l)\| = \begin{cases} 1 & \text{if } k = 0 \\ 1 + \|[t_1] \ldots [t_{k-1}]\gamma_F\| & \text{if } k > 0 \end{cases}$
  for fluent $F \in \mathcal{F}$;
- $\|[t_1] \ldots [t_k]\text{IF}(t)\| = 1 + \|[t_1] \ldots [t_k]\varphi\|$;
- $\|[t_1] \ldots [t_k](t'_1 = t'_2)\| = 1$;
- $\|[t_1] \ldots [t_k]\neg\alpha\| = 1 + \|[t_1] \ldots [t_k]\alpha\|$;
- $\|[t_1] \ldots [t_k](\alpha \vee \beta)\| = 1 + \|[t_1] \ldots [t_k]\alpha\| + \|[t_1] \ldots [t_k]\beta\|$;
- $\|[t_1] \ldots [t_k]\exists x\, \alpha\| = 1 + \|[t_1] \ldots [t_k]\alpha\|$;

- $\|[t_1]\dots[t_k]\mathbf{B}(\alpha \Rightarrow \beta)\| = \begin{cases} 1 + \|(\alpha \supset \beta)\| & \text{if } k = 0 \\ 1 + \|[t_1]\dots[t_{k-1}]\sigma\| & \text{if } k > 0 \end{cases}$
  where σ is the right-hand side of Theorem 5.5.4 or 5.5.5 depending on the sort of t_k.

Observe that $\|[t_1]\dots[t_k]\alpha\|$ reflects the regression operator $\mathcal{R}[\langle t_1,\dots,t_k\rangle,\alpha]$ from Definitions 5.5.2 and 5.5.6. For example, the definition $\|[t]F(t')\| = 1 + \|\gamma_F\|$ corresponds to $\mathcal{R}[[t]F(t')] = \mathcal{R}[\gamma_F{}^{x\ a}_{t'\ t}]$; similarly for the other cases. This makes $\|\cdot\|$ useful for induction proofs involving regression: the base cases are $\|[t_1]\dots[t_k]R(t'_1,\dots,t'_l)\|$ for rigid R, $\|F(t'_1,\dots,t'_l)\|$ for fluent $F \in \mathcal{F}$, and $\|[t_1]\dots[t_k](t = t')\|$, whose regression is trivial; all other cases are proved by induction.

We first need to show that $\|\cdot\|$ is a well-defined function from the regressable formulas to the natural numbers. Intuitively this is true because the right-hand sides of $\|\cdot\|$ for fluent atoms and beliefs eliminate an action or push it inside the belief, respectively, and the right-hand side for IF mentions no IF itself. Given the construction of $\|\cdot\|$ it then follows immediately that the measure for expressions on the left-hand side of the equations in Definition B.2.9 is always bigger than the measure of expressions on the right-hand side.

Lemma B.2.10 $\|\cdot\|$ *is a well-defined function from the regressable formulas to the natural numbers.*

Proof. Let $|\alpha|_\mathrm{B}$ be the nesting depth of $\mathbf{B}$ operators: $|R(t_1,\dots,t_k)|_\mathrm{B} = |F(t_1,\dots,t_k)|_\mathrm{B} = |(t = t')|_\mathrm{B} = 0$ for rigid R and fluent F; $|\neg\alpha|_\mathrm{B} = |\exists x\,\alpha|_\mathrm{B} = |[t]\alpha|_\mathrm{B} = |\alpha|_\mathrm{B}$; $|(\alpha \vee \beta)|_\mathrm{B} = \max\{|\alpha|_\mathrm{B}, |\beta|_\mathrm{B}\}$; and $|\mathbf{B}(\alpha \Rightarrow \beta)|_\mathrm{B} = 1 + \max\{|\alpha|_\mathrm{B}, |\beta|_\mathrm{B}\}$.

Let $|\alpha|_\mathrm{A}$ be as follows: $|R(t_1,\dots,t_k)|_\mathrm{A} = |F(t_1,\dots,t_k)|_\mathrm{A} = |(t = t')|_\mathrm{A} = 0$ for rigid R and fluent F; $|\neg\alpha|_\mathrm{A} = |\exists x\,\alpha|_\mathrm{A} = |\alpha|_\mathrm{A}$; $|(\alpha \vee \beta)|_\mathrm{A} = |\mathbf{B}(\alpha \Rightarrow \beta)|_\mathrm{A} = \max\{|\alpha|_\mathrm{A}, |\beta|_\mathrm{A}\}$; and $|[t]\alpha|_\mathrm{A} = 2^{|\alpha|_\mathrm{B}} + |\alpha|_\mathrm{A}$. Note that for objective ϕ, $|\phi|_\mathrm{A}$ is just the number of nested action operators in ϕ. In subjective formulas every action is additionally penalized with $|\cdot|_\mathrm{B}$.

First we show that $|[t_1]\dots[t_k]\mathbf{B}(\alpha \Rightarrow \beta)|_\mathrm{A} > |[t_1]\dots[t_{k-1}]\sigma|_\mathrm{A}$ for $k > 0$ where σ is the right-hand side of Theorem 5.5.4 or 5.5.5 (*). Let $|\mathbf{B}(\alpha \Rightarrow \beta)|_\mathrm{B} = n$. Then $|[t_1]\dots[t_k]\mathbf{B}(\alpha \Rightarrow \beta)|_\mathrm{A} = k \cdot 2^n + \max\{|\alpha|_\mathrm{A}, |\beta|_\mathrm{A}\}$. On the other hand, $|\sigma|_\mathrm{B} = n$. Hence $|[t_1]\dots[t_{k-1}]\sigma|_\mathrm{A} = (k-1)\cdot 2^n + |\sigma|_\mathrm{A}$. It is immediate from Theorems 5.5.4 and 5.5.5 that $|\sigma|_\mathrm{A} = \max\{|[t_k]\alpha|_\mathrm{A}, |[t_k]\beta|_\mathrm{A}\}$. Since $|\alpha|_\mathrm{B} \le n-1$ and $|\beta|_\mathrm{B} \le n-1$, we have $\max\{|[t_k]\alpha|_\mathrm{A}, |[t_k]\beta|_\mathrm{A}\} \le 2^{n-1} + \max\{|\alpha|_\mathrm{A}, |\beta|_\mathrm{A}\}$. Thus $|[t_1]\dots[t_{k-1}]\sigma|_\mathrm{A} \le (k-1)\cdot 2^n + 2^{n-1} + \max\{|\alpha|_\mathrm{A}, |\beta|_\mathrm{A}\}$. Since $k \cdot 2^n > (k-1)\cdot 2^n + 2^{n-1}$, (*) holds.

Now we prove the lemma by induction on $|\alpha|_\mathrm{A}$. For the base case, consider regressable

α with $|\alpha|_{\mathrm{A}} = 0$. We show that $\|\alpha\|$ is well-defined by subinduction on the length of α, where we take the length of $\mathrm{IF}(t)$ to be the length of φ plus 1 (which is well-behaved because φ contains no IF), and the length of $\mathbf{B}(\alpha \Rightarrow \beta)$ to be the length of $(\alpha \supset \beta)$ plus 1. The subinduction base cases $\|R(t'_1, \ldots, t'_l)\| = 1$ for rigid R, $\|F(t'_1, \ldots, t'_l)\| = 1$ for fluent $F \in \mathcal{F}$, and $\|(t = t')\| = 1$ are obviously well-defined. For the subinduction steps, $\|\mathrm{IF}(t)\|$ is well-defined iff $\|\varphi\|$ is well-defined, $\|\neg\alpha\|$ is well-defined iff $\|\alpha\|$ is well-defined, $\|(\alpha \vee \beta)\|$ is well-defined if $\|\alpha\|$ and $\|\beta\|$ are well-defined, $\|\exists x\, \alpha\|$ is well-defined iff $\|\alpha\|$ is well-defined, $\|\mathbf{B}(\alpha \Rightarrow \beta)\|$ is well-defined iff $\|(\alpha \supset \beta)\|$ is well-defined, all of which is the case by subinduction.

For the induction step consider α with $|\alpha|_{\mathrm{A}} = m > 0$ and suppose that $\|\beta\|$ is well defined for all regressable β with $|\beta|_{\mathrm{A}} < m$. We show that $\|\alpha\|$ is well-defined by a subinduction in the same vein as in the main base case. As for the first base case, $\|[t_1] \ldots [t_m] F(t'_1, \ldots, t'_l)\|$ for fluent $F \in \mathcal{F}$ is well-defined iff $\|[t_1] \ldots [t_{m-1}]\gamma_F\|$ is well-defined, which holds by induction since γ_F is fluent and thus mentions neither actions, beliefs, nor IF, so $|[t_1] \ldots [t_{m-1}]\gamma_F|_{\mathrm{A}} = m - 1$. The other base cases $\|[t_1] \ldots [t_m] R(t'_1, \ldots, t'_l)\| = 1$ for rigid R and $\|[t_1] \ldots [t_m](t = t')\| = 1$ are immediate. For the first subinduction step, $\|[t_1] \ldots [t_m]\mathrm{IF}(t)\|$ is well-defined iff $\|[t_1] \ldots [t_m]\varphi\|$ is well-defined, which holds by subinduction. For disjunction with $|[t_1] \ldots [t_k](\alpha \vee \beta)|_{\mathrm{A}} = m$, $\|[t_1] \ldots [t_k](\alpha \vee \beta)\|$ is well-defined iff $\|[t_1] \ldots [t_k]\alpha\|$ and $\|[t_1] \ldots [t_k]\beta\|$ are well-defined, which for α holds by induction in case $|[t_1] \ldots [t_k]\alpha|_{\mathrm{A}} < m$ and otherwise by subinduction, and likewise for β. The subinduction steps $\|[t_1] \ldots [t_k]\neg\alpha\|$ and $\|[t_1] \ldots [t_k]\exists x\, \alpha\|$ trivially hold by subinduction. For the subinduction step for beliefs, let $|[t_1] \ldots [t_k]\mathbf{B}(\alpha \Rightarrow \beta)|_{\mathrm{A}} = m$; then $\|[t_1] \ldots [t_k]\mathbf{B}(\alpha \Rightarrow \beta)\|$ is well-defined iff $\|[t_1] \ldots [t_{k-1}]\sigma\|$ is well-defined where σ is the right-hand side of Theorem 5.5.4 or 5.5.5 depending on the sort of t_k, which holds by induction since $|[t_1] \ldots [t_k]\mathbf{B}(\alpha \Rightarrow \beta)|_{\mathrm{A}} > |[t_1] \ldots [t_{k-1}]\sigma|_{\mathrm{A}}$ by (*). ◻

Since $\|\alpha\|$ is a natural number for every regressable α, we can prove properties of α by induction over $\|\alpha\|$, as we do in the next two lemmas. For example, for the induction step for a fluent atom $[t_1] \ldots [t_k] F(t'_1, \ldots, t'_l)$ after $k \geq 1$ actions, we have $\mathcal{R}[[t_1] \ldots [t_k] F(t'_1, \ldots, t'_l)] = \mathcal{R}[[t_1] \ldots [t_{k-1}]\gamma_F{}^{x_1 \ldots x_l\, a}_{t'_1 \ldots t'_l\, t_k}]$ by the definition of $\mathcal{R}$, and then use the induction assumption since $\|[t_1] \ldots [t_k] F(t'_1, \ldots, t'_l)\| > \|[t_1] \ldots [t_{k-1}]\gamma_F\|$ by Definition B.2.9.

Lemma B.2.11 *Let α be a regressable sentence. Then $\mathcal{R}[\alpha^x_n] = \mathcal{R}[\alpha]^x_n$.*

Proof. By induction on $\|\alpha\|$. For the base case let $\|\alpha\| = 1$. For rigid R and $k \geq 0$, $\mathcal{R}[([t_1] \ldots [t_k] R(t'_1, \ldots, t'_l))^x_n] = R(t'_1, \ldots, t'_l)^x_n = \mathcal{R}[[t_1] \ldots [t_k] R(t'_1, \ldots, t'_l)]^x_n$; analogously for $[t_1]\ldots[t_k](t = t')$. For fluent $F \in \mathcal{F}$, $\mathcal{R}[F(t'_1,\ldots,t'_l)^x_n] = F(t'_1,\ldots,t'_l)^x_n = \mathcal{R}[F(t'_1,\ldots,t'_l)]^x_n$.

For the induction step let $\|\alpha\| = m > 1$ and suppose the lemma holds for all β with $\|\beta\| < m$. For fluent $F \in \mathcal{F}$ and $k \geq 1$, $\mathcal{R}[([t_1]\ldots[t_k]F(t'_1,\ldots,t'_l))^x_n]$ = (by Rules R2 and R8) $\mathcal{R}[[t_1{}^x_n]\ldots[t_{k-1}{}^x_n]\gamma_F{}^{x_1\,\ldots\,x_l\,a}_{t'_1{}^x_n\,\ldots\,t'_l\,t_k{}^x_n}]$ = (since γ_F does not mention x due to rectification) $\mathcal{R}[([t_1]\ldots[t_{k-1}]\gamma_F{}^{x_1\,\ldots\,x_l\,a}_{t'_1\,\ldots\,t'_l\,t_k})^x_n]$ = (by induction) $\mathcal{R}[[t_1]\ldots[t_{k-1}]\gamma_F{}^{x_1\,\ldots\,x_l\,a}_{t'_1\,\ldots\,t'_l\,t_k}]^x_n$ = (by Rules R2 and R8) $\mathcal{R}[[t_1]\ldots[t_k]F(t'_1,\ldots,t'_l)]^x_n$. Similarly, $\mathcal{R}[([t_1]\ldots[t_k]\mathrm{IF}(t))^x_n] = \mathcal{R}[[t_1{}^x_n]\ldots[t_k{}^x_n]\mathrm{IF}(t^x_n)]$ = (by Rules R3 and R8) $\mathcal{R}[[t_1{}^x_n]\ldots[t_k{}^x_n]\varphi^a_{t^x_n}]$ = (since x does not occur in φ) $\mathcal{R}[([t_1]\ldots[t_k]\varphi^a_t)^x_n]$ = (by induction) $\mathcal{R}[[t_1]\ldots[t_k]\varphi^a_t]^x_n$ = (by Rules R3 and R8) $\mathcal{R}[\mathrm{IF}(t)]^x_n$.

For a quantifier and $k \geq 1$, $\mathcal{R}[([t_1]\ldots[t_k]\exists x'\alpha)^x_n]$ = (since x', x are distinct and by Rules R7 and R8) $\exists x'\mathcal{R}[[t_1{}^x_n]\ldots[t_k{}^x_n]\alpha^x_n]$ = (by induction) $\exists x'\mathcal{R}[[t_1]\ldots[t_k]\alpha]^x_n$ = (since x', x are distinct and by Rules R8 and R7) $\mathcal{R}[[t_1]\ldots[t_k]\exists x'\alpha]^x_n$. We omit the similar induction steps for $[t_1]\ldots[t_k]\neg\alpha$ and $[t_1]\ldots[t_k](\alpha \vee \beta)$.

For static belief, $\mathcal{R}[\mathbf{B}(\alpha \Rightarrow \beta)^x_n] = \mathcal{R}[\mathbf{B}(\alpha^x_n \Rightarrow \beta^x_n)]$ = (by Rule R9) $\mathbf{B}(\mathcal{R}[\alpha^x_n] \Rightarrow \mathcal{R}[\beta^x_n])$ = (by induction) $\mathbf{B}(\mathcal{R}[\alpha]^x_n \Rightarrow \mathcal{R}[\beta]^x_n) = \mathbf{B}(\mathcal{R}[\alpha] \Rightarrow \mathcal{R}[\beta])^x_n$ = (by Rule R9) $\mathcal{R}[\mathbf{B}(\alpha \Rightarrow \beta)]^x_n$. Finally, for beliefs after $k \geq 1$ actions, $\mathcal{R}[([t_1]\ldots[t_k]\mathbf{B}(\alpha \Rightarrow \beta))^x_n]$ = (by Rules R9 and R8, where σ is the right-hand side of Theorem 5.5.4 or 5.5.5 depending on the sort of t_k) $\mathcal{R}[[t_1{}^x_n]\ldots[t_{k-1}{}^x_n]\sigma^{x\;a}_{n\;t_k{}^x_n}] = \mathcal{R}[([t_1]\ldots[t_{k-1}]\sigma^a_{t_k})^x_n]$ = (by induction) $\mathcal{R}[[t_1]\ldots[t_{k-1}]\sigma^a_{t_k}]^x_n$ = (by Rules R9 and R8) $\mathcal{R}[[t_1]\ldots[t_k]\mathbf{B}(\alpha \Rightarrow \beta)]^x_n$. □

Lemma B.2.12 *Let α be a regressable sentence. Then $\vec{e}, w \models \mathcal{R}[\alpha]$ iff $\vec{e}_{\Sigma_{\mathrm{dyn}}}, w_{\Sigma_{\mathrm{dyn}}} \models \alpha$.*

Proof. By induction on $\|\alpha\|$. For the base case let $\|\alpha\| = 1$. For rigid R and $k \geq 0$, $w_{\Sigma_{\mathrm{dyn}}} \models [t_1]\ldots[t_k]R(t'_1,\ldots,t'_l)$ iff $w_{\Sigma_{\mathrm{dyn}}}[R(n'_1,\ldots,n'_l)] = 1$ where $n'_i = w_{\Sigma_{\mathrm{dyn}}}(t'_i)$ iff $w[R(n'_1,\ldots,n'_l)] = 1$ where $n'_i = w(t'_i)$ iff $w \models R(t'_1,\ldots,t'_l)$ iff (by Rule R1) $w \models \mathcal{R}[R(t'_1,\ldots,t'_l)]$; similarly for $[t_1]\ldots[t_k](t = t')$. For fluent $F \in \mathcal{F}$, $w_{\Sigma_{\mathrm{dyn}}} \models F(t'_1,\ldots,t'_l)$ iff $w_{\Sigma_{\mathrm{dyn}}}[F(n'_1,\ldots,n'_l),\langle\rangle] = 1$ where $n'_i = w_{\Sigma_{\mathrm{dyn}}}(t'_i)$ iff (by definition of $w_{\Sigma_{\mathrm{dyn}}}$) $w[F(n'_1,\ldots,n'_l),\langle\rangle] = 1$ where $n'_i = w(t'_i)$ iff $w \models F(t'_1,\ldots,t'_l)$ iff (by Rule R2) $w \models \mathcal{R}[F(t'_1,\ldots,t'_l)]$.

For the induction step let $\|\alpha\| = m > 1$ and suppose the lemma holds for all regressable β with $\|\beta\| < m$. For fluent $F \in \mathcal{F}$ and $k \geq 1$, $w_{\Sigma_{\mathrm{dyn}}} \models [t_1]\ldots[t_k]F(t'_1,\ldots,t'_l)$ iff (by Rule $\mathcal{ESB}$7) $(w_{\Sigma_{\mathrm{dyn}}} \gg n_1 \gg \ldots \gg n_k)[F(n'_1,\ldots,n'_l),\langle\rangle] = 1$ where $n_i = w_{\Sigma_{\mathrm{dyn}}}(t_i)$ and $n'_i = w_{\Sigma_{\mathrm{dyn}}}(t'_i)$ iff (by definition of $w_{\Sigma_{\mathrm{dyn}}}$ and Rule $\mathcal{ESB}$7) $w_{\Sigma_{\mathrm{dyn}}} \models [t_1]\ldots[t_{k-1}]$ $\gamma_F{}^{x_1\,\ldots\,x_l\,a}_{t'_1\,\ldots\,t'_l\,t_k}$ iff (by induction) $w \models \mathcal{R}[[t_1]\ldots[t_{k-1}]\gamma_F{}^{x_1\,\ldots\,x_l\,a}_{t'_1\,\ldots\,t'_l\,t_k}]$ iff (by Rules R2 and R8) $w \models \mathcal{R}[[t_1]\ldots[t_k]F(t'_1,\ldots,t'_l)]$. Similarly for IF, $w_{\Sigma_{\mathrm{dyn}}} \models [t_1]\ldots[t_k]\mathrm{IF}(t)$ iff (by Rule $\mathcal{ESB}$7) $(w_{\Sigma_{\mathrm{dyn}}} \gg n_1 \gg \ldots \gg n_k)[\mathrm{IF}(n),\langle\rangle] = 1$ where $n_i = w_{\Sigma_{\mathrm{dyn}}}(t_i)$ and

$n = w_{\Sigma_{\mathrm{dyn}}}(t)$ iff (by definition of $w_{\Sigma_{\mathrm{dyn}}}$ and Rule $\mathcal{ESB}7$) $w_{\Sigma_{\mathrm{dyn}}} \models [t_1] \ldots [t_k]\varphi^a_t$ iff (by induction) $w \models \mathcal{R}[[t_1] \ldots [t_k]\varphi^a_t]$ iff (by Rules R3 and R8) $w \models \mathcal{R}[[t_1] \ldots [t_k]\mathrm{IF}(t)]$.

For a quantifier and $k \geq 1$, $\vec{e}_{\Sigma_{\mathrm{dyn}}}, w_{\Sigma_{\mathrm{dyn}}} \models [t_1] \ldots [t_k]\exists x\, \alpha$ iff (by Rules $\mathcal{ESB}6$ and $\mathcal{ESB}7$) $\vec{e}_{\Sigma_{\mathrm{dyn}}}, w_{\Sigma_{\mathrm{dyn}}} \models ([t_1] \ldots [t_k]\alpha)^x_n$ for some standard name n iff (by induction) $\vec{e}, w \models \mathcal{R}[([t_1] \ldots [t_k]\alpha)^x_n]$ for some n iff (by Lemma B.2.11) $\vec{e}, w \models \mathcal{R}[[t_1] \ldots [t_k]\alpha]^x_n$ for some n iff (by Rule $\mathcal{ESB}6$) $\vec{e}, w \models \exists x \mathcal{R}[[t_1] \ldots [t_k]\alpha]$ iff (by Rules R7 and R8) $\vec{e}, w \models \mathcal{R}[[t_1] \ldots [t_k]\exists x\, \alpha]$. We omit the similar induction steps for $[t_1] \ldots [t_k]\neg\alpha$ and $[t_1] \ldots [t_k](\alpha \vee \beta)$.

For static belief, $\vec{e}_{\Sigma_{\mathrm{dyn}}} \models \mathbf{B}(\alpha \Rightarrow \beta)$ iff (by Theorem 5.3.12) $\lfloor \vec{e}_{\Sigma_{\mathrm{dyn}}} \,|\, \alpha \rfloor = \infty$ or $\vec{e}_{\Sigma_{\mathrm{dyn}}}, w \models (\alpha \supset \beta)$ for all $w \in (\vec{e}_{\Sigma_{\mathrm{dyn}}})_{\lfloor \vec{e}_{\Sigma_{\mathrm{dyn}}} | \alpha \rfloor}$ iff (by definition of $\vec{e}_{\Sigma_{\mathrm{dyn}}}$) $\lfloor \vec{e}_{\Sigma_{\mathrm{dyn}}} \,|\, \alpha \rfloor = \infty$ or $\vec{e}_{\Sigma_{\mathrm{dyn}}}, w_{\Sigma_{\mathrm{dyn}}} \models (\alpha \supset \beta)$ for all $w \in e_{\lfloor \vec{e}_{\Sigma_{\mathrm{dyn}}} | \alpha \rfloor}$ iff (by induction) $\lfloor \vec{e} \,|\, \mathcal{R}[\alpha] \rfloor = \infty$ or $\vec{e}, w \models \mathcal{R}[(\alpha \supset \beta)]$ for all $w \in e_{\lfloor \vec{e} | \mathcal{R}[\alpha] \rfloor}$ iff (by Rules R5 and R6) $\lfloor \vec{e} \,|\, \mathcal{R}[\alpha] \rfloor = \infty$ or $\vec{e}, w \models (\mathcal{R}[\alpha] \supset \mathcal{R}[\beta])$ for all $w \in e_{\lfloor \vec{e} | \mathcal{R}[\alpha] \rfloor}$ iff (by Theorem 5.3.12) $\vec{e} \models \mathbf{B}(\mathcal{R}[\alpha] \Rightarrow \mathcal{R}[\beta])$ iff (by Rule R9) $\vec{e} \models \mathcal{R}[\mathbf{B}(\alpha \Rightarrow \beta)]$. Finally, for belief after $k \geq 1$ actions, $\vec{e}_{\Sigma_{\mathrm{dyn}}}, w_{\Sigma_{\mathrm{dyn}}} \models [t_1] \ldots [t_k]\mathbf{B}(\alpha \Rightarrow \beta)$ iff (by Rule $\mathcal{ESB}7$ and, depending on the sort of t_k, Theorem 5.5.4 or 5.5.5, where σ is that theorem's right-hand side) $\vec{e}_{\Sigma_{\mathrm{dyn}}}, w_{\Sigma_{\mathrm{dyn}}} \models [t_1] \ldots [t_{k-1}]\sigma^a_{n_k}$ where $n_k = w_{\Sigma_{\mathrm{dyn}}}(t_k)$ iff (since by assumption action terms in formulas to be regressed only have variables or names as arguments) $\vec{e}_{\Sigma_{\mathrm{dyn}}}, w_{\Sigma_{\mathrm{dyn}}} \models [t_1] \ldots [t_{k-1}]\sigma^a_{t_k}$ iff (by induction) $\vec{e}, w \models \mathcal{R}[[t_1] \ldots [t_{k-1}]\sigma^a_{t_k}]$ iff (by Rules R9 and R8) $\vec{e}, w \models \mathcal{R}[[t_1] \ldots [t_k]\mathbf{B}(\alpha \Rightarrow \beta)]$. □

Theorem 5.5.3 *Let ϕ be a fluent sentence and ψ be an objective regressable sentence. Then* $\Sigma_{\mathrm{dyn}} \wedge \phi \models \psi$ *iff* $\phi \models \mathcal{R}[\psi]$.

Proof. For the only-if direction suppose $\Sigma_{\mathrm{dyn}} \wedge \phi \models \psi$ and $w \models \phi$. By Lemma B.2.7 and the assumption, $w_{\Sigma_{\mathrm{dyn}}} \models \psi$, and by Lemma B.2.12, $w \models \mathcal{R}[\psi]$. Conversely, suppose $\phi \models \mathcal{R}[\psi]$ and $w \models \Sigma_{\mathrm{dyn}} \wedge \phi$. Then $w \models \mathcal{R}[\psi]$ by assumption, and thus $w_{\Sigma_{\mathrm{dyn}}} \models \psi$ by Lemma B.2.12. By Lemma B.2.6, $w_{\Sigma_{\mathrm{dyn}}} = w$, so $w \models \psi$. □

Theorem 5.5.7 *Let α be a regressable sentence. Then* $\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}}) \models \alpha$ *iff* $\mathbf{O}\Sigma_{\mathrm{bel}} \models \mathcal{R}[\alpha]$.

Proof. For the only-if direction suppose $\mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}}) \models \alpha$ and $\vec{e} \models \mathbf{O}\Sigma_{\mathrm{bel}}$. Then by Lemma B.2.8, $\vec{e}_{\Sigma_{\mathrm{dyn}}} \models \mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}})$. By assumption, $\vec{e}_{\Sigma_{\mathrm{dyn}}} \models \alpha$. By Lemma B.2.12, $\vec{e} \models \mathcal{R}[\alpha]$. Conversely, suppose $\mathbf{O}\Sigma_{\mathrm{bel}} \models \mathcal{R}[\alpha]$ and $\vec{e} \models \mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}})$. Let $\vec{e}\,' \models \mathbf{O}\Sigma_{\mathrm{bel}}$. By assumption, $\vec{e}\,' \models \mathcal{R}[\alpha]$. By Lemma B.2.12, $\vec{e}\,'_{\Sigma_{\mathrm{dyn}}} \models \alpha$. By Lemma B.2.8, $\vec{e}\,'_{\Sigma_{\mathrm{dyn}}} \models \mathbf{O}(\Sigma_{\mathrm{dyn}}, \Sigma_{\mathrm{bel}})$. By Theorem 5.3.16, $\vec{e}\,'_{\Sigma_{\mathrm{dyn}}} = \vec{e}$, so $\vec{e} \models \alpha$. □

To generalize the regression result for the extended only-believing operator from Section 5.6, suppose $\mathcal{S}$ is a finite set of object function and predicate symbols and that Σ_{dyn} is $\mathcal{S}$-free from now on. It suffices to generalize Lemma B.2.8, as the other lemmas involved do not refer to only-believing.

Lemma B.2.13 *Let ϕ be objective and $\mathcal{S}$-free, and $w \approx_{\mathcal{S}} w'$. Then $w \models \phi$ iff $w' \models \phi$.*

Proof. Follows by a trivial induction on the length of ϕ. □

Lemma B.2.14 *If $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Sigma_{\text{bel}}$, then $\vec{e}_{\Sigma_{\text{dyn}}} \models \mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$.*

Proof. Let $\vec{e} \models \mathbf{O}\Sigma_{\text{bel}}$. By Lemma B.2.8, $\vec{e}_{\Sigma_{\text{dyn}}} \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$. Then $\vec{e}_{\mathcal{S}} \models \mathbf{O}_{\mathcal{S}}\Sigma_{\text{bel}}$, and $(\vec{e}_{\Sigma_{\text{dyn}}})_{\mathcal{S}} \models \mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$. We need to show that $(\vec{e}_{\mathcal{S}})_{\Sigma_{\text{dyn}}} = (\vec{e}_{\Sigma_{\text{dyn}}})_{\mathcal{S}}$. Let $w \in ((\vec{e}_{\mathcal{S}})_{\Sigma_{\text{dyn}}})_p$. Then for some $w' \in (\vec{e}_{\mathcal{S}})_p$, $w'_{\Sigma_{\text{dyn}}} = w$. Then for some $w'' \in e_p$, $w'' \approx_{\mathcal{S}} w'$, and therefore $w''_{\Sigma_{\text{dyn}}} \approx_{\mathcal{S}} w'_{\Sigma_{\text{dyn}}} = w$. Since $w''_{\Sigma_{\text{dyn}}} \in (\vec{e}_{\Sigma_{\text{dyn}}})_p$, $w \in ((\vec{e}_{\Sigma_{\text{dyn}}})_{\mathcal{S}})_p$. Conversely, let $w \in ((\vec{e}_{\Sigma_{\text{dyn}}})_{\mathcal{S}})_p$. Then for some $w' \in (\vec{e}_{\Sigma_{\text{dyn}}})_p$, $w' \approx_{\mathcal{S}} w$. Then $w' \in e_p$, and so $w \in (\vec{e}_{\mathcal{S}})_p$. By Lemma B.2.5, $w' \models \Sigma_{\text{dyn}}$. Thus by Lemma B.2.13 and since Σ_{dyn} is $\mathcal{S}$-free, $w \models \Sigma_{\text{dyn}}$. By Lemma B.2.6, $w = w_{\Sigma_{\text{dyn}}}$, so $w \in ((\vec{e}_{\mathcal{S}})_{\Sigma_{\text{dyn}}})_p$. □

Theorem 5.6.5 *Let α be a regressable sentence. Then $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models \alpha$ iff $\mathbf{O}_{\mathcal{S}}\Sigma_{\text{bel}} \models \mathcal{R}[\alpha]$.*

Proof. Proceeds by the exact same argument as the proof of Theorem 5.5.7, with $\mathbf{O}$ replaced by $\mathbf{O}_{\mathcal{S}}$, Lemma B.2.8 replaced by Lemma B.2.14, and Theorem 5.3.16 replaced by Corollary 5.6.4. □

B.3 Proof of the revision theorems

In this section we prove Theorems 5.7.3 and 5.7.5, which claim correctness of our weak and strong revision of a conditional knowledge base.

Lemma B.3.1 *If $\lfloor \vec{e} \mid \alpha \rfloor > \lceil \vec{e} \rceil$, then $\lfloor \vec{e} \mid \alpha \rfloor = \infty$.*

Proof. Suppose $\lfloor \vec{e} \mid \alpha \rfloor \neq \infty$. Then $\vec{e}, w \models \alpha$ for some $w \in e_p$ and $p \in \mathbb{P}$. Since $e_p \subseteq e_{\lceil \vec{e} \rceil}$, $\lfloor \vec{e} \mid \alpha \rfloor \leq \lceil \vec{e} \rceil$. □

Lemma B.3.2 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be objective, $\vec{e} \models \mathbf{O}\Gamma$, and let $p \in \mathbb{P} \cup \{\infty\}$ such that $p \geq \lfloor \vec{e} \mid \alpha \rfloor$.*
Then $w \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)$ iff $w \models \bigwedge_{\phi \Rightarrow \psi \in \Gamma_\alpha \text{ with } \max\{\lfloor \vec{e} \mid \phi \rfloor, \lfloor \vec{e} \mid \alpha \rfloor\} \geq p}(\phi \supset \psi)$ for all w.

Proof. By Theorems 5.3.16 and 5.3.14, $\phi \Rightarrow \psi \in \Gamma_\alpha$ iff $\lfloor \vec{e} \mid \alpha \rfloor = \lfloor \vec{e} \mid \neg(\phi \supset \psi) \rfloor = \infty$ or $\lfloor \vec{e} \mid \alpha \rfloor < \lfloor \vec{e} \mid \neg(\phi \supset \psi) \rfloor$ (*).

For the only-if direction, suppose $w \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)$ and let $\phi \Rightarrow \psi \in \Gamma_\alpha$ with $\max\{\lfloor \vec{e} \mid \phi \rfloor, \lfloor \vec{e} \mid \alpha \rfloor\} \geq p$. Then $w \in e_{\min\{p, \lceil \vec{e} \rceil\}}$ by Rule $\mathcal{ESB}$10. Note that $w' \models (\phi \supset \psi)$ for all $w' \in e_{p'}$ and $p' \in \mathbb{P}$ with $p' \leq \lfloor \vec{e} \mid \phi \rfloor$ by Rule $\mathcal{ESB}$10. Likewise, $w' \models (\phi \supset \psi)$ for all $w' \in e_{p'}$ and $p' \in \mathbb{P}$ with $p' \leq \lfloor \vec{e} \mid \alpha \rfloor$ by (*). Hence, since $p \leq \max\{\lfloor \vec{e} \mid \phi \rfloor, \lfloor \vec{e} \mid \alpha \rfloor\}$, $w \models (\phi \supset \psi)$.

Conversely, suppose $w \not\models (\phi_i \supset \psi_i)$ for some i with $\lfloor \vec{e} \mid \phi_i \rfloor \geq p$. Then trivially $\max\{\lfloor \vec{e} \mid \phi_i \rfloor, \lfloor \vec{e} \mid \alpha \rfloor\} \geq p$, so we only need to show that $\phi_i \Rightarrow \psi_i \in \Gamma_\alpha$. By Rule $\mathcal{ESB}10$, $w' \models (\phi_i \supset \psi_i)$ for all $w' \in e_{p'}$ and $p' \in \mathbb{P}$ with $p' \leq p$. Hence $p < \lfloor \vec{e} \mid \neg(\phi_i \supset \psi_i) \rfloor$. If $\lfloor \vec{e} \mid \alpha \rfloor \leq \lceil \vec{e} \rceil$, then $\lfloor \vec{e} \mid \alpha \rfloor \leq p < \lfloor \vec{e} \mid \neg(\phi_i \supset \psi_i) \rfloor$. Otherwise, if $\lceil \vec{e} \rceil < \lfloor \vec{e} \mid \alpha \rfloor$, then $\lceil \vec{e} \rceil \leq p < \lfloor \vec{e} \mid \neg(\phi_i \supset \psi_i) \rfloor$, and by Lemma B.3.1, $\lfloor \vec{e} \mid \alpha \rfloor = \lfloor \vec{e} \mid \neg(\phi_i \supset \psi_i) \rfloor = \infty$. In both cases, by (*), $\phi_i \Rightarrow \psi_i \in \Gamma_\alpha$. □

Lemma B.3.3 *Let ϕ be objective and $\mathcal{S}$-free. Then $(\vec{e} * \phi)_{\mathcal{S}} = \vec{e}_{\mathcal{S}} * \phi$.*

Proof. By Lemma B.2.13, $\lfloor \vec{e} \mid \phi \rfloor = \lfloor \vec{e}_{\mathcal{S}} \mid \phi \rfloor$ and $(\vec{e}_{\mathcal{S}} \mid \phi)_p = ((\vec{e} \mid \phi)_p)_{\mathcal{S}}$. Thus by Definition 5.3.4 the lemma follows. □

Theorem 5.7.3 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ and υ be objective and $\mathcal{S}$-free. Let R be the nullary rigid predicate newly introduced in $\Gamma *_{\mathrm{w}} \upsilon$.*
If $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$, then $\vec{e} *_{\mathrm{w}} \upsilon \models \mathbf{O}_{\mathcal{S}\cup\{R\}}\Gamma *_{\mathrm{w}} \upsilon$.

Proof. We show the theorem for the case $\mathcal{S} = \{\}$ first. Let $\vec{e} \models \mathbf{O}\Gamma$. We construct an $\vec{e}'$ such that $\vec{e}' \models \mathbf{O}\Gamma *_{\mathrm{w}} \upsilon$ and $\vec{e}'_{\{R\}} = \vec{e} *_{\mathrm{w}} \upsilon$, which gives $\vec{e} *_{\mathrm{w}} \upsilon \models \mathbf{O}_{\{R\}}\Gamma *_{\mathrm{w}} \upsilon$. If $\lfloor \vec{e} \mid \upsilon \rfloor = \infty$, we let $e'_p = \{\}$ for all $p \in \mathbb{P}$. Otherwise, we let $e'_1 = ((\vec{e} \mid \upsilon) \mid R)_{\lfloor \vec{e} \mid \upsilon \rfloor}$ and $e'_p = e'_1 \cup (\vec{e} \mid \neg R)_{p-1}$ for $p > 1$.

First suppose $\lfloor \vec{e} \mid \upsilon \rfloor = \infty$. Then $\vec{e} *_{\mathrm{w}} \upsilon = \langle\{\}\rangle$. Clearly, $\lfloor \langle\{\}\rangle \mid \phi \rfloor = \infty$ for all ϕ, so $\langle\{\}\rangle \models \mathbf{O}\Gamma *_{\mathrm{w}} \upsilon$ if (by Rule $\mathcal{ESB}10$) $\bigwedge_{\phi\Rightarrow\psi\in\Gamma_\upsilon}(R \supset (\phi \supset \psi)) \land (R \supset \upsilon) \land (\textsc{true} \supset R)$ is unsatisfiable iff $\bigwedge_{\phi\Rightarrow\psi\in\Gamma_\upsilon}(\phi \supset \psi) \land \upsilon \land R$ is unsatisfiable if (by Lemma B.3.2) $\bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor=\infty}(\phi_i \supset \psi_i) \land \upsilon$ is unsatisfiable, which holds by Rule $\mathcal{ESB}10$ and $\lfloor \vec{e} \mid \upsilon \rfloor = \infty$.

Now suppose $\lfloor \vec{e} \mid \upsilon \rfloor \neq \infty$. We show that $\vec{e}' \models \mathbf{O}\Gamma *_{\mathrm{w}} \upsilon$ for the following plausibilities of the conditionals in $\Gamma *_{\mathrm{w}} \upsilon$.

- $\lfloor \vec{e}' \mid \textsc{true} \rfloor = 1$; because by assumption $\lfloor \vec{e} \mid \upsilon \rfloor \neq \infty$ and thus $e'_1 \neq \{\}$.
- $\lfloor \vec{e}' \mid \neg(R \supset \upsilon) \rfloor = \infty$; because $w \models \upsilon$ for all $w \in e'_1$, and $w \not\models R$ for all $w \in e'_p \setminus e'_1$.
- $\lfloor \vec{e}' \mid \neg(R \supset (\phi \supset \psi)) \rfloor = \infty$ for all $\phi \Rightarrow \psi \in \Gamma_\upsilon$; because by Lemma B.3.2, $w \models (\phi \supset \psi)$ for all $w \in e_{\lfloor \vec{e} \mid \upsilon \rfloor} \supseteq e'_1$, and $w \not\models R$ for all $w \in e'_p \setminus e'_1$.
- $\lfloor \vec{e}' \mid (\neg R \land \phi) \rfloor = \lfloor \vec{e} \mid \phi \rfloor + 1$ for all $\phi \Rightarrow \psi \in \Gamma$; because $w \models R$ for all $w \in e'_1$, and so for all $p \in \mathbb{P}$ we have $p + 1 \geq \lfloor \vec{e}' \mid (\neg R \land \phi) \rfloor$ iff $w \models (\neg R \land \phi)$ for some $w \in e'_{p+1}$ iff $w \models (\neg R \land \phi)$ for some $w \in (\vec{e} \mid \neg R)_p$ iff (since Γ, υ are $\{R\}$-free and by Rule $\mathcal{ESB}10$ and Lemma B.2.13) $w \models \phi$ for some $w \in e_p$ iff $p \geq \lfloor \vec{e} \mid \phi \rfloor$.

Then $w \in e'_1$ iff $w \in ((\vec{e} \mid \upsilon) \mid R)_{\lfloor \vec{e} \mid \upsilon \rfloor}$ iff $w \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq \lfloor \vec{e} \mid \upsilon \rfloor}(\phi_i \supset \psi_i) \land \upsilon \land R$ iff (by Lemma B.3.2) $w \models \bigwedge_{\phi\Rightarrow\psi\in\Gamma_\upsilon}(\phi \supset \psi) \land \upsilon \land R$ iff $w \models \bigwedge_{\phi\Rightarrow\psi\in\Gamma_\upsilon}(R \supset (\phi \supset \psi)) \land (R \supset \upsilon) \land (\textsc{true} \supset R) \land \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor+1\geq 1}((\neg R \land \phi_i) \supset \psi_i)$. For $p > 1$, $w \in e'_p$ iff

$w \in ((\vec{e} \mid \upsilon) \mid R)_{\lfloor \vec{e} \mid \upsilon \rfloor}$ or $w \in (\vec{e} \mid \neg R)_{p-1}$ iff $w \models \bigwedge_{i : \lfloor \vec{e} \mid \phi_i \rfloor \geq \lfloor \vec{e} \mid \upsilon \rfloor} (\phi_i \supset \psi_i) \wedge \upsilon \wedge R$ or $w \models \bigwedge_{i : \lfloor \vec{e} \mid \phi_i \rfloor \geq p-1} (\phi_i \supset \psi_i) \wedge \neg R$ iff (by Lemma B.3.2) $w \models \bigwedge_{\phi \Rightarrow \psi \in \Gamma_\upsilon} (\phi \supset \psi) \wedge \upsilon \wedge R$ or $w \models \bigwedge_{i : \lfloor \vec{e} \mid \phi_i \rfloor + 1 \geq p} (\phi_i \supset \psi_i) \wedge \neg R$ iff $w \models \bigwedge_{\phi \Rightarrow \psi \in \Gamma_\upsilon} (R \supset (\phi \supset \psi)) \wedge (R \supset \upsilon) \wedge \bigwedge_{i : \lfloor \vec{e} \mid \phi_i \rfloor + 1 \geq p} ((\neg R \wedge \phi_i) \supset \psi_i)$. Thus the right-hand side of Rule $\mathcal{ESB}$10 holds, and hence $\vec{e}' \models \mathbf{O}\Gamma *_{\mathrm{w}} \upsilon$.

Since Γ and υ are $\{R\}$-free, for each $w \in (\vec{e} \mid \upsilon)_{\lfloor \vec{e} \mid \upsilon \rfloor}$ there is a $w' \in e'_1$ with $w \approx_{\{R\}} w'$ by Lemma B.2.13 and Rule $\mathcal{ESB}$10. Likewise, for each $w \in e_p$ there is a $w' \in (\vec{e} \mid \neg R)_p$ with $w \approx_{\{R\}} w'$. Thus $\vec{e} *_{\mathrm{w}} \upsilon = \vec{e}'_{\{R\}}$, and hence $\vec{e} *_{\mathrm{w}} \upsilon \models \mathbf{O}_{\{R\}}\Gamma *_{\mathrm{w}} \upsilon$.

Now let $\mathcal{S} \neq \{\}$. Let $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$ and $\vec{e}' \models \mathbf{O}\Gamma$. By Rule $\mathcal{ESB}$11 and Corollary 5.6.4, $\vec{e} = \vec{e}'_{\mathcal{S}}$. By the case for $\mathcal{S} = \{\}$, $\vec{e}' *_{\mathrm{w}} \upsilon \models \mathbf{O}_{\{R\}}\Gamma *_{\mathrm{w}} \upsilon$. By Rule $\mathcal{ESB}$11, $(\vec{e}' *_{\mathrm{w}} \upsilon)_{\mathcal{S}} \models \mathbf{O}_{\mathcal{S} \cup \{R\}}\Gamma *_{\mathrm{w}} \upsilon$. By Lemma B.3.3, $\vec{e} *_{\mathrm{w}} \upsilon \models \mathbf{O}_{\mathcal{S} \cup \{R\}}\Gamma *_{\mathrm{w}} \upsilon$. □

Next, we turn to the strong-revision theorem from Section 5.7. Showing that requires even more work than the weak-revision result.

Definition B.3.4 We say $*$ is a *symbol involution* when it maps object function symbols to object function symbols and predicate symbols to predicate symbols of corresponding arities, and $S = S^{**}$ for all object function or predicate symbols S. We denote by α^* the formula obtained from α by simultaneously replacing each object function or predicate symbol S with S^*. For a world w, we define the world w^* such that

- for all object function symbols g, $w^*[g^*(n_1, \ldots, n_k)] = w[g(n_1, \ldots, n_k)]$;
- for all rigid predicate symbols R and action sequences $z \neq \langle\rangle$,
  - $w^*[R^*(n_1, \ldots, n_k)] = w[R(n_1, \ldots, n_k)]$ if R^* is rigid;
  - $w^*[R^*(n_1, \ldots, n_k), \langle\rangle] = w[R(n_1, \ldots, n_k)]$ and
    $w^*[R^*(n_1, \ldots, n_k), z] = w[R^*(n_1, \ldots, n_k), z]$ if R^* is fluent;
- for all fluent predicate symbols F and action sequences z,
  - $w^*[F^*(n_1, \ldots, n_k)] = w[F(n_1, \ldots, n_k), \langle\rangle]$ if F^* is rigid;
  - $w^*[F^*(n_1, \ldots, n_k), z] = w[F(n_1, \ldots, n_k), z]$ if F^* is fluent.

For a set of worlds W and an epistemic state $\vec{e}$, we let $W^* = \{w^* \mid w \in W\}$ and $\vec{e}^* = \langle e^*_1, \ldots, e^*_{\lceil \vec{e} \rceil} \rangle$.

We use symbol involutions to rename the symbols of a formula: when $\mathcal{S}'$ contains all symbols of α, and $*$ maps each of them to a new symbol from a set $\mathcal{S}''$ disjoint with $\mathcal{S}'$, then α^* is just the result of replacing every symbol from $\mathcal{S}'$ with the corresponding symbol from $\mathcal{S}''$.

Lemma B.3.5 *Let $*$ be a symbol involution. Then $w = w^{**}$, $W = W^{**}$, and $\vec{e} = \vec{e}^{**}$.*

Proof. For $w = w^{**}$, the only non-trivial cases are when a fluent F is mapped to rigid F^* or the other way around. If R is rigid and R^* is fluent, then $w^{**}[R(n_1, \ldots, n_k)] = w^{**}[R^{**}(n_1, \ldots, n_k)] = w^*[R^*(n_1, \ldots, n_k), \langle\rangle] = w[R(n_1, \ldots, n_k)]$. If F is fluent and F^* is rigid, then $w^{**}[F(n_1, \ldots, n_k), \langle\rangle] = w^{**}[F^{**}(n_1, \ldots, n_k), \langle\rangle] = w^*[F^*(n_1, \ldots, n_k)] = w[F(n_1, \ldots, n_k), \langle\rangle]$, and similarly $w^{**}[F(n_1, \ldots, n_k), z] = w^{**}[F^{**}(n_1, \ldots, n_k), z] = w^*[F^*(n_1, \ldots, n_k), z] = w[F(n_1, \ldots, n_k), z]$ for all $z \neq \langle\rangle$. Thus $w = w^{**}$. $W = W^{**}$ and $\vec{e} = \vec{e}^{**}$ then follow immediately from the definition. □

Lemma B.3.6 *Let ϕ be objective static and let $*$ be a symbol involution. Then $w \models \phi$ iff $w^* \models \phi^*$.*

Proof. By induction on the length of ϕ. We show the base case only for fluent predicate symbols F with rigid image F^* (the other cases are analogous): $w \models F(t_1, \ldots, t_k)$ iff $w[F(n_1, \ldots, n_k), \langle\rangle] = 1$ where $n_i = w(t_i)$ iff $w^*[F^*(n_1, \ldots, n_k)] = 1$ where $n_i = w^*(t_i^*)$ iff $w^* \models (F(t_1, \ldots, t_k))^*$. The induction steps for $\neg\phi$, $(\phi \vee \beta)$, and $\exists x\, \phi$ are trivial. □

Lemma B.3.7 *Let ϕ be objective static and let $*$ be a symbol involution. Then $\{w \mid w \models \phi\}^* = \{w \mid w \models \phi^*\}$.*

Proof. Let $w^* \in \{w' \mid w' \models \phi\}^*$. Then $w \models \phi$, and by Lemma B.3.6, $w^* \models \phi^*$, so $w^* \in \{w' \mid w' \models \phi^*\}$. Conversely, let $w \in \{w' \mid w' \models \phi^*\}$. Then $w \models \phi^*$. By Lemma B.3.6, $w^{**} \models \phi$, and by Lemma B.3.5, $w = w^{**} \in \{w' \mid w' \models \phi\}^*$. □

Lemma B.3.8 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be objective and static, and let $*$ be a symbol involution. Then $\vec{e} \models \mathbf{O}\Gamma$ iff $\vec{e}^* \models \mathbf{O}\Gamma^*$.*

Proof. Let $\vec{e} \models \mathbf{O}\Gamma$ and $\vec{e}' \models \mathbf{O}\Gamma^*$. Then $e_p = \{w \mid w \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)\}$, and $e'_p = \{w' \mid w' \models \bigwedge_{i:\lfloor \vec{e}' \mid \phi_i^* \rfloor \geq p}(\phi_i^* \supset \psi_i^*)\}$ by Rule $\mathcal{ESB}$10. We show by induction on p that $e_p^* = e'_p$, and $\lfloor \vec{e} \mid \phi_i \rfloor > p$ iff $\lfloor \vec{e}' \mid \phi_i^* \rfloor > p$. For $p = 1$, this follows from Lemmas B.3.7 and B.3.6, respectively. For $p > 1$, by induction $\lfloor \vec{e} \mid \phi_i \rfloor \geq p$ iff $\lfloor \vec{e}' \mid \phi_i^* \rfloor \geq p$. Thus, just like in the base case, $e_p^* = e'_p$ by Lemma B.3.7, and $\lfloor \vec{e} \mid \phi_i \rfloor > p$ iff $\lfloor \vec{e}' \mid \phi_i^* \rfloor > p$ by Lemma B.3.6. Thus $\vec{e}^* = \vec{e}'$. The only-if direction of the lemma thus holds. Conversely, the if direction holds because $*$ is an involution: if $\vec{e}^* \models \mathbf{O}\Gamma^*$ for some $\vec{e}$, then $\vec{e}^{**} \models \mathbf{O}\Gamma^{**}$ by the only-if direction, and since $\vec{e}^{**} = \vec{e}$ by Lemma B.3.5 and $\mathbf{O}\Gamma^{**} = \mathbf{O}\Gamma$, $\vec{e} \models \mathbf{O}\Gamma$. □

Lemma B.3.9 *Let ϕ_1, ϕ_2 be objective sentences over object function and predicate symbols $\mathcal{S}_1, \mathcal{S}_2$ and let $\mathcal{S}_1, \mathcal{S}_2$ be disjoint. Suppose $w_1 \models \phi_1$, $w_2 \models \phi_2$. Then there is some w such that $w \models \phi_1 \wedge \phi_2$ and $w \approx_{\mathcal{S}_2} w_1$ and $w \approx_{\mathcal{S}_1} w_2$.*

Proof. Since $\mathcal{S}_1$ and $\mathcal{S}_2$ are disjoint, there clearly is a w with $w \approx_{\mathcal{S}_2} w_1$ and $w \approx_{\mathcal{S}_1} w_2$. By a simple induction on ϕ_1, $w \models \phi_1$ because ϕ_1 does not mention any symbol from

$\mathcal{S}_2$. Analogously, $w \models \phi_2$. □

Theorem 5.7.5 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ and v be objective and static. Let $\mathcal{S}$ be a set of object function and predicate symbols, and let v be $\mathcal{S}$-free. Let $\mathcal{S}''$ be the symbols newly introduced in $\Gamma *_s v$. If $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$, then $\vec{e} *_s v \models \mathbf{O}_{\mathcal{S}\cup\mathcal{S}''}\Gamma *_s v$.*

Proof. Let $\mathcal{S}'$ be the object function and predicate symbols that occur in Γ and $*$ be the symbol involution that maps each symbol from $\mathcal{S}'$ to the corresponding symbol from $\mathcal{S}''$. Recall that $\Delta = \{\phi \Rightarrow \psi \in \Gamma_v \mid \mathbf{O}\Gamma \not\models \mathbf{K}(\phi \supset \psi)\}$. Then $\Gamma *_s v = \Gamma'_v \cup \Gamma'_{\neg v} \cup \Delta' \cup \Lambda_1 \cup \Lambda_2 \cup \Lambda_3$, where

$$\begin{aligned}
\Gamma'_v &= \{\phi^* \Rightarrow \psi^* \mid \phi \Rightarrow \psi \in \Gamma_v\};\\
\Gamma'_{\neg v} &= \{(\phi^* \wedge \neg v) \Rightarrow \psi^* \mid \phi \Rightarrow \psi \in \Gamma_{\neg v}\};\\
\Delta' &= \{\text{TRUE} \Rightarrow v\} \cup \{\neg(\phi^* \supset \psi^*) \vee \neg v \Rightarrow v \mid \phi \Rightarrow \psi \in \Delta\};\\
\Lambda'_1 &= \{\neg((v \wedge \neg v^*) \supset (\phi \supset \psi)) \Rightarrow \text{FALSE} \mid \phi \Rightarrow \psi \in \Gamma_v\};\\
\Lambda'_2 &= \{\neg((\neg v \wedge v^*) \supset (\phi \supset \psi)) \Rightarrow \text{FALSE} \mid \phi \Rightarrow \psi \in \Gamma_{\neg v}\};\\
\Lambda'_3 &= \{\neg((v \equiv v^*) \supset (\phi \equiv \phi^*) \wedge (\psi \equiv \psi^*)) \Rightarrow \text{FALSE} \mid \phi \Rightarrow \psi \in \Gamma\}.
\end{aligned}$$

We show the theorem for the case $\mathcal{S} = \{\}$ first. Let $\vec{e} \models \mathbf{O}\Gamma$. We show that $\vec{e} *_s v \models \mathbf{O}_{\mathcal{S}''}\Gamma *_s v$.

First suppose $\lfloor \vec{e} \mid v \rfloor = \infty$. Then $\vec{e} *_s v = \langle\{\}\rangle$. Clearly, $\lfloor \langle\{\}\rangle \mid \phi \rfloor = \infty$ for all ϕ, so $\langle\{\}\rangle \models \mathbf{O}\Gamma *_s v$ if (by Rule $\mathcal{ESB}$10) $\bigwedge_{\phi\Rightarrow\psi\in\Gamma_v}(\phi^* \supset \psi^*) \wedge \bigwedge_{\phi\Rightarrow\psi\in\Gamma_v} \neg(((v \wedge \neg v^*) \supset (\phi \supset \psi)) \supset \text{FALSE}) \wedge v$ is unsatisfiable. By Rule $\mathcal{ESB}$10 and $\lfloor \vec{e} \mid v \rfloor = \infty$, $\bigwedge_{i:\lfloor \vec{e} \mid \phi \rfloor = \infty}(\phi \supset \psi) \wedge v$ is unsatisfiable, and by Lemma B.3.2, $\bigwedge_{\phi\Rightarrow\psi\in\Gamma_v}(\phi \supset \psi) \wedge v$ is unsatisfiable, too, and by Lemma B.3.6, $\bigwedge_{\phi\Rightarrow\psi\in\Gamma_v}(\phi^* \supset \psi^*) \models \neg v^*$. Hence, the formula mentioned before is indeed inconsistent.

Now suppose $\lfloor \vec{e} \mid v \rfloor \neq \infty$. We show $\vec{e} *_s v \models \mathbf{O}_{\mathcal{S}''}\Gamma *_s v$ by constructing a model of $\mathbf{O}\Gamma *_s v$ and showing that forgetting $\mathcal{S}''$ in this epistemic state yields $\vec{e} *_s v$. The proof proceeds in three steps. Step 1 is to show $\vec{e}' \models \mathbf{O}\Gamma'_v \cup \Gamma'_{\neg v} \cup \Delta'$ where

$$\begin{aligned}
\vec{e}' = \langle &(\vec{e}^* \mid v)_{\lfloor \vec{e} \mid v \rfloor}, \ldots, (\vec{e}^* \mid v)_{\lceil \vec{e} \rceil},\\
&(\vec{e}^* \mid \neg v)_{\lfloor \vec{e} \mid \neg v \rfloor} \cup (\vec{e}^* \mid v)_{\lceil \vec{e} \rceil}, \ldots, (\vec{e}^* \mid \neg v)_{\lceil \vec{e} \rceil} \cup (\vec{e}^* \mid v)_{\lceil \vec{e} \rceil}\rangle.
\end{aligned}$$

Step 2 is to show $(\vec{e}' \mid \lambda) \models \mathbf{O}\Gamma *_s v$, where $\lambda = \bigwedge_{\phi\Rightarrow\psi\in\Lambda'_1\cup\Lambda'_2\cup\Lambda'_3}(\phi \supset \psi)$. Step 3 is to prove $(\vec{e}' \mid \lambda)_{\mathcal{S}''} = \vec{e} *_s v$.

Step 1. We now prove that $\vec{e}' \models \mathbf{O}\Pi$ where $\Pi = \Gamma'_v \cup \Gamma'_{\neg v} \cup \Delta'$ for the following plausibilities of the conditionals in Π.

- $\lfloor \vec{e}' \mid \phi^* \rfloor = \max\{\lfloor \vec{e} \mid \upsilon \rfloor, \lfloor \vec{e} \mid \phi \rfloor\} - \lfloor \vec{e} \mid \upsilon \rfloor + 1$ for all $\phi \Rightarrow \psi \in \Gamma_\upsilon$; because for all $p \in \mathbb{P}$ with $p \geq \lfloor \vec{e} \mid \upsilon \rfloor$ we have $p \geq \lfloor \vec{e} \mid \phi \rfloor$ iff $w \models \phi$ for some $w \in e_p$ iff (by Lemma B.3.6) $w \models \phi^*$ for some $w \in \vec{e}^*_p$ iff (by Rule $\mathcal{ESB}$10 and Lemma B.3.9) $w \models \phi^* \wedge \upsilon$ for some $w \in \vec{e}^*_p$ iff $w \models \phi^*$ for some $w \in e'_{p - \lfloor \vec{e} \mid \upsilon \rfloor + 1}$ iff $p - \lfloor \vec{e} \mid \upsilon \rfloor + 1 \geq \lfloor \vec{e}' \mid \phi^* \rfloor$.

- $\lfloor \vec{e}' \mid \phi^* \wedge \neg\upsilon \rfloor = \max\{\lfloor \vec{e} \mid \neg\upsilon \rfloor, \lfloor \vec{e} \mid \phi \rfloor\} + \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor - \lfloor \vec{e} \mid \neg\upsilon \rfloor + 2$ for all $\phi \Rightarrow \psi \in \Gamma_{\neg\upsilon}$; because, very similar to the above, for all $p \in \mathbb{P}$ with $p \geq \lfloor \vec{e} \mid \neg\upsilon \rfloor$ we have $p \geq \lfloor \vec{e} \mid \phi \rfloor$ iff $w \models \phi$ for some $w \in e_p$ iff (by Lemma B.3.6) $w \models \phi^*$ for some $w \in \vec{e}^*_p$ iff (by Rule $\mathcal{ESB}$10 and Lemma B.3.9) $w \models \phi^* \wedge \neg\upsilon$ for some $w \in \vec{e}^*_p$ iff (since $w \models \upsilon$ for all $w \in e'_{p'}$ and $p' \leq \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor + 1$) $w \models \phi^* \wedge \neg\upsilon$ for some $w \in e'_{p - \lfloor \vec{e} \mid \neg\upsilon \rfloor + 1 + \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor + 1}$ iff $p + \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor - \lfloor \vec{e} \mid \neg\upsilon \rfloor + 2 \geq \lfloor \vec{e}' \mid \phi^* \wedge \neg\upsilon \rfloor$.

- $\max(\{\lfloor \vec{e}' \mid \textsc{true} \rfloor\} \cup \{\lfloor \vec{e}' \mid \neg(\phi^* \supset \psi^*) \vee \neg\upsilon \rfloor \mid \phi \Rightarrow \psi \in \Delta\}) = \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor + 1$; for the following reason. If $\Delta = \{\}$, then $\lfloor \vec{e} \mid \upsilon \rfloor = \lceil \vec{e} \rceil$, and since $\lfloor \vec{e} \mid \upsilon \rfloor \neq \infty$, $\lfloor \vec{e}' \mid \textsc{true} \rfloor = 1$. Now suppose $\phi \Rightarrow \psi \in \Delta$. Then $\lfloor \vec{e} \mid \neg(\phi \supset \psi) \rfloor \neq \infty$, and by Lemma B.3.1, $\lfloor \vec{e} \mid \neg(\phi \supset \psi) \rfloor \leq \lceil \vec{e} \rceil$. Moreover, $w \models (\phi \supset \psi)$ for all $w \in e_{\lfloor \vec{e} \mid \phi \rfloor}$, so $\lfloor \vec{e} \mid \neg(\phi \supset \psi) \rfloor \geq \lfloor \vec{e} \mid \phi \rfloor + 1$. In particular, there is some $\phi \Rightarrow \psi \in \Delta$ with $\lfloor \vec{e} \mid \phi \rfloor = \lceil \vec{e} \rceil - 1$, for otherwise $e_{\lceil \vec{e} \rceil - 1} = e_{\lceil \vec{e} \rceil}$ by Rule $\mathcal{ESB}$10; hence $\lfloor \vec{e} \mid \neg(\phi \supset \psi) \rfloor = \lceil \vec{e} \rceil$. By Lemma B.3.6 and since $w \not\models \neg\upsilon$ for all $w \in e'_p$ and $p \leq \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor + 1$, the equality follows.

First consider $p \leq \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor + 1$. Then $w \in e'_p$ iff $w \in (\vec{e}^* \mid \upsilon)_{p + \lfloor \vec{e} \mid \upsilon \rfloor - 1}$ iff (by Rule $\mathcal{ESB}$10 and Lemma B.3.6) $w \models \bigwedge_{i : \lfloor \vec{e} \mid \phi_i \rfloor \geq p + \lfloor \vec{e} \mid \upsilon \rfloor - 1} (\phi^*_i \supset \psi^*_i) \wedge \upsilon$ iff (by Lemma B.3.2) $w \models \bigwedge_{\phi \Rightarrow \psi \in \Gamma_\upsilon \text{ with } \max\{\lfloor \vec{e} \mid \upsilon \rfloor, \lfloor \vec{e} \mid \phi \rfloor\} \geq p + \lfloor \vec{e} \mid \upsilon \rfloor - 1} (\phi^* \supset \psi^*) \wedge \upsilon$ iff (since $\models \upsilon \supset (\phi^* \wedge \neg\upsilon \supset \psi^*)$ as well as $\models \upsilon \equiv ((\neg(\phi^* \supset \psi^*) \vee \neg\upsilon) \supset \upsilon)$ and by the above plausibilities)

$$w \models \bigwedge_{\substack{\phi \Rightarrow \psi \in \Gamma_\upsilon \\ \lfloor \vec{e}' \mid \phi^* \rfloor \geq p}} (\phi^* \supset \psi^*) \wedge \bigwedge_{\substack{\phi \Rightarrow \psi \in \Gamma_{\neg\upsilon} \\ \lfloor \vec{e}' \mid \phi^* \wedge \neg\upsilon \rfloor \geq p}} (\phi^* \wedge \neg\upsilon \supset \psi^*) \wedge \bigwedge_{\substack{\phi \Rightarrow \psi \in \Delta \\ \lfloor \vec{e}' \mid \neg(\phi^* \supset \psi^*) \vee \neg\upsilon \rfloor \geq p}} ((\neg(\phi^* \supset \psi^*) \vee \neg\upsilon) \supset \upsilon) \wedge \bigwedge_{\lfloor \vec{e}' \mid \textsc{true} \rfloor \geq p} (\textsc{true} \supset \upsilon)$$

iff $w \models \bigwedge_{\phi \Rightarrow \psi \in \Pi \text{ with } \lfloor \vec{e}' \mid \phi \rfloor \geq p} (\phi \supset \psi)$.

Now consider $p > \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor + 1$. Then $w \in e'_p$ iff $w \in (\vec{e}^* \mid \upsilon)_{\lceil \vec{e} \rceil}$ or $w \in (\vec{e}^* \mid \neg\upsilon)_{p - \lceil \vec{e} \rceil + \lfloor \vec{e} \mid \upsilon \rfloor + \lfloor \vec{e} \mid \neg\upsilon \rfloor - 2}$ iff (by Lemma B.3.6) $w \models \bigwedge_{i : \lfloor \vec{e} \mid \phi_i \rfloor = \infty} (\phi^*_i \supset \psi^*_i) \wedge \upsilon$ or $w \models \bigwedge_{\lfloor \vec{e} \mid \phi_i \rfloor \geq p - \lceil \vec{e} \rceil + \lfloor \vec{e} \mid \upsilon \rfloor + \lfloor \vec{e} \mid \neg\upsilon \rfloor - 2} (\phi^*_i \supset \psi^*_i) \wedge \neg\upsilon$ iff (by Lemma B.3.2 and by the above plausibilities)

$$w \models \bigwedge_{\substack{\phi \Rightarrow \psi \in \Gamma_{\neg\upsilon} \\ \lfloor \vec{e}' \mid \phi^* \rfloor = \infty}} (\phi^* \supset \psi^*) \wedge \upsilon \text{ or } w \models \bigwedge_{\substack{\phi \Rightarrow \psi \in \Gamma_{\neg\upsilon} \\ \lfloor \vec{e}' \mid \phi^* \wedge \neg\upsilon \rfloor \geq p}} (\phi^* \supset \psi^*) \wedge \neg\upsilon$$

iff

$$w \models \bigwedge_{\substack{\phi \Rightarrow \psi \in \Gamma_{\neg\upsilon} \\ \lfloor \vec{e}' \mid \phi^* \rfloor = \infty}} (\phi^* \supset \psi^*) \wedge \bigwedge_{\substack{\phi \Rightarrow \psi \in \Gamma_{\neg\upsilon} \\ \lfloor \vec{e}' \mid \phi^* \wedge \neg\upsilon \rfloor \geq p}} (\phi^* \wedge \neg\upsilon \supset \psi^*)$$

iff $w \models \bigwedge_{\phi \Rightarrow \psi \in \Pi \text{ with } \lfloor \vec{e}' \mid \phi \rfloor \geq p} (\phi \supset \psi)$. Therefore $\vec{e} \models \mathbf{O}\Pi$ and Step 1 is completed.

Step 2. The second step of the proof is to show $(\vec{e}' \mid \lambda) \models \mathbf{O}\Gamma *_s \upsilon$. The conditionals added in $\Gamma *_s \upsilon$ over Π are those from $\Lambda'_1 \cup \Lambda'_2 \cup \Lambda'_3$, which simply assert knowledge of λ as they have unsatisfiable consequents. Having shown $\vec{e}' \models \mathbf{O}\Pi$ in Step 1, it is immediate from Rule $\mathcal{ESB}$10 that $(\vec{e}' \mid \lambda) \models \mathbf{O}\Gamma *_s \upsilon$ if $\lfloor \vec{e}' \mid \tau \rfloor = \lfloor (\vec{e}' \mid \lambda) \mid \tau \rfloor$ for the antecedents $\tau \in \{\phi^*, \phi^* \wedge \neg\upsilon, (\phi^* \wedge \psi^*) \vee \neg\upsilon\}$ of the conditionals in Π. Clearly, $\lfloor \vec{e}' \mid \tau \rfloor \leq \lfloor (\vec{e}' \mid \lambda) \mid \tau \rfloor$. To show the converse, we let $w \in e'_p$ be arbitrary and show that there is some $w' \in (\vec{e}' \mid \lambda)_p$ such that $w' \models \tau$ iff $w \models \tau$. First suppose $w \models \neg\upsilon \wedge \upsilon^*$. Since $w \models \neg\upsilon$, by definition of $\vec{e}'$, $\lfloor \vec{e} \mid \neg\upsilon \rfloor \neq \infty$. Hence $\bigwedge_{\phi \Rightarrow \psi \in \Gamma_{\neg\upsilon}} (\phi \supset \psi) \wedge \neg\upsilon$ is satisfiable by Lemma B.3.2. By Lemma B.3.9 there is some w' with $w' \approx_{\mathcal{S}'} w$ with $w' \models \bigwedge_{\phi \Rightarrow \psi \in \Gamma_{\neg\upsilon}} ((\neg\upsilon \wedge \upsilon^*) \supset (\phi \supset \psi)) \wedge \neg\upsilon \wedge \upsilon^*$, so $w' \models \lambda$. Since $w' \models \upsilon$ iff $w \models \upsilon$ and by Lemma B.2.13 we have that for all $\phi \Rightarrow \psi \in \Pi$, $w' \models \phi$ iff $w \models \phi$ as well as $w' \models \psi$ iff $w \models \psi$, so $w' \in e'_p$ by Rule $\mathcal{ESB}$10. Therefore $w' \in (\vec{e}' \mid \lambda)_p$, and $w' \models \tau$ iff $w \models \tau$. The case for $w \models \upsilon \wedge \neg\upsilon^*$ is analogous. Now suppose $w \models \upsilon \equiv \upsilon^*$. By Lemma B.3.6, $\bigwedge_{\phi \Rightarrow \psi \in \Gamma} (\pm\phi \wedge \pm\psi) \wedge \pm\upsilon$ is satisfiable where $\pm\beta$ stands for β if $w \models \beta^*$ and for $\neg\beta$ otherwise. By Lemma B.3.9 there is some w' with $w \approx_{\mathcal{S}'} w'$ with $w' \models \bigwedge_{\phi \Rightarrow \psi \in \Gamma} (\pm\phi \wedge \pm\psi) \wedge \pm\upsilon$. Since $w' \models \upsilon$ iff $w \models \upsilon$ and by Lemma B.2.13 we have that for all $\phi \Rightarrow \psi \in \Pi$, $w' \models \phi$ iff $w \models \phi$ as well as $w' \models \psi$ iff $w \models \psi$, so $w' \in e'_p$ by Rule $\mathcal{ESB}$10. Therefore $w' \in (\vec{e}' \mid \lambda)_p$, and $w' \models \tau$ iff $w \models \tau$.

Step 3. Lastly, we need to show that $(\vec{e}' \mid \lambda)_{\mathcal{S}''} = \vec{e} *_s \upsilon$. Let $w \in ((\vec{e}' \mid \lambda)_{\mathcal{S}''})_p$. Then there is some $w' \in (\vec{e}' \mid \lambda)_p$ with $w \approx_{\mathcal{S}''} w'$. Thus $w' \models \bigwedge_{\phi \Rightarrow \psi \in \Gamma *_s \upsilon \text{ with } \lfloor \vec{e}' \mid \phi \rfloor \geq p} (\phi \supset \psi)$. First suppose $p \leq \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor + 1$. Then $w \models \upsilon$. If $w' \models \upsilon \equiv \upsilon^*$, then by Lemma B.2.13 and Λ_3, $w \models \bigwedge_{i : \lfloor \vec{e} \mid \phi_i \rfloor \geq p + \lfloor \vec{e} \mid \upsilon \rfloor - 1} (\phi_i \supset \psi_i) \wedge \upsilon$, and thus $w \in (\vec{e} \mid \upsilon)_{p + \lfloor \vec{e} \mid \upsilon \rfloor - 1} = (\vec{e} *_s \upsilon)_p$. Otherwise, $w' \models \upsilon \wedge \neg\upsilon^*$, and then by Lemma B.2.13 and Λ_1, $w \models \bigwedge_{i : \lfloor \vec{e} \mid \phi_i \rfloor \geq \lfloor \vec{e} \mid \upsilon \rfloor} (\phi_i \supset \psi_i) \wedge \upsilon$, and thus $w \in (\vec{e} \mid \upsilon)_{\lfloor \vec{e} \mid \upsilon \rfloor} \subseteq (\vec{e} \mid \upsilon)_{p + \lfloor \vec{e} \mid \upsilon \rfloor - 1} = (\vec{e} *_s \upsilon)_p$. Now suppose $p > \lceil \vec{e} \rceil - \lfloor \vec{e} \mid \upsilon \rfloor + 1$. The cases for $w' \models \upsilon \equiv \upsilon^*$ and $w' \models \upsilon \wedge \neg\upsilon^*$ are analogous. If $w' \models \neg\upsilon \wedge \upsilon^*$, then by Lemma B.2.13 and Λ_2, $w \models \bigwedge_{i : \lfloor \vec{e} \mid \phi_i \rfloor \geq \lfloor \vec{e} \mid \neg\upsilon \rfloor} (\phi_i \supset \psi_i) \wedge \neg\upsilon$, and thus $w \in (\vec{e} \mid \neg\upsilon)_{\lfloor \vec{e} \mid \neg\upsilon \rfloor} \subseteq (\vec{e} \mid \neg\upsilon)_{p-\lceil\vec{e}\rceil+\lfloor\vec{e}\mid\upsilon\rfloor+\lfloor\vec{e}\mid\neg\upsilon\rfloor-2} \subseteq (\vec{e} *_{\mathrm{s}} \upsilon)_p$. For the converse direction, let $w' \in (\vec{e} *_{\mathrm{s}} \upsilon)_p$. Since $*$ swaps the (initial) values of $\mathcal{S}'$ and $\mathcal{S}''$, there clearly is a w such that $w \approx_{\mathcal{S}''} w'$ and $w^* = w$. Thus $w \models \upsilon$ iff $w^* \models \upsilon$, and therefore $w \in e'_p$. For all $\phi \Rightarrow \psi \in \Gamma$, $w \models \phi$ iff (by Lemma B.3.6) $w^* \models \phi^*$ iff $w \models \phi^*$, and likewise for ψ. Thus $w \models \lambda$, and hence $w \in (\vec{e}' \mid \lambda)_p$. As $w \approx_{\mathcal{S}''} w'$, we have $w' \in ((\vec{e}' \mid \lambda)_{\mathcal{S}''})_p$.

Now let $\mathcal{S} \neq \{\}$. Let $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$ and $\vec{e}' \models \mathbf{O}\Gamma$. By Rule $\mathcal{ESB}$11 and Corollary 5.6.4, $\vec{e} = \vec{e}'_{\mathcal{S}}$. By the case for $\mathcal{S} = \{\}$, $\vec{e}' *_{\mathrm{s}} \upsilon \models \mathbf{O}_{\mathcal{S}''}\Gamma *_{\mathrm{s}} \upsilon$. By Rule $\mathcal{ESB}$11, $(\vec{e}' *_{\mathrm{s}} \upsilon)_{\mathcal{S}} \models \mathbf{O}_{\mathcal{S}\cup\mathcal{S}''}\Gamma *_{\mathrm{s}} \upsilon$. By Lemma B.3.3, $\vec{e} *_{\mathrm{s}} \upsilon \models \mathbf{O}_{\mathcal{S}\cup\mathcal{S}''}\Gamma *_{\mathrm{s}} \upsilon$. □

B.4 Proof of the progression theorems

Here we prove the correctness of progression, that is, Theorems 5.8.2 and 5.8.3. We use the following assumptions.

- Let $\Sigma_{\mathrm{bel}}, \Sigma_{\mathrm{dyn}}$ be a basic action theory over fluents $\mathcal{F} = \{F_1, \ldots, F_l\}$, and let n be an action standard name. Recall that Σ_{dyn} contains the successor state axioms $\square[a]F(x_1, \ldots, x_k) \equiv \gamma_F$ for $F \in \mathcal{F}$, and the informed fluent axiom $\square\mathrm{IF}(a) \equiv \varphi$.

- Let $\mathcal{S}'$ be the symbols newly introduced in $\Sigma_{\mathrm{bel}} \gg n$, which is partitioned into two subsets: $\mathcal{R} = \{R_1, \ldots, R_l\} \subseteq \mathcal{S}'$ contains the rigid predicates for the physical progression as in Definition 5.8.1; $\mathcal{S}' \setminus \mathcal{R}$ contains the rigid symbols introduced by the revision as in Definition 5.7.2 or 5.7.4.

- Let $*$ be the symbol involution that maps F_i to R_i and leaves the rest unchanged.

Definition B.4.1 For a world w and an action n, w^n is a world such that $w^n \approx_{\mathcal{F}} (w \gg n)$ and

- $w^n[F(n_1, \ldots, n_k), \langle\rangle] = 1$ iff $w^n \models (\gamma_F{}^{x_1 \ldots x_k\, a}_{n_1 \ldots n_k\, n})^*$ for all $F \in \mathcal{F}$;

- $w^n[F(n_1, \ldots, n_k), z] = w[F(n_1, \ldots, n_k), z]$ for all $F \in \mathcal{F}$ and $z \neq \langle\rangle$.

For a set of worlds W and an epistemic state $\vec{e}$, we let $W^n = \{w^n \mid w \in W\}$ and $\vec{e}^n = \langle e_1^n, \ldots, e_{\lceil\vec{e}\rceil}^n\rangle$.

Intuitively, w^n sets the initial values of every fluent $F \in \mathcal{F}$ to the its value after n when the values before n are memorized in $\mathcal{R}$: $w^n \models F(n_1, \ldots, n_k)$ iff $w^n \models (\gamma_F{}^{x_1 \ldots x_k\, a}_{n_1 \ldots n_k\, n})^*$, where $*$ replaces all fluents in γ_F with the corresponding rigid predicates from $\mathcal{R}$.

Lemma B.4.2 *w^n is uniquely defined.*

Proof. Let $w' \approx_{\mathcal{F}} (w \gg n)$ be such that for every $F \in \mathcal{F}$, $w'[F(n_1,\dots,n_k),\langle\rangle] = 1$ iff $w \gg n \models (\gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n})^*$, and moreover $w'[F(n_1,\dots,n_k),z] = w[F(n_1,\dots,n_k),z]$ for all $z \neq \langle\rangle$. Clearly such a w' exists and is uniquely defined. By Lemma B.2.13 and since γ_F^* is $\mathcal{F}$-free, $w' \models (\gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n})^*$ iff $w \gg n \models (\gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n})^*$. Thus $w' = w^n$. □

Lemma B.4.3 *Let ϕ be fluent. Then $w \models \phi^*$ iff $w^n \models \phi^*$.*

Proof. Since ϕ is fluent and by definition of $*$, ϕ^* mentions only rigid predicates. By Lemma B.4.2, w^n is uniquely defined, and since w and w^n agree on all rigids, a simple induction on the length of ϕ shows that the lemma holds. □

Lemma B.4.4 *If $\vec{e} \models \mathbf{O}(\Sigma_{\text{bel}} * \varphi^a_n)^*$, then $\vec{e}^{\,n} \models \mathbf{O}\Sigma_{\text{bel}} \gg n$.*

Proof. Let $(\Sigma_{\text{bel}} * \varphi^a_n)^* = \{\phi_1 \Rightarrow \psi_1, \dots, \phi_m \Rightarrow \psi_m\}$ and $\vec{e} \models \mathbf{O}(\Sigma_{\text{bel}} * \varphi^a_n)^*$. We show that $\vec{e}^{\,n} \models \mathbf{O}\Sigma_{\text{bel}} \gg n$ for the same plausibilities and plausibility ∞ for the added conditionals. We first show that $\lfloor \vec{e} \mid \phi_i \rfloor = \lfloor \vec{e}^{\,n} \mid \phi_i \rfloor$ for all i and $\lfloor \vec{e}^{\,n} \mid \phi \rfloor = \infty$ for the newly added conditionals $\phi \Rightarrow \psi \in (\Sigma_{\text{bel}} \gg n) \setminus (\Sigma_{\text{bel}} * \varphi^a_n)$. If $p < \lfloor \vec{e} \mid \phi_i \rfloor$, then $w \not\models \phi_i$ for all $w \in e_p$, and by Lemma B.4.3, $w \not\models \phi_i$ for all $w \in \vec{e}^{\,n}_p$, so $p < \lfloor \vec{e}^{\,n} \mid \phi_i \rfloor$. Analogously, if $p \geq \lfloor \vec{e} \mid \phi_i \rfloor$, then $p \geq \lfloor \vec{e}^{\,n} \mid \phi_i \rfloor$. Thus $\lfloor \vec{e} \mid \phi_i \rfloor = \lfloor \vec{e} \mid \phi_i \rfloor$. All other conditionals from $\Sigma_{\text{bel}} \gg n$ are of the form $\neg\big(F(x_1,\dots,x_k) \equiv (\gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n})^*\big) \Rightarrow \textsc{false}$ for some $F \in \mathcal{F}$, and by definition of w^n, $\lfloor \vec{e}^{\,n} \mid \neg(F(x_1,\dots,x_k) \equiv (\gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n})^*)\rfloor = \infty$.

Now we prove $w \in \vec{e}^{\,n}_p$ iff $w \models \bigwedge_{i:\lfloor \vec{e}^{\,n} \mid \phi_i \rfloor \geq p}(\phi_i \supset \psi_i) \land \bigwedge_{F \in \mathcal{F}}(\neg\big(F(x_1,\dots,x_k) \equiv (\gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n})^*\big) \supset \textsc{false})$, that is, $\vec{e}^{\,n}$ satisfies Rule $\mathcal{ESB}$10. For the only-if direction suppose $w' \in \vec{e}^{\,n}_p$. Then for some $w \in e_p$, $w^n = w'$. By Rule $\mathcal{ESB}$10 for $\vec{e}$ we have $w \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)$. Since $w' = w^n$ and by $\lfloor \vec{e} \mid \phi_i \rfloor = \lfloor \vec{e}^{\,n} \mid \phi_i \rfloor$ and Lemma B.4.3, $w' \models \bigwedge_{i:\lfloor \vec{e}^{\,n} \mid \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)$, and by of w^n also $w' \models F(x_1,\dots,x_k) \equiv (\gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n})^*$ for every $F \in \mathcal{F}$. Thus the right-hand side holds. Conversely, suppose $w' \notin \vec{e}^{\,n}_p$. If there is some w with $w^n = w'$, then $w \notin e_p$, and hence by $\lfloor \vec{e} \mid \phi_i \rfloor = \lfloor \vec{e}^{\,n} \mid \phi_i \rfloor$, $w' \not\models \bigwedge_{i:\lfloor \vec{e}^{\,p} \mid \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)$. Otherwise, for all w, $w^n \neq w'$, and hence $w' \not\models F(x_1,\dots,x_k) \equiv (\gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n})^*$ for some $F \in \mathcal{F}$. In either case the right-hand side is false. □

Lemma B.4.5 $(w_{\Sigma_{\text{dyn}}} \gg n) \approx_{\mathcal{R}} ((w^*)^n)_{\Sigma_{\text{dyn}}}$.

Proof. The cases for object function symbols and rigid predicate symbols $R \notin \mathcal{R}$ are trivial as they are left unchanged by the involved Definitions 5.8.1, B.2.3, B.3.4, B.4.1. For $F \in \mathcal{F}$ and $z = \langle\rangle$, $(w_{\Sigma_{\text{dyn}}} \gg n)[F(n_1,\dots,n_k),\langle\rangle] = 1$ iff $w_{\Sigma_{\text{dyn}}}[F(n_1,\dots,n_k),n] = 1$ iff $w_{\Sigma_{\text{dyn}}} \models \gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n}$ iff (by Lemma B.2.7) $w \models \gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n}$ iff (by Lemma B.3.6) $w^* \models (\gamma_F{}^{x_1 \dots x_k\, a}_{n_1 \dots n_k\, n})^*$ iff $(w^*)^n[F(n_1,\dots,n_k),\langle\rangle] = 1$ iff $((w^*)^n)_{\Sigma_{\text{dyn}}}[F(n_1,\dots,n_k),\langle\rangle] = 1$. The cases for $F \in \mathcal{F}$ with $z \neq \langle\rangle$ and for IF follow from the definition of $w_{\Sigma_{\text{dyn}}}$ and φ, γ_F being $\mathcal{R}$-free and fluent. Finally, $F \notin \mathcal{F} \cup \{\text{IF}\}$ follows because both $w \gg n$ and w^n progress F by n. □

Lemma B.4.6 *Let* $\vec{e} \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} * \varphi_n^a)$ *and* $\vec{e}' \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$*. Then for all* $p \in \mathbb{P}$*,* $(e_p \gg n) = (\vec{e}'_{\mathcal{R}})_p$*.*

Proof. Let $\vec{e}'' \models \mathbf{O}\Sigma_{\text{bel}} * \varphi_n^a$, which exists by Theorem 5.3.16. Then by Lemma B.2.8, $\vec{e}''_{\Sigma_{\text{dyn}}} \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} * \varphi_n^a)$, and by Theorem 5.3.16, $\vec{e} = \vec{e}''_{\Sigma_{\text{dyn}}}$ (*). By Lemma B.3.8, $\vec{e}''^{*} \models \mathbf{O}(\Sigma_{\text{bel}} * \varphi_n^a)^*$, and by Lemma B.4.4, $(\vec{e}''^{*})^n \models \mathbf{O}\Sigma_{\text{bel}} \gg n$, and by Lemma B.2.8, $((\vec{e}''^{*})^n)_{\Sigma_{\text{dyn}}} \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$, and by Theorem 5.3.16, $((\vec{e}''^{*})^n)_{\Sigma_{\text{dyn}}} = \vec{e}'$ (**).

For the $\subseteq$ direction let $w' \in (e_p \gg n)$. By (*) there is some $w \in \vec{e}''_p$ such that $(w_{\Sigma_{\text{dyn}}} \gg n) = w'$. Also, by (**), $((w^*)^n)_{\Sigma_{\text{dyn}}} \in e'_p$. By Lemma B.4.5, $(w_{\Sigma_{\text{dyn}}} \gg n) \approx_{\mathcal{R}} ((w^*)^n)_{\Sigma_{\text{dyn}}}$. Thus $w' = (w_{\Sigma_{\text{dyn}}} \gg n) \in (\vec{e}'_{\mathcal{R}})_p$, so $(e_p \gg n) \subseteq (\vec{e}'_{\mathcal{R}})_p$.

Conversely, let $w \in (\vec{e}'_{\mathcal{R}})_p$. Then $w \approx_{\mathcal{R}} w'$ for some $w' \in e'_p$. By (**) there is some $w'' \in \vec{e}''_p$ such that $((w''^{*})^n)_{\Sigma_{\text{dyn}}} = w'$. By Lemma B.4.5, $(w''_{\Sigma_{\text{dyn}}} \gg n) \approx_{\mathcal{R}} w'$, and thus $(w''_{\Sigma_{\text{dyn}}} \gg n) \approx_{\mathcal{R}} w$. Hence $w \in ((\vec{e}''_p)_{\Sigma_{\text{dyn}}} \gg n)_{\mathcal{R}} = (e_p \gg n)_{\mathcal{R}}$ with (*). Thus $(\vec{e}'_{\mathcal{R}})_p \subseteq (e_p \gg n)_{\mathcal{R}}$. Since $\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} * \varphi_n^a$ are $\mathcal{R}$-free and by Lemma B.2.13, $(e_p)_{\mathcal{R}} = e_p$, and so $(e_p \gg n)_{\mathcal{R}} = (e_p \gg n)$. Thus $(\vec{e}'_{\mathcal{R}})_p \subseteq (e_p \gg n)$. □

Lemma B.4.7 *If* $\vec{e} \models \mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$*, then* $\vec{e} * \mathrm{IF}(n) \models \mathbf{O}_{\mathcal{S} \cup (\mathcal{S}' \setminus \mathcal{R})}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} * \varphi_n^a)$*.*

Proof. Since $w \models \Box\mathrm{IF}(n) \equiv \varphi_n^a$ for all $w \in e_p$ and $p \in \mathbb{P}$, we have $\vec{e} * \mathrm{IF}(n) = \vec{e} * \varphi_n^a$. Hence $\vec{e} * \mathrm{IF}(n) \models \mathbf{O}_{\mathcal{S} \cup (\mathcal{S}' \setminus \mathcal{R})}((\{\neg\Sigma_{\text{dyn}} \Rightarrow \textsc{false}\} \cup \Sigma_{\text{bel}}) * \varphi_n^a)$ by Theorems 5.7.3 and 5.7.5. Let $\Theta = ((\{\neg\Sigma_{\text{dyn}} \Rightarrow \textsc{false}\} \cup \Sigma_{\text{bel}}) * \varphi_n^a) \setminus (\Sigma_{\text{bel}} * \varphi_n^a)$. It is easy to see that Σ_{dyn} is not affected by revision, that is, $\bigwedge_{\phi \Rightarrow \psi \in \Theta}(\phi \supset \psi)$ is equivalent to $\neg\Sigma_{\text{dyn}} \supset \textsc{false}$. Therefore $\vec{e} * \mathrm{IF}(n) \models \mathbf{O}_{\mathcal{S} \cup (\mathcal{S}' \setminus \mathcal{R})}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} * \varphi_n^a)$. □

Theorem 5.8.2 $\models \mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \supset [n]\mathbf{O}_{\mathcal{S} \cup \mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$*.*

Proof. Suppose $\vec{e} \models \mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$. By Theorem 5.3.16, $\vec{e}' \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} * \varphi_n^a)$ and $\vec{e}'' \models \mathbf{O}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$ exist. By Lemma B.4.7, $\vec{e} * \mathrm{IF}(n) \models \mathbf{O}_{\mathcal{S} \cup (\mathcal{S}' \setminus \mathcal{R})}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} * \varphi_n^a)$. By Rule $\mathcal{ESB}$11 and Theorem 5.3.16, $\vec{e} * \mathrm{IF}(n) = \vec{e}'_{\mathcal{S} \cup (\mathcal{S}' \setminus \mathcal{R})}$. By Lemma B.4.6 we have $(e'_p \gg n) = (\vec{e}''_{\mathcal{R}})_p$. Thus $(\vec{e}'_{\mathcal{S} \cup (\mathcal{S}' \setminus \mathcal{R})})_p \gg n = (\vec{e}''_{\mathcal{S} \cup \mathcal{S}'})_p$, and so $\vec{e} \gg n = \vec{e}''_{\mathcal{S} \cup \mathcal{S}'}$. Moreover by assumption and Rule $\mathcal{ESB}$11, $\vec{e}''_{\mathcal{S} \cup \mathcal{S}'} \models \mathbf{O}_{\mathcal{S} \cup \mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$. Thus by Rule $\mathcal{ESB}$7, $\vec{e} \models [n]\mathbf{O}_{\mathcal{S} \cup \mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$. □

Theorem 5.8.3 $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models [n]\alpha$ *iff* $\mathbf{O}_{\mathcal{S} \cup \mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n) \models \alpha$*.*

Proof. For the if direction, suppose $\mathbf{O}_{\mathcal{S} \cup \mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n) \models \alpha$. Let $\vec{e} \models \mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$. By Theorem 5.8.2, $\vec{e} \gg n \models \mathbf{O}_{\mathcal{S} \cup \mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$. By assumption, $\vec{e} \gg n, w \gg n \models \alpha$ for all w. By Rule $\mathcal{ESB}$7, $\vec{e}, w \models [n]\alpha$ for all w.

Conversely, suppose $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models [n]\alpha$. Let $\vec{e} \models \mathbf{O}_{\mathcal{S} \cup \mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$. By Corollary 5.6.4, there is some $\vec{e}' \models \mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}})$. By assumption, $\vec{e}', w \models [n]\alpha$ for all w. Thus $\vec{e}' \gg n, w \models \alpha$ for all w. By Theorem 5.8.2, $\vec{e}' \gg n \models \mathbf{O}_{\mathcal{S} \cup \mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)$, and again by Corollary 5.6.4, $\vec{e} = \vec{e}' \gg n$. Thus $\vec{e}, w \models \alpha$ for all w. □

B.5 Proof of the representation theorems

Here we prove the representation theorem for $\mathcal{ESB}$, Theorem 5.9.7. The representation theorem for $\mathcal{BO}$, Theorem 4.8.5, is just a special case of Theorem 5.9.7. Corollaries 5.9.8 and 5.9.9, which combine the representation theorem with regression or progression, respectively, follow immediately as well.

Our proof is similar to the one of the representation theorem in $\mathcal{OL}$ in (Levesque and Lakemeyer 2001). In fact, the following lemmas generalize their Lemma 2.8.5, Corollary 2.8.6, and Lemma 7.2.2 from (Levesque and Lakemeyer 2001), respectively, to our notions of beliefs and action standard names. Theorem 5.9.7 and Corollaries 5.9.8 and 5.9.9 then follow easily.

Definition B.5.1 A function $*$ between standard names is called a *preserving involution* with respect to α iff

- $n^{**} = n$ and n^* is of the same sort as n for all standard names n;
- $A(n_1, \ldots, n_k)^* = A(n_1^*, \ldots, n_k^*)$ for all n_i for every action function A which is mentioned in α as a non-standard-name term $A(t_1, \ldots, t_k)$.

For non-standard-names terms, we extend the definition inductively by letting $x^* = x$ for variables x, $A(t_1, \ldots, t_k)^* = A(t_1^*, \ldots, t_k^*)$ for non-standard-name action terms, and $g(t_1, \ldots, t_k)^* = g(t_1^*, \ldots, t_k^*)$ for object terms. For formulas, we define α^* inductively by $R(t_1, \ldots, t_k)^* = R(t_1^*, \ldots, t_k^*)$ for rigid R; $F(t_1, \ldots, t_k)^* = F(t_1^*, \ldots, t_k^*)$ for fluent F; $(t_1 = t_2)^* = (t_1^* = t_2^*)$; $(\neg\alpha)^* = \neg\alpha^*$; $(\alpha \lor \beta)^* = (\alpha^* \lor \beta^*)$; $(\exists x\, \alpha)^* = \exists x\, \alpha^*$; $([t]\alpha)^* = [t^*]\alpha^*$; $(\mathbf{B}(\alpha \Rightarrow \beta))^* = \mathbf{B}(\alpha^* \Rightarrow \beta^*)$. For a world w, we let w^* be such that $w^*[F(n_1, \ldots, n_k), z] = w[F(n_1^*, \ldots, n_k^*), z^*]$ for all fluent predicate symbols F, $w^*[R(n_1, \ldots, n_k)] = w[R(n_1^*, \ldots, n_k^*)]$ for all rigid predicate symbols R, and $w^*[g(n_1, \ldots, n_k)] = w[g(n_1^*, \ldots, n_k^*)]^*$ for all object function symbols g.

Lemma B.5.2 *Let $*$ be a preserving involution with respect to α.*

(i) $(\alpha^x_n)^* = (\alpha^*)^x_{n^*}$.

(ii) *$*$ is a preserving involution with respect to α^x_n.*

Proof. (i) We first show that $(t^x_n)^* = (t^*)^x_{n^*}$ for any term t whose action function symbols occur in α by induction on the size of t. First consider the base cases. For the variable x, $(x^x_n)^* = n^* = x^x_{n^*}$. For variables distinct from x, for object constants, and for object and action standard names the claim trivially holds. Now we do the induction step. For an object function g, $(g(t_1, \ldots, t_k)^x_n)^* = (g(t_1{}^x_n, \ldots, t_k{}^x_n))^* = g((t_1{}^x_n)^*, \ldots, (t_k{}^x_n)^*)$

= (by induction) $g((t_1^*)^x_{n^*}, \ldots, (t_k^*)^x_{n^*}) = g(t_1^*, \ldots, t_k^*)^x_{n^*} = (g(t_1, \ldots, t_k)^*)^x_{n^*}$. For a non-standard-name action term, $(A(t_1, \ldots, t_k)^x_n)^* = A(t_1{}^x_n, \ldots, t_k{}^x_n)^* =$ (if $A(t_1{}^x_n, \ldots, t_k{}^x_n)$ is a standard name: because $*$ is preserving) $A((t_1{}^x_n)^*, \ldots, (t_k{}^x_n)^*) =$ (by induction) $A((t_1^*)^x_n, \ldots, (t_k^*)^x_n) = A(t_1^*, \ldots, t_k^*)^x_n =$ (if $A(t_1{}^x_n, \ldots, t_k{}^x_n)$ is a standard name: because $*$ is preserving) $(A(t_1, \ldots, t_k)^*)^x_n$. With that, (i) follows by an easy induction on the length of α.

(ii) Suppose α^x_n mentions an action function symbol A that does not occur in α. Then this occurrence can only be in the standard name n, that is, $n = A(n_1, \ldots, n_k)$. Thus $*$ is a preserving involution with respect to α^x_n. □

Lemma B.5.3 *Let ϕ be objective and let $*$ be a preserving involution with respect to ϕ. Then $\models \phi$ iff $\models \phi^*$.*

Proof. We need to show that $w^* \models \phi$ iff $w \models \phi^*$ for any arbitrary world w.

We first show that $w^*(t) = w(t^*)^*$ for all ground terms t whose action function symbols occur in ϕ (*). The proof is by induction on the nesting depth of t. The base cases are standard names and object constant symbols. For a standard name, $w^*(n) = n = n^{**} = w(n^*)^*$. For an object constant symbol g, $w^*(g) = w[g]^* = w(g)^* = w(g^*)^*$. For the induction step let $g(t_1, \ldots, t_k)$ be an object term of nesting depth l and suppose the $w^*(t) = w(t^*)^*$ for all t of nesting depth $< l$, which particularly includes the t_i. Then $w^*(g(t_1, \ldots, t_k)) = n$ iff $w^*[g(n_1, \ldots, n_k)] = n$ where $n_i = w^*(t_i)$ iff (by definition of w^*) $w[g(n_1^*, \ldots, n_k^*)]^* = n$ where $n_i = w^*(t_i)$ iff (by induction) $w[g(n_1^*, \ldots, n_k^*)]^* = n$ where $n_i = w(t_i^*)^*$ iff $w[g(n_1, \ldots, n_k)]^* = n$ where $n_i = w(t_i^*)$ iff $w(g(t_1^*, \ldots, t_k^*))^* = n$ iff (by definition of t^*) $w(g(t_1, \ldots, t_k)^*)^* = n$. Consider an action function symbol A that occurs as non-standard-name in ϕ. Then $w^*(A(t_1, \ldots, t_k)) = n$ iff $A(n_1, \ldots, n_k) = n$ where $n_i = w^*(t_i)$ iff (by induction) $A(n_1, \ldots, n_k) = n$ where $n_i = w(t_i^*)^*$ iff $A(n_1^*, \ldots, n_k^*) = n$ where $n_i = w(t_i^*)$ iff (by definition of n^*) $A(n_1, \ldots, n_k)^* = n$ where $n_i = w(t_i^*)$ iff $w(A(t_1^*, \ldots, t_k^*))^* = n$ iff (by definition of t^*) $w(A(t_1, \ldots, t_k)^*)^* = n$.

Now we do the induction on the length of ϕ to show the lemma. Consider a rigid predicate symbol R. Then $w^* \models R(t_1, \ldots, t_k)$ iff $w^*[R(n_1, \ldots, n_k)] = 1$ where $n_i = w^*(t_i)$ iff (by (*)) $w^*[R(n_1, \ldots, n_k)] = 1$ where $n_i = w(t_i^*)^*$ iff (by definition of w^*) $w[R(n_1^*, \ldots, n_k^*)] = 1$ where $n_i = w(t_i^*)^*$ iff $w[R(n_1, \ldots, n_k)] = 1$ where $n_i = w(t_i^*)$ iff $w \models R(t_1^*, \ldots, t_k^*)$. The base case for fluent predicates and equality expressions is analogous. The induction steps for $\neg\phi$ and $(\phi \vee \psi)$ are trivial. For actions, $w^* \models [t]\phi$ iff $w^* \gg n \models \phi$ for $n = w^*(t)$ iff (by definition of w^* and by (*)) $(w \gg n^*)^* \models \phi$ for $n = w(t^*)^*$ iff (by induction) $(w \gg n^*) \models \phi^*$ for $n = w(t^*)^*$ iff $(w \gg n) \models [t^*]\phi^*$. For quantifiers, $w^* \models \exists x\, \phi$ iff $w^* \models \phi^x_n$ for some n iff (by induction) $w \models (\phi^x_n)^*$ for some n iff (by Lemma B.5.2) $w \models (\phi^*)^x_{n^*}$ for some n iff $w \models (\exists x\, \phi)^*$. □

Corollary B.5.4 *Let ϕ be an objective formula with free variables $x_1, \ldots, x_k$ and let $*$ be a preserving involution with respect to ϕ which leaves the names in ϕ unchanged. Then for any names $n_1, \ldots, n_k$ of corresponding sorts, $\models \phi^{x_1 \ldots x_k}_{n_1 \ldots n_k}$ iff $\models \phi^{x_1 \ldots x_k}_{n_1^* \ldots n_k^*}$.*

Proof. $\models \phi^{x_1 \ldots x_k}_{n_1 \ldots n_k}$ iff (by Lemma B.5.3) $\models (\phi^{x_1 \ldots x_k}_{n_1 \ldots n_k})^*$ iff (by Lemma B.5.2) $\models (\phi^*)^{x_1 \ldots x_k}_{n_1^* \ldots n_k^*}$ iff (since $*$ leaves all names in ϕ unchanged) $\models \phi^{x_1 \ldots x_k}_{n_1^* \ldots n_k^*}$. □

Lemma B.5.5 *Let ϕ be an objective sentence. Let ψ be an objective formula with free variables $x_1, \ldots, x_k$ and let $n_1, \ldots, n_k$ be standard names of corresponding sorts. Then $\models \mathrm{RES}[\![\psi, \phi]\!]^{x_1 \ldots x_k}_{n_1 \ldots n_k}$ iff $\models (\phi \supset \psi^{x_1 \ldots x_k}_{n_1 \ldots n_k})$.*

Proof. The proof is by induction on the number of free variables in ψ. If there are none, the lemma holds immediately by Definition 5.9.1.

For the induction step let ψ have k free variables and suppose that the lemma holds for $\psi^{x_1}_{n_1}$ for arbitrary n_1. We consider here only the case where x_1 is of sort action; the proof for object variables is actually simpler and matches Lemma 7.2.2 from (Levesque and Lakemeyer 2001). Suppose $n_1 = A(\hat{n}_1, \ldots, \hat{n}_l)$ for arbitrary A and $\hat{n}_i$ for the rest of the proof. Let $\mathcal{A}$, K, $\mathcal{N}'$, $\mathcal{N}$, $\mathcal{M}$, $\mathcal{M}'$, A' be as in Definition 5.9.1 for the case of a free action variables. In analogy to the three disjuncts in $\mathrm{RES}[\![\psi, \phi]\!]$ in case of a free variable a in ψ, we consider three different cases.

Suppose $n_1 \in \mathcal{M}$. Then all but one of the disjuncts in $\mathrm{RES}[\![\psi, \phi]\!]^{x_1}_{n_1}$ are certainly false; the remaining one is the one which contains $(x_1 = n_1)$. Then $\models \mathrm{RES}[\![\psi, \phi]\!]^{x_1 \ldots x_k}_{n_1 \ldots n_k}$ iff (by Definition 5.9.1) $\models \mathrm{RES}[\![\psi^{x_1}_{n_1}, \phi]\!]^{x_2 \ldots x_k}_{n_2 \ldots n_k}$ iff (by induction) $\models (\phi \supset \psi^{x_1\, x_2 \ldots x_k}_{n_1\, n_2 \ldots n_k})$.

Suppose $A \in \mathcal{A}$ and at least one $\hat{n}_i \notin \mathcal{N}$. Then there is some $n_1^\circ = A(\hat{n}_1^\circ, \ldots, \hat{n}_l^\circ) \in \mathcal{M}'$ so that either $\hat{n}_i = \hat{n}_i^\circ \in \mathcal{N}$, or $\hat{n}_i \notin \mathcal{N}$, $\hat{n}_i^\circ \notin \mathcal{N}$, and $\hat{n}_i^\circ \in \mathcal{N}'$ (*), and moreover $\hat{n}_i^\circ = \hat{n}_j^\circ$ iff $\hat{n}_i = \hat{n}_j$ (**). For this n_1°, $\exists y_1 \ldots \exists y_K ((n_1 = n_1^\circ)^{n'_1 \ldots n'_K}_{y_1 \ldots y_K} \wedge \bigwedge_{1 \leq i \leq K, n'' \in \mathcal{N}} (y_i \neq n'') \wedge \bigwedge_{1 \leq i < j \leq K} (y_i \neq y_j))$ is valid. Namely, for every $\hat{n}_i^\circ \notin \mathcal{N}$, there is some j such that $\hat{n}_i^\circ = n'_j \in \mathcal{N}'$; so $\hat{n}_i^\circ$ is replaced with y_j in the mentioned formula; the formula thus comes out true in the semantics by substituting the corresponding $\hat{n}_i$ for y_j. Therefore $\models \mathrm{RES}[\![\psi, \phi]\!]^{x_1 \ldots x_k}_{n_1 \ldots n_k}$ iff (by Definition 5.9.1) $\models \mathrm{RES}[\![\psi^{x_1}_{n_1^\circ}, \phi]\!]^{n_1^\circ\, x_1\, x_2 \ldots x_k}_{x_1\, n_1\, n_2 \ldots n_k}$. Let $*$ be the bijection so that $(\hat{n}_i^\circ)^* = \hat{n}_i$ and $\hat{n}_i^* = \hat{n}_i^\circ$ for all i where $\hat{n}_i \notin \mathcal{N}$, and $A(\vec{n}) = A(\vec{n}^*)$, and the rest is left unchanged. By (**), the bijection is well-defined. By (*) and since RES does not introduce new names, these $\hat{n}_i^\circ$ and $\hat{n}_i$ do not occur in $\mathrm{RES}[\![\psi^{x_1}_{n_1^\circ}, \phi]\!]^{n_1^\circ}_{x_1}$; hence $*$ is preserving with respect to $\mathrm{RES}[\![\psi^{x_1}_{n_1^\circ}, \phi]\!]^{n_1^\circ}_{x_1}$. Therefore, $\models \mathrm{RES}[\![\psi^{x_1}_{n_1^\circ}, \phi]\!]^{n_1^\circ\, x_1\, x_2 \ldots x_k}_{x_1\, n_1\, n_2 \ldots n_k}$ iff (by Corollary B.5.4) $\models \mathrm{RES}[\![\psi^{x_1}_{n_1^\circ}, \phi]\!]^{n_1^\circ\, x_1\, x_2 \ldots x_k}_{x_1\, n_1^*\, n_2^* \ldots n_k^*}$ iff (since by definition of $*$, $n_1^* = n_1^\circ$) $\models \mathrm{RES}[\![\psi^{x_1}_{n_1^\circ}, \phi]\!]^{x_2 \ldots x_k}_{n_2^* \ldots n_k^*}$ iff (by induction) $\models (\phi \supset \psi^{x_1\, x_2 \ldots x_k}_{n_1^\circ\, n_2^* \ldots n_k^*})$ iff (by Lemma B.5.2 and Corollary B.5.4) $\models (\phi \supset \psi^{x_1\;\, x_2\; \ldots x_k}_{n_1^{\circ *}\, n_2^{**} \ldots n_k^{**}})$ iff (since by definition of $*$, $n_1^{\circ *} = n_1$) $\models (\phi \supset$

$\psi^{x_1\,x_2\,\dots\,x_k}_{n_1\,n_2\,\dots\,n_k})$.

Suppose $A \notin \mathcal{A}$. Then $\models \mathrm{RES}[\![\psi,\phi]\!]^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$ iff $\models \mathrm{RES}[\![\psi^{x_1}_{A'},\phi]\!]^{A'\,x_1\,x_2\,\dots\,x_k}_{x_1\,n_1\,n_2\,\dots\,n_k}$. Let $*$ be a bijection so that $n_1^* = A'$ and $A'^* = n_1$ and leaves the rest unchanged. As neither A nor A' occur in ψ or ϕ, $*$ is preserving with respect to $\mathrm{RES}[\![\psi^{x_1}_{A'},\phi]\!]^{A'}_{x_1}$. Then $\models \mathrm{RES}[\![\psi^{x_1}_{A'},\phi]\!]^{A'\,x_1\,x_2\,\dots\,x_k}_{x_1\,n_1\,n_2\,\dots\,n_k}$ iff (by Corollary B.5.4) $\models \mathrm{RES}[\![\psi^{x_1}_{A'},\phi]\!]^{A'\,x_1\,x_2\,\dots\,x_k}_{x_1\,n_1^*\,n_2^*\,\dots\,n_k^*}$ iff $\models \mathrm{RES}[\![\psi^{x_1}_{A'},\phi]\!]^{x_2\,\dots\,x_k}_{n_2^*\,\dots\,n_k^*}$ iff (by induction) $\models (\phi \supset \psi^{x_1\,x_2\,\dots\,x_k}_{A'\,n_2^*\,\dots\,n_k^*})$ iff (by Corollary B.5.4) $\models (\phi \supset \psi^{x_1\;\;x_2\;\;\dots\,x_k}_{A'^*\,n_2^{**}\,\dots\,n_k^{**}})$ iff $\models (\phi \supset \psi^{x_1\,x_2\,\dots\,x_k}_{n_1\,n_2\,\dots\,n_k})$. □

Lemma B.5.6 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1,\dots,\phi_m \Rightarrow \psi_m\}$ be objective. Let $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$ and $\vec{\gamma}$ be an objective representation of $\mathbf{O}_{\mathcal{S}}\Gamma$. Let α be a belief-static formula without* **O** *with free variables $x_1,\dots,x_k$ and let $n_1,\dots,n_k$ be standard names of corresponding sorts. Then $\vec{e}, w \models \alpha^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$ iff $w \models (\|\alpha\|_{\vec{\gamma}})^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$.*

Proof. The proof is by induction on the length α where the length of $\mathbf{B}(\alpha \Rightarrow \beta)$ is taken to be the length of $(\alpha \supset \beta)$ plus 1. For objective α (which includes the base cases for rigid and fluent atoms and equalities, and, by the belief-static assumption, also the cases for $[t]\alpha$ and $\square\alpha$), the lemma clearly holds since $\|\alpha\|_{\vec{\gamma}} = \alpha$. The induction steps for negation, conjunction, and quantification are straightforward. The induction step for $\mathbf{B}(\alpha \Rightarrow \beta)$ uses the fact from Theorem 5.3.16 that only the first $m+1$ levels of $\vec{e}$ may differ and we thus may limit our consideration to levels $1 \le p \le m+1$, each of which corresponds to γ_p. More precisely, $\vec{e} \models \mathbf{B}(\alpha \Rightarrow \beta)^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$ iff for all $p \in \mathbb{P}$, $\vec{e}, w' \models (\alpha \supset \beta)$ for all $p \le \lfloor \vec{e} \mid \alpha \rfloor$ iff (by Theorem 5.3.16) for all $1 \le p \le m+1$, if for all $p' < p$, for all $w' \in e_{p'}$, $\vec{e}, w' \models \neg\alpha$, then for all $w' \in e_p$, $\vec{e}, w' \models (\alpha \supset \beta)$ iff (by induction) for all $1 \le p \le m+1$, if for all $p' < p$, for all $w' \in e_{p'}$, $w' \models (\|\neg\alpha\|_{\vec{\gamma}})^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$, then for all $w' \in e_p$, $w' \models (\|(\alpha \supset \beta)\|_{\vec{\gamma}})^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$ iff (by Lemma B.5.5 since $e_p = \{w \mid w \models \gamma_p\}$ for all $p \in \mathbb{P}$) for all $1 \le p \le m+1$, if for all $1 \le p' < p$, $w \models \mathrm{RES}[\![\|\neg\alpha\|_{\vec{\gamma}}, \gamma'_p]\!]^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$, then $w \models \mathrm{RES}[\![\|(\alpha \supset \beta)\|_{\vec{\gamma}}, \gamma_p]\!]^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$ iff $w \models \bigwedge_{p=1}^{m+1}((\bigwedge_{p'=1}^{p-1} \mathrm{RES}[\![\|\neg\alpha\|_{\vec{\gamma}}, \gamma_{p'}]\!]) \supset \mathrm{RES}[\![\|(\alpha \supset \beta)\|_{\vec{\gamma}}, \gamma_p]\!])^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$ iff (by Definition 5.9.6) $w \models (\|\mathbf{B}(\alpha \Rightarrow \beta)\|_{\vec{\gamma}})^{x_1\,\dots\,x_k}_{n_1\,\dots\,n_k}$. □

Theorem 5.9.7 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1,\dots,\phi_m \Rightarrow \psi_m\}$ be objective and α be belief-static without* **O**. *Then $\mathbf{O}_{\mathcal{S}}\Gamma \models \alpha$ iff $\models \|\alpha\|_{\mathbf{O}_{\mathcal{S}}\Gamma}$.*

Proof. $\mathbf{O}_{\mathcal{S}}\Gamma \models \alpha$ iff $\vec{e} \models \alpha$ for every $\vec{e} \models \mathbf{O}_{\mathcal{S}}\Gamma$ iff (by Lemma B.5.6 and since $\vec{e}$ exists and is unique by Corollary 5.6.4) $\models \|\alpha\|_{\vec{\gamma}}$. □

Theorem 4.8.5 *Let $\Gamma = \{\phi_1 \Rightarrow \psi_1,\dots,\phi_m \Rightarrow \psi_m\}$ be objective where ϕ_i, ψ_i are formulas of $\mathcal{BO}$, and let α be a formula of $\mathcal{BO}$ without* **O**. *Then $\mathbf{O}\Gamma \models_{\mathcal{BO}} \alpha$ iff $\models_{\mathcal{BO}} \|\alpha\|_{\mathbf{O}\Gamma}$.*

Proof. By Theorem 5.3.9, the representation theorem of $\mathcal{BO}$ is just a special case of Theorem 5.9.7. □

Corollary 5.9.8 $\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}$ *be an $\mathcal{S}$-free basic action theory and let α be a regressable sentence. Then* $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models \alpha$ *iff* $\models \|\mathcal{R}[\alpha]\|_{\mathbf{O}_{\mathcal{S}}\Sigma_{\text{bel}}}$.

Proof. $\models \|\mathcal{R}[\alpha]\|_{\mathbf{O}_{\mathcal{S}}\Sigma_{\text{bel}}}$ iff (by Theorem 5.9.7) $\mathbf{O}_{\mathcal{S}}\Sigma_{\text{bel}} \models \mathcal{R}[\alpha]$ iff (by Theorem 5.3.16) $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models \alpha$. □

Corollary 5.9.9 $\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}$ *be a basic action theory, $\mathcal{S}'$ be the symbols newly introduced by $\Sigma_{\text{bel}} \gg n$, and let α be a belief-static sentence without* $\mathbf{O}$.
Then $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models [n]\alpha$ *iff* $\models \|\alpha\|_{\mathbf{O}_{\mathcal{S}\cup\mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)}$.

Proof. $\models \|\alpha\|_{\mathbf{O}_{\mathcal{S}\cup\mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n)}$ iff (by Theorem 5.9.7) $\mathbf{O}_{\mathcal{S}\cup\mathcal{S}'}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}} \gg n) \models \alpha$ iff (by Theorem 5.8.3) $\mathbf{O}_{\mathcal{S}}(\Sigma_{\text{dyn}}, \Sigma_{\text{bel}}) \models [n]\alpha$. □

C Long Proofs for $\mathcal{L}^-$

C.1 Proof of the decidability theorems

Here we show correctness of the decision procedures for $\mathrel{|\!\!\approx^{\circ}}$ and $\mathrel{|\!\!\approx^{\bullet}}$ from Section 6.8 for the case of proper$^+$ knowledge bases. The proof is quite tedious. After proving some general lemmas, we show that only a finite number of literals needs to be considered when splitting or adding a literal (Lemmas C.1.5 and C.1.13). Then we show that finitely many names suffice for the quantifiers in the query (Lemma C.1.14), and that the same holds for the quantifiers in the knowledge base (Lemma C.1.16). With these results, we can finally prove Theorems 6.8.7 and 6.8.11. At the end we show the complexity results Theorems 6.8.8 and 6.8.12.

For the rest of this section, let π denote a proper$^+$ sentence.

Definition C.1.1 Let $*$ be a bijection between standard names. For a set of standard names N, we let $N^* = \{n^* \mid n \in N\}$, and we write c^*, s^*, ϕ^* for the corresponding clause, setup, formula where every name n is replaced with n^*.

Note that in general N and N^* may be distinct. When N is the set of all standard names, however, $N = N^*$.

Lemma C.1.2 *Let $*$ be a bijection between standard names. Let N be a set of standard names. Then* $\mathsf{gnd}_N(\pi)^* = \mathsf{gnd}_{N^*}(\pi^*)$.

Proof. Let $\pi = \bigwedge \forall \vec{x}_j c_j$. Then $c^* \in \mathsf{gnd}_N(\pi)^*$ iff $c^* = ({c_j}^{\vec{x}_j}_{\vec{n}})^*$ for some $n_i \in N$ for some j iff $c^* = (c_j^*)^{\vec{x}_j}_{\vec{n}^*}$ for some $n_i \in N$ for some j iff $c^* \in \mathsf{gnd}_{N^*}(\pi^*)$. □

Lemma C.1.3 *Let $*$ be a bijection between standard names.*

(i) $c \in \mathsf{UP}(s)$ *iff* $c^* \in \mathsf{UP}(s^*)$*;*

(ii) $c \in \mathsf{UP}^+(s)$ *iff* $c^* \in \mathsf{UP}^+(s^*)$*;*

(iii) $c \in \mathsf{UP}^-(s)$ *iff* $c^* \in \mathsf{UP}^-(s^*)$*;*

(iv) $c \in \mathsf{XP}(s)$ *iff* $c^* \in \mathsf{XP}(s^*)$*;*

(v) $c \in \mathsf{L}(\ell, s)$ *iff* $c^* \in \mathsf{L}(\ell^*, s^*)$.

Proof. (i) By induction on the length of the derivation of c. The base case holds trivially. For the induction step, $c \in \mathsf{UP}(s) \setminus (\mathsf{EQ} \cup s)$ iff $c \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s)$ for some $[\ell]$ iff (by induction) $c^* \cup [\ell^*], [\overline{\ell^*}] \in \mathsf{UP}(s^*)$ iff $c^* \in \mathsf{UP}(s^*)$.

(ii) $c \in \mathsf{UP}^+(s)$ iff $c' \in \mathsf{UP}(s)$ for some $c' \subseteq c$ iff (by (i)) $c'^* \in \mathsf{UP}(s^*)$ for some $c' \subseteq c$ iff $c' \in \mathsf{UP}(s^*)$ for some $c' \subseteq c^*$ iff $c^* \in \mathsf{UP}^+(s^*)$.

(iii) $c \in \mathsf{UP}^-(s)$ iff $c \in \mathsf{UP}(s)$ and for all $c' \supsetneq c$, $c' \notin \mathsf{UP}(s)$ iff (by (i)) $c^* \in \mathsf{UP}(s^*)$ and for all $c' \supsetneq c$, $c'^* \notin \mathsf{UP}(s^*)$ iff $c^* \in \mathsf{UP}(s^*)$ and for all $c' \supsetneq c^*$, $c' \notin \mathsf{UP}(s^*)$ iff $c^* \in \mathsf{UP}^-(s^*)$.

(iv) $c \in \mathsf{XP}(s)$ iff $c \in \mathsf{UP}^-(s \uplus \ell \mid \text{for some } c, c \cup [\ell] \in \mathsf{UP}^-(s)\})$ iff (by (iii)) $c^* \in \mathsf{UP}^-(s^* \uplus \ell^* \mid \text{for some } c, c \cup [\ell] \in \mathsf{UP}^-(s)\})$ iff (by (iii)) $c^* \in \mathsf{UP}^-(s^* \uplus \ell^* \mid \text{for some } c, c^* \cup [\ell^*] \in \mathsf{UP}^-(s^*)\})$ iff $c^* \in \mathsf{UP}^-(s^* \uplus \ell \mid \text{for some } c, c \cup [\ell] \in \mathsf{UP}^-(s^*)\})$ iff $c^* \in \mathsf{XP}(s^*)$.

(v) $c \in \mathsf{L}(\ell, s)$ iff $c = [\ell']$ and $[\ell'] \in \mathsf{gnd}([\ell])$ and $[\overline{\ell'}] \notin \mathsf{UP}(s)$ for some ℓ' iff (by Lemma C.1.2 and (i)) $c = [\ell']$ and $[\ell'^*] \in \mathsf{gnd}([\ell^*])$ and $[\overline{\ell'^*}] \notin \mathsf{UP}(s^*)$ for some ℓ' iff $c^* = [\ell']$ and $[\ell'] \in \mathsf{gnd}([\ell^*])$ and $[\overline{\ell'}] \notin \mathsf{UP}(s^*)$ for some ℓ' iff $c^* \in \mathsf{L}(\ell^*, s^*)$. □

Lemma C.1.4 *Let $*$ be a bijection between standard names.*

(i) $s, k \mathrel{|\!\approx} \phi$ iff $s^*, k \mathrel{|\!\approx} \phi^*$;

(ii) $s, l \mathrel{|\!\approx} \phi$ iff $s^*, l \mathrel{|\!\approx} \phi^*$.

Proof. (i) By induction on k. For the base case $k = 0$ we do a subinduction on the length of ϕ. For a clause, $s, 0 \mathrel{|\!\approx} c$ iff $c \in \mathsf{UP}^+(s)$ iff (by Lemma C.1.3) $c^* \in \mathsf{UP}^+(s^*)$ iff $s^*, 0 \mathrel{|\!\approx} c^*$. The subinduction steps for a non-clausal disjunction, negated disjunction, and double negation are trivial. For an existential, $s, 0 \mathrel{|\!\approx} \exists x\, \phi$ iff $s, 0 \mathrel{|\!\approx} \phi^x_n$ for some n iff (by subinduction) $s^*, 0 \mathrel{|\!\approx} (\phi^x_n)^*$ for some n iff $s^*, 0 \mathrel{|\!\approx} (\phi^*)^x_{n^*}$ for some n iff $s^*, 0 \mathrel{|\!\approx} \exists x\, \phi^*$ iff $s^*, 0 \mathrel{|\!\approx} (\exists x\, \phi)^*$. The case for a negated existential is analogous. For the main induction, suppose (i) holds for k. Then $s, k+1 \mathrel{|\!\approx} \phi$ iff $s \uplus \ell, k \mathrel{|\!\approx} \phi$ and $s \uplus \overline{\ell}, k \mathrel{|\!\approx} \phi$ for some ℓ iff (by induction) $(s \uplus \ell)^*, k \mathrel{|\!\approx} \phi^*$ and $(s \uplus \overline{\ell})^*, k \mathrel{|\!\approx} \phi^*$ for some ℓ iff $s^* \uplus \ell^*, k \mathrel{|\!\approx} \phi^*$ and $s^* \uplus \overline{\ell^*}, k \mathrel{|\!\approx} \phi^*$ for some ℓ iff $s^*, k+1 \mathrel{|\!\approx} \phi^*$.

(ii) By induction on l very similar to (i). For the base case $l = 0$ we do a subinduction on the length of ϕ. For a negated clause, $s, 0 \mathrel{|\!\approx} \neg c$ iff $[] \in \mathsf{XP}(s)$ or $c \notin \mathsf{UP}^+(s)$ iff (by Lemma C.1.3) $[] \in \mathsf{XP}(s^*)$ or $c^* \notin \mathsf{UP}^+(s^*)$ iff $s^*, 0 \mathrel{|\!\approx} c^*$. The subinduction steps for a literal, a disjunction, negated non-clausal disjunction, and double negation are trivial. The subinduction steps for an existential and a negated existential are analogous to (i). For the main induction, suppose (ii) holds for l. Then $s, l+1 \mathrel{|\!\approx} \phi$ iff $s \otimes \ell, l \mathrel{|\!\approx} \phi$ for all ℓ iff (by induction) $(s \otimes \ell)^*, l \mathrel{|\!\approx} \phi^*$ for all ℓ iff (by Lemma C.1.3) $s^* \otimes \ell^*, l \mathrel{|\!\approx} \phi^*$ for all ℓ iff $s^*, l+1 \mathrel{|\!\approx} \phi^*$. □

Finitely many literals suffice for splitting and adding

Here we observe that the number of relevant split literals in Rule $\mathcal{L}^{\circ}1$ and added literals in Rule $\mathcal{L}^{\bullet}1$ are in fact finite.

Lemma C.1.5 *Let $v \geq |\pi|_{\mathrm{w}}$ and $v \geq |\phi|_{\mathrm{w}}$, and let N_j contain the names from π and ϕ plus new names $n_1, \ldots, n_j$. Then $\mathsf{gnd}(\pi), k+1 \mathrel{|\!\!\approx} \phi$ iff $\mathsf{gnd}(\pi) \uplus \ell, k \mathrel{|\!\!\approx} \phi$ and $\mathsf{gnd}(\pi) \uplus \overline{\ell}, k \mathrel{|\!\!\approx} \phi$ for some ℓ whose symbol occurs in π or ϕ and whose names are from $N_{(k+1)\cdot v}$.*

Proof. The if direction holds immediately by the semantics. For the only-if direction suppose $\mathsf{gnd}(\pi), k+1 \mathrel{|\!\!\approx} \phi$. Then $\mathsf{gnd}(\pi) \uplus \ell, k \mathrel{|\!\!\approx} \phi$ and $\mathsf{gnd}(\pi) \uplus \overline{\ell}, k \mathrel{|\!\!\approx} \phi$ for some ℓ. We first show that it suffices to consider for ℓ only symbols from π or ϕ; in a second step we show that considering in ℓ only names from $N_{(k+1)\cdot v}$ suffices as well.

For the first step suppose the symbol of ℓ occurs neither in π nor in ϕ or is an equality literal. We show that then $\mathsf{gnd}(\pi), k \mathrel{|\!\!\approx} \phi$. By Lemma 6.5.3 this implies that we can split an arbitrary literal, that is, choose any symbol that occurs in π or ϕ for the split literal. The proof is by induction on k. The base case $k = 0$ needs a subinduction on the length of ϕ. For a clause, if ℓ is an equality literal, $\mathsf{gnd}(\pi) \uplus \ell, 0 \mathrel{|\!\!\approx} c$ and $\mathsf{gnd}(\pi) \uplus \overline{\ell}, 0 \mathrel{|\!\!\approx} c$ iff $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi) \uplus \ell)$ and $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi) \uplus \overline{\ell})$ iff (since $\ell \in \mathsf{UP}^+(\mathsf{gnd}(\pi))$ or $\overline{\ell} \in \mathsf{UP}^+(\mathsf{gnd}(\pi))$) $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi))$ iff $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} c$; and otherwise, if the symbol of ℓ is not equality and does not occur in π or ϕ, $\mathsf{gnd}(\pi) \uplus \ell, 0 \mathrel{|\!\!\approx} c$ and $\mathsf{gnd}(\pi) \uplus \overline{\ell}, 0 \mathrel{|\!\!\approx} c$ iff $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi) \uplus \ell)$ and $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi) \uplus \overline{\ell})$ iff (since ℓ cannot trigger unit propagation) $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi)) \uplus \ell$ and $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi)) \uplus \overline{\ell}$ iff (since the symbol of ℓ does not occur in c) $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi))$ iff $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} c$ and $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} c$. The subinduction steps are trivial. For the induction step, suppose the claim holds for k. Then $\mathsf{gnd}(\pi) \uplus \ell, k \mathrel{|\!\!\approx} \phi$ and $\mathsf{gnd}(\pi) \uplus \overline{\ell}, k \mathrel{|\!\!\approx} \phi$ iff $\mathsf{gnd}(\pi \wedge \ell), k \mathrel{|\!\!\approx} \phi$ and $\mathsf{gnd}(\pi \wedge \overline{\ell}), k \mathrel{|\!\!\approx} \phi$ only if (by induction) $\mathsf{gnd}(\pi \wedge \ell), k \mathrel{|\!\!\approx} \phi$ and $\mathsf{gnd}(\pi \wedge \overline{\ell}), k \mathrel{|\!\!\approx} \phi$ where the k more split literals have symbols that occur in π, ϕ, or ℓ iff (since for every split literal ℓ', only ℓ' or $\overline{\ell'}$ can resolve with ℓ) $\mathsf{gnd}(\pi \wedge \ell), k \mathrel{|\!\!\approx} \phi$ and $\mathsf{gnd}(\pi \wedge \overline{\ell}), k \mathrel{|\!\!\approx} \phi$ where the k more split literals have symbols that occur in π or ϕ only if $\mathsf{gnd}(\pi), k \mathrel{|\!\!\approx} \phi$.

For the second step suppose the symbol of ℓ occurs in π or ϕ but it mentions names not in $N_{(k+1)\cdot v}$. Let $n'_1, \ldots, n'_l \notin N_{(k+1)\cdot v}$ be those names. If ℓ is an equality expression, then $N_{(k+1)\cdot v}$ must be non-empty, and by the first case we can split an arbitrary equality literal formed from $N_{(k+1)\cdot v}$ instead. Otherwise, the arity of ℓ is at most v, and thus $l \leq v$. Without loss of generality suppose $n_1, \ldots, n_l \in N_{(k+1)\cdot v}$ do not occur in ℓ. Let $*$ be the bijection that swaps the n_i and n'_i and leaves the rest unchanged. Then $\mathsf{gnd}(\pi) \uplus \ell, k \mathrel{|\!\!\approx} \phi$ and $\mathsf{gnd}(\pi) \uplus \overline{\ell}, k \mathrel{|\!\!\approx} \phi$ iff (by Lemma C.1.4) $(\mathsf{gnd}(\pi) \uplus \ell)^*, k \mathrel{|\!\!\approx} \phi^*$ and $(\mathsf{gnd}(\pi) \uplus \overline{\ell})^*, k \mathrel{|\!\!\approx} \phi^*$ iff $\mathsf{gnd}(\pi)^* \uplus \ell^*, k \mathrel{|\!\!\approx} \phi^*$ and $\mathsf{gnd}(\pi)^* \uplus \overline{\ell^*}, k \mathrel{|\!\!\approx} \phi^*$ iff (by

Lemma C.1.2) $\mathsf{gnd}(\pi) \uplus \ell^*, k \mathrel{|\!\!\approx} \phi^*$ and $\mathsf{gnd}(\pi) \uplus \overline{\ell^*}, k \mathrel{|\!\!\approx} \phi^*$. Since $\phi^* = \phi$ and ℓ^* mentions names only from $N_{(k+1)\cdot v}$, this obtains the lemma. □

Now we prove similar results for the unsound semantics, Lemma C.1.13. To this end, we first need a syntactic representation of $\mathsf{gnd}(\pi) \otimes \ell$.

Definition C.1.6 Let N contain all and only the names from π and ℓ. Let $x_1, \ldots, x_l$ be the variables in ℓ. We define $\pi \otimes \ell$ as $\pi \wedge \Pi$ where Π is the least set such that, if $[\overline{\ell^{x_1 \ldots x_l}_{n_1 \ldots n_l}}] \notin \mathsf{UP}(\mathsf{gnd}(\pi))$ for $n_1, \ldots, n_k \notin N$ and $n_{k+1}, \ldots, n_l \in N$, then $e \vee e_= \vee e_{\neq} \vee \ell^{x_{k+1} \ldots x_l}_{n_{k+1} \ldots n_l} \in \Pi$ where $e = \bigvee_{1 \le i \le k \text{ and } n \in N} (x_i = n)$ and $e_= = \bigvee_{1 \le i < j \le k \text{ and } n_i = n_j} (x_i \neq x_j)$ and $e_{\neq} = \bigvee_{1 \le i < j \le k \text{ and } n_i \neq n_j} (x_i = x_j)$.

A few words about this definition are in order. The idea is to represent every ground instance added in $\mathsf{gnd}(\pi) \otimes \ell$ by one clause in Π. More intuitively, this clause can be read as material implication

$$\bigwedge_{1 \le i \le k,\, n \in N} (x_i \neq n) \wedge \bigwedge_{1 \le i < j \le k,\, n_i = n_j} (x_i = x_j) \wedge \bigwedge_{1 \le i < j \le k,\, n_i \neq n_j} (x_i \neq x_j) \supset [\ell^{x_{k+1} \ldots x_l}_{n_{k+1} \ldots n_l}].$$

(We do not use $\supset$ in the definition of Π in order to keep Π proper$^+$. For example, $P \wedge Q \supset R$ stands for $\neg\neg(\neg P \vee \neg Q) \vee R$, which is not proper$^+$ due to the double negation.) Notice that this clause does not mention the names $n_1, \ldots, n_k \notin N$, and hence Π introduces no new names. Instead, the names $n_1, \ldots, n_k$ are replaced with variables, and the formulas e, $e_=$, $e_{\neq}$ ensure that any instance of ℓ derived from this clause corresponds to an instance of ℓ added to $\mathsf{gnd}(\pi) \otimes \ell$. For example, let $\pi = \forall x \neg P(x, x)$. Then $\mathsf{gnd}(\pi) = \{[\neg P(n, n)] \mid n \text{ is a name}\}$, and $\mathsf{gnd}(\pi) \otimes P(x, x) = \{[\neg P(n, n) \mid n \text{ is a name}\} \cup \{[P(n, n') \mid n, n' \text{ are distinct names}\}$. Therefore Π contains the single clause $(x_1 = x_2) \vee P(x_1, x_2)$ that represents all $[P(n, n')] \notin \mathsf{gnd}(\pi)$: grounding Π and doing unit propagation with EQ just obtains $\{[P(n, n')] \mid n, n' \text{ are distinct names}\}$.

Lemma C.1.7 *$\pi \otimes \ell$ is a well-defined sentence, proper$^+$, contains only names that occur in π or ℓ, and* $|\pi \otimes \ell|_{\mathrm{w}} \le \max\{|\pi|_{\mathrm{w}}, |\ell|_{\mathrm{w}}\}$.

Proof. Let Π and N be as in Definition C.1.6. Since N is finite, every element in Π is a well-defined formula. Since there are only finitely many variables in $\ell^{x_{k+1} \ldots x_l}_{n_{k+1} \ldots n_l}$, the set Π must be finite as well. Hence $\pi \otimes \ell$ is well-defined. Since the formulas in Π are clauses with universally quantified variables, $\pi \otimes \ell$ is proper$^+$. Π introduces no new names, so $\pi \otimes \ell$ does not either. Moreover, Π does not mention more variables than ℓ and does not mention more of them freely in any subformula than ℓ does, so $|\Pi|_{\mathrm{w}} \le |\ell|_{\mathrm{w}}$. □

Lemma C.1.8 $\mathsf{UP}^-(\mathsf{gnd}_{N'}(\pi) \otimes_{N'} \ell) = \mathsf{UP}^-(\mathsf{gnd}_{N'}(\pi \otimes \ell))$.

Proof. Let Π and N be as in Definition C.1.6. It suffices to show $\mathsf{UP}(\mathsf{L}_{N'}(\ell, \mathsf{gnd}_{N'}(\pi))) = \mathsf{UP}^-(\mathsf{gnd}_{N'}(\Pi))$, as the lemma then follows from Lemma 6.8.2.

First we show that $\mathsf{UP}^-(\mathsf{gnd}_{N'}(\Pi)) \setminus \mathsf{EQ}$ only contains unit clauses of the form $[\ell']$ where $\ell' = \ell^{x_1 \ldots x_l}_{n_1 \ldots n_l}$ (*). Clearly every clause in $\mathsf{UP}^-(\mathsf{gnd}_{N'}(\Pi)) \setminus \mathsf{EQ}$ is a subclause of some clause from $\mathsf{gnd}_{N'}(\Pi)$. Hence it suffices to show that $[] \notin \mathsf{UP}(\mathsf{gnd}_{N'}(\Pi))$ and that every clause from $\mathsf{gnd}_{N'}(\Pi)$ is subsumed by some unit clause in $\mathsf{UP}(\mathsf{gnd}_{N'}(\Pi))$. Let $c \in \mathsf{gnd}_{N'}(\Pi)$. Then $c = c' \cup [\ell']$ where c' only contains equality literals. If c' is subsumed by EQ, then so is c. Otherwise, unit resolution of c' with EQ yields $[\ell'] \in \mathsf{UP}(\mathsf{gnd}_{N'}(\Pi))$, which subsumes c. Moreover, $c \in \mathsf{gnd}_{N'}(e \vee e_= \vee e_{\neq} \vee \ell^{x_{k+1} \ldots x_l}_{n_{k+1} \ldots n_l})$ for some clause from Π. By Definition C.1.6, $[\overline{\ell''}] \notin \mathsf{UP}(\mathsf{gnd}(\pi))$ where $\ell'' = \ell^{x_1 \ldots x_k x_{k+1} \ldots x_l}_{n'_1 \ldots n'_k n_{k+1} \ldots n_l}$ for some $n'_1, \ldots, n'_k \notin N$. Let $*$ be the bijection that swaps $n_1, \ldots, n_k$ for $n'_1, \ldots, n'_k$ and leaves the rest unchanged. Then $[\overline{\ell''}]^* = [\overline{\ell'}] \notin \mathsf{UP}(\mathsf{gnd}(\pi))$ by Lemmas C.1.3 and C.1.2. Thus $[\overline{\ell'}] \notin \mathsf{EQ}$, so if ℓ' is an equality literal, then $[\ell'] \in \mathsf{EQ}$. Otherwise, $[\overline{\ell'}] \notin \mathsf{UP}(\mathsf{gnd}_{N'}(\Pi))$. Hence $[] \notin \mathsf{UP}(\mathsf{gnd}_{N'}(\Pi))$.

For the only-if direction let $c \in \mathsf{UP}(\mathsf{L}_N(\ell, \mathsf{gnd}_{N'}(\pi)))$. Then by definition of L_N, either $c \in \mathsf{EQ}$ or $c = [\ell'] \in \mathsf{L}_N(\ell, \mathsf{gnd}_{N'}(\pi))$ where $\ell' = \ell^{x_1 \ldots x_l}_{n_1 \ldots n_l}$ for $n_1, \ldots, n_l \in N'$. In the former case, $c \in \mathsf{UP}^-(\mathsf{gnd}_{N'}(\Pi))$ by (*). In the latter case, $[\overline{\ell'}] \notin \mathsf{gnd}_{N'}(\pi)$. Hence $e \vee e_= \vee e_{\neq} \vee \ell^{x_{k+1} \ldots x_l}_{n_{k+1} \ldots n_l} \in \Pi$ for some $k \leq l$, and by construction, $c' = (e \vee e_= \vee e_{\neq})^{x_1 \ldots x_k}_{n_1 \ldots n_k}$ is invalid. Hence $c' \cup [\ell'] \in \mathsf{gnd}_{N'}(\Pi)$, and since c' only mentions invalid equality literals, they are eliminated by unit resolution with EQ, so that $[\ell'] \in \mathsf{UP}(\mathsf{gnd}_{N'}(\Pi))$, and by (*), $c = [\ell'] \in \mathsf{UP}^-(\mathsf{gnd}_{N'}(\Pi))$.

For the if direction let $c \in \mathsf{UP}^-(\mathsf{gnd}_{N'}(\Pi))$. If $c \in \mathsf{EQ}$, then $c \in \mathsf{UP}(\mathsf{L}_N(\ell, \mathsf{gnd}_{N'}(\pi)))$. Otherwise, by (*), $c = [\ell']$ where $\ell' = \ell^{x_1 \ldots x_l}_{n_1 \ldots n_l}$ for $n_1, \ldots, n_l \in N'$. By Definition C.1.6, $e \vee e_= \vee e_{\neq} \vee \ell^{x_{k+1} \ldots x_l}_{n_{k+1} \ldots n_l} \in \Pi$ for some $k \leq l$ and $(e \vee e_= \vee e_{\neq})^{x_1 \ldots x_k}_{n_1 \ldots n_k}$ is invalid. Again by Definition C.1.6, some $n'_1, \ldots, n'_k \notin N$ exist such that $[\overline{\ell''}] \notin \mathsf{UP}(\mathsf{gnd}(\pi))$ where $\ell'' = \ell^{x_1 \ldots x_k x_{k+1} \ldots x_l}_{n'_1 \ldots n'_k n_{k+1} \ldots n_l}$ and such that $n_i = n_j$ iff $n'_i = n'_j$. Therefore there is a bijection $*$ between standard names that swaps $n_1, \ldots, n_k$ and $n'_1, \ldots, n'_k$ and leaves the rest unchanged. Then $[\overline{\ell''}]^* = [\overline{\ell'}] \notin \mathsf{UP}(\mathsf{gnd}_{N'}(\pi))$ by Lemmas C.1.3, C.1.2, and 6.5.2. Therefore $[\ell'] \in \mathsf{L}_{N'}(\ell, \mathsf{gnd}_{N'}(\pi)) \subseteq \mathsf{UP}(\mathsf{L}_{N'}(\ell, \mathsf{gnd}_{N'}(\pi)))$. □

Lemma C.1.9 $\mathsf{UP}^-(\mathsf{gnd}_{N'}(\pi) \otimes_{N'} \ell_1 \otimes_{N'} \ldots \otimes_{N'} \ell_k) = \mathsf{UP}^-(\mathsf{gnd}_{N'}(\pi \otimes \ell_1 \otimes \ldots \otimes \ell_k))$.

Proof. By Lemma 6.8.2, $\mathsf{L}_{N'}(\ell, \mathsf{UP}(s)) = \mathsf{L}_{N'}(\ell, s)$, and therefore again by Lemma 6.8.2, $\mathsf{UP}^-(\mathsf{UP}^-(s) \otimes_{N'} \ell) = \mathsf{UP}^-(s \otimes_{N'} \ell)$. The lemma then follows by a simple induction on k using Lemma C.1.8. □

Lemma C.1.10 *Let the symbol of ℓ not occur in s and* $[] \notin \mathsf{XP}(s \otimes \ell \otimes \ell_1 \otimes \ldots \otimes \ell_l)$. *Then* $[] \notin \mathsf{XP}(s \otimes \ell_1 \otimes \ldots \otimes \ell_l)$.

Proof. Suppose the opposite. Since the symbol of ℓ does not occur in s, there must be some ℓ', ℓ_i, ℓ_j with $j > i$ with the same symbol as ℓ and so that $[\ell'] \in \mathsf{L}(\ell_i, s \otimes \ell_1 \otimes \ldots \otimes \ell_{i-1})$ and $[\overline{\ell'}] \in \mathsf{L}(\ell_j, s \otimes \ell_1 \otimes \ldots \otimes \ell_{j-1})$. However, by definition $[\overline{\ell'}] \in \mathsf{L}(\ell_j, s \otimes \ell_1 \otimes \ldots \otimes \ell_{j-1})$ only if $[\ell'] \notin \mathsf{UP}^+(s \otimes \ell_1 \otimes \ldots \otimes \ell_{i-1})$. Contradiction. □

Lemma C.1.11 *Let the symbol of ℓ_1 not occur in s or c. Then $c \in \mathsf{UP}(s \otimes \ell_1 \otimes \ldots \otimes \ell_l)$ iff $c \in \mathsf{UP}(s \otimes \ell_2 \otimes \ldots \otimes \ell_l)$.*

Proof. By induction on the length of the derivation of c. If $c \in s \otimes \ell_1 \otimes \ldots \otimes \ell_l$, then also $c \in s \otimes \ell_2 \otimes \ldots \otimes \ell_l$ since c does not mention the symbol of ℓ_1. Now suppose $c \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s \otimes \ell_1 \otimes \ldots \otimes \ell_l)$. Suppose the symbol of ℓ is the same as the one of ℓ_1. Then this symbol does not occur in s, so $c = []$ and $[\ell] \in \mathsf{L}(\ell_i, s \otimes \ell_1 \otimes \ldots \otimes \ell_{i-1})$ and $[\overline{\ell}] \in \mathsf{L}(\ell_j, s \otimes \ell_1 \otimes \ldots \otimes \ell_{j-1})$ for some $i \neq j$, which contradicts the definition of L. Hence the symbols of ℓ and ℓ_1 are distinct, so by induction, $c \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(s \otimes \ell_2 \otimes \ldots \otimes \ell_l)$, and thus $c \in \mathsf{UP}(s \otimes \ell_2 \otimes \ldots \otimes \ell_l)$. □

Lemma C.1.12 $\mathsf{gnd}(\pi), l \not\mathrel{|\!\approx} \phi$ *iff* $\mathsf{gnd}(\pi) \otimes \ell_1 \otimes \ldots \otimes \ell_l, 0 \not\mathrel{|\!\approx} \phi$ *for some $\ell_1, \ldots, \ell_l$ whose symbols occur in π or ϕ.*

Proof. The if direction holds immediately by the semantics. For the only-if direction, suppose $\mathsf{gnd}(\pi), l \not\mathrel{|\!\approx} \phi$. Then $\mathsf{gnd}(\pi) \otimes \ell_1 \otimes \ldots \otimes \ell_l, 0 \not\mathrel{|\!\approx} \phi$ for some $\ell_1, \ldots, \ell_l$. We show by induction on i that $\mathsf{gnd}(\pi) \otimes \ell'_1 \otimes \ldots \otimes \ell'_i \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l, 0 \not\mathrel{|\!\approx} \phi$ for some $\ell'_1, \ldots, \ell'_i$ whose symbols occur in π or ϕ. The base case $i = 0$ holds trivially.

For the induction step, suppose $\mathsf{gnd}(\pi) \otimes \ell'_1 \otimes \ldots \otimes \ell'_{i-1} \otimes \ell_i \otimes \ldots \otimes \ell_l, 0 \not\mathrel{|\!\approx} \phi$ for some $\ell'_1, \ldots, \ell'_{i-1}$ whose symbols occur in π or ϕ, and suppose the symbol of ℓ_i occurs neither in π nor in ϕ. Before we prove the induction step, we need two observations ((*) and (**) below).

By assumption, $[] \notin \mathsf{XP}(\mathsf{gnd}(\pi) \otimes \ell'_1 \otimes \ldots \otimes \ell'_{i-1} \otimes \ell_i \otimes \ldots \otimes \ell_l)$. Let $\pi' = \pi \otimes \ell'_1 \otimes \ldots \otimes \ell'_{i-1}$. By Lemmas C.1.9 and 6.8.2, we can move $\ell'_1, \ldots, \ell'_{i-1}$ inside gnd and obtain that $[] \notin \mathsf{XP}(\mathsf{gnd}(\pi') \otimes \ell_i \otimes \ldots \otimes \ell_l)$. By Lemma C.1.10, $[] \notin \mathsf{XP}(\mathsf{gnd}(\pi') \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l)$.

Now let ℓ'_i be as follows. If $i > 1$, let ℓ'_i be just ℓ'_{i-1}. If $i = 1$ and there is an ℓ_j with a symbol that occurs in π or ϕ for with minimal $j > i$, let ℓ'_i be ℓ_j. Otherwise (that is, if $i = 1$ and all $\ell_2, \ldots, \ell_l$ have symbols not from π or ϕ), let ℓ'_i some literal that occurs in $\mathsf{UP}^-(\mathsf{gnd}(\pi')) \setminus \mathsf{EQ}$, or if there is none, some ground literal whose symbol occurs in ϕ that is not of the form $(n \neq n)$ or $(n = n')$ for distinct names n, n'. In all three cases, we have that the symbol of ℓ'_i occurs in π or ϕ, $[] \notin \mathsf{XP}(\mathsf{gnd}(\pi') \otimes \ell'_i \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l)$, and $\mathsf{UP}^+(\mathsf{gnd}(\pi') \otimes \ell'_i \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l) = \mathsf{UP}^+(\mathsf{gnd}(\pi') \otimes \ell'_i \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l)$: in the first case, $\mathsf{gnd}(\pi') \otimes \ell'_i \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l = \mathsf{gnd}(\pi') \otimes \ell'_i \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l$ by Lemma C.1.9; in the second case, it holds by Lemmas C.1.8, C.1.10, and C.1.11; in the third case, it holds since $[\overline{\ell_i'}]$ does not occur in $\mathsf{UP}^-(\mathsf{gnd}(\pi'))$.

By Lemma C.1.11, $\mathsf{UP}(\mathsf{gnd}(\pi') \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l) = \mathsf{UP}(\mathsf{gnd}(\pi') \otimes \ell_i \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l)$. By applying Lemmas C.1.9 and 6.8.2 again to pull $\ell_1', \ldots, \ell_{i-1}'$ back out of gnd, we obtain the observations for the induction step: $[] \notin \mathsf{XP}(\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_i' \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l)$ (*) and $\mathsf{UP}(\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_{i-1}' \otimes \ell_i \otimes \ldots \otimes \ell_l) \subseteq \mathsf{UP}(\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_i' \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l)$ (**).

With (*) and (**), we can show by subinduction on the length of ϕ that $\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_i' \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l, 0 \mathrel{|\!\!\!\approx} \phi$, which proves the induction step. For a negated clause, $\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_{i-1}' \otimes \ell_i \otimes \ldots \otimes \ell_l, 0 \mathrel{|\!\!\!\approx} \neg c$ iff $[] \notin \mathsf{XP}(\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_{i-1}' \otimes \ell_i \otimes \ldots \otimes \ell_l)$ and $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_{i-1}' \otimes \ell_i \otimes \ldots \otimes \ell_l)$ iff (by (*) and (**) and Lemma 6.8.2) $[] \notin \mathsf{XP}(\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_i' \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l)$ and $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_i' \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l)$ iff $\mathsf{gnd}(\pi) \otimes \ell_1' \otimes \ldots \otimes \ell_i' \otimes \ell_{i+1} \otimes \ldots \otimes \ell_l, 0 \mathrel{|\!\!\!\approx} \neg c$. The other cases for the subinduction are trivial. □

Lemma C.1.13 *Let $v \geq |\pi|_{\mathrm{w}}$ and $v \geq |\phi|_{\mathrm{w}}$, and let N_j contain the names from π and ϕ plus new names $n_1, \ldots, n_j$. Then* $\mathsf{gnd}(\pi), l+1 \mathrel{|\!\!\!\approx} \phi$ *iff* $\mathsf{gnd}(\pi) \otimes \ell, l \mathrel{|\!\!\!\approx} \phi$ *for some ℓ whose symbol occurs in π or ϕ and whose names are from $N_{(l+1)\cdot v}$.*

Proof. The if direction holds immediately by the semantics. For the only-if direction suppose $\mathsf{gnd}(\pi), l+1 \mathrel{|\!\!\!\approx} \phi$. Then $\mathsf{gnd}(\pi) \otimes \ell, l \mathrel{|\!\!\!\approx} \phi$ for some ℓ. By Lemma C.1.12 we already have that only symbols from π or ϕ are relevant. So we only need to show that considering names from $N_{(l+1)\cdot v}$ suffices.

Suppose the symbol of ℓ occurs in π or ϕ but it mentions names not in $N_{(l+1)\cdot v}$. Let $n_1', \ldots, n_k' \notin N_{(l+1)\cdot v}$ be those names. If ℓ is an equality expression, then $N_{(l+1)\cdot v}$ must be non-empty, and since $\mathsf{gnd}(\pi) \otimes \ell' = \mathsf{gnd}(\pi)$ for any equality literal ℓ', we have $\mathsf{gnd}(\pi) \otimes (n = n), l \mathrel{|\!\!\!\approx} \phi$ for arbitrary $n \in N_{(l+1)\cdot v}$. Otherwise, the arity of ℓ is at most v, and thus $k \leq v$. Without loss of generality suppose $n_1, \ldots, n_k \in N_{(l+1)\cdot v}$. Let $*$ be the bijection that swaps the n_i and n_i' and leaves the rest unchanged. Then $\mathsf{gnd}(\pi) \otimes \ell, l \mathrel{|\!\!\!\approx} \phi$ iff (by Lemmas C.1.3 and C.1.4) $\mathsf{gnd}(\pi)^* \otimes \ell^*, l \mathrel{|\!\!\!\approx} \phi^*$ iff (Lemma C.1.2) $\mathsf{gnd}(\pi) \otimes \ell^*, l \mathrel{|\!\!\!\approx} \phi^*$. Since $\phi^* = \phi$ and ℓ^* mentions only names from $N_{(l+1)\cdot v}$, this obtains the base case. □

Finitely many names suffice for the quantifiers in the query

The following Lemma C.1.14 allows us to consider only finitely many names when dealing with quantifiers.

Lemma C.1.14 *Let x be a free variable in ϕ, and let n, n' be two names that do not occur in π or ϕ.*

(i) $\mathsf{gnd}(\pi), k \mathrel{|}\!\overset{\circ}{\approx} \phi^x_n$ *iff* $\mathsf{gnd}(\pi), k \mathrel{|}\!\overset{\circ}{\approx} \phi^x_{n'}$*;*

(ii) $\mathsf{gnd}(\pi), l \mathrel{|}\!\overset{\bullet}{\approx} \phi^x_n$ *iff* $\mathsf{gnd}(\pi), l \mathrel{|}\!\overset{\bullet}{\approx} \phi^x_{n'}$*.*

Proof. Let N be the set of all standard names, and $*$ be the bijection that swaps n and n' and leaves the rest unchanged. Note that $N = N^*$.

(i) $\mathsf{gnd}(\pi), k \mathrel{|}\!\overset{\circ}{\approx} \phi^x_n$ iff (by Lemma C.1.4) $\mathsf{gnd}(\pi)^*, k \mathrel{|}\!\overset{\circ}{\approx} (\phi^x_n)^*$ iff (by Lemma C.1.2) $\mathsf{gnd}_{N^*}(\pi^*), k \mathrel{|}\!\overset{\circ}{\approx} (\phi^x_n)^*$ iff (since π and ϕ contain no n') $\mathsf{gnd}(\pi), k \mathrel{|}\!\overset{\circ}{\approx} \phi^x_{n'}$.

(ii) $\mathsf{gnd}(\pi), l \mathrel{|}\!\overset{\bullet}{\approx} \phi^x_n$ iff (by Lemma C.1.4) $\mathsf{gnd}(\pi)^*, l \mathrel{|}\!\overset{\bullet}{\approx} (\phi^x_n)^*$ iff (by Lemma C.1.2) $\mathsf{gnd}_{N^*}(\pi^*), l \mathrel{|}\!\overset{\bullet}{\approx} (\phi^x_n)^*$ iff (since π and ϕ contain no n') $\mathsf{gnd}(\pi), l \mathrel{|}\!\overset{\bullet}{\approx} \phi^x_{n'}$. □

Finitely many names suffice for the quantifiers in the knowledge base

Here we show that grounding with respect to a finite set of names is sufficient.

Lemma C.1.15 *Let $v \geq |\pi|_{\mathrm{w}}$ and let N contain all names from π. Then any clause $c \in \mathsf{UP}(\mathsf{gnd}(\pi)) \setminus \mathsf{EQ}$ mentions at most v names not in N.*

Proof. By induction on the length of the derivation of c. For the base case, let $\pi = \bigwedge \forall \vec{x}_j c_j$ and $c \in \mathsf{gnd}(\pi)$. If $c \in \mathsf{gnd}(\pi)$, $c = {c_j}^{\vec{x}_j}_{\vec{n}}$ for some j and $\vec{n}$, and since c_j has at most v variables by assumption, the lemma holds again. For the induction step, $c \in \mathsf{UP}(\mathsf{gnd}(\pi)) \setminus (\mathsf{EQ} \cup \mathsf{gnd}(\pi))$ only if $c \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(\mathsf{gnd}(\pi))$ for some ℓ. By induction, either $c \cup [\ell] \in \mathsf{EQ}$, in which case $c = []$ satisfies the lemma, or $c \cup [\ell]$ mentions at most v names not in N, and hence c satisfies the lemma as well. □

Lemma C.1.16 *Let $v \geq |\pi|_{\mathrm{w}}$ and let N contain all names from π plus new names $n_1, \ldots, n_v$. Let c mention names only from $(N \setminus \{n_1, \ldots, n_l\}) \cup \{n'_1, \ldots, n'_l\}$ where $n'_1, \ldots, n'_l \notin N$, $l \leq v$. Let $*$ be the bijection that swaps n_i with n'_i.*

(i) $c \in \mathsf{UP}(\mathsf{gnd}(\pi))$ *iff* $c^* \in \mathsf{UP}(\mathsf{gnd}_N(\pi))$*;*

(ii) $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi))$ *iff* $c^* \in \mathsf{UP}^+(\mathsf{gnd}_N(\pi))$*;*

(iii) $c \in \mathsf{UP}^-(\mathsf{gnd}(\pi))$ *iff* $c^* \in \mathsf{UP}^-(\mathsf{gnd}_N(\pi))$*;*

(iv) $c \in \mathsf{XP}(\mathsf{gnd}(\pi))$ *iff* $c^* \in \mathsf{XP}(\mathsf{gnd}_N(\pi))$*.*

Proof. (i) By induction on the length of the derivation of c. For the base case, let $\pi = \bigwedge \forall \vec{x}_j c_j$. If $c \in \mathsf{EQ}$, clearly $c^* \in \mathsf{EQ}$. Otherwise, $c \in \mathsf{gnd}(\pi)$ iff $c = {c_j}^{\vec{x}_j}_{\vec{n}}$ for some j and $\vec{n}$ iff (since $*$ leaves the names in c_j unchanged) $c^* = {c_j}^{\vec{x}_j}_{\vec{n}^*}$ for some j and $\vec{n}$ iff $c^* \in \mathsf{gnd}(\pi)$ iff (since c^* contains names only from N) $c^* \in \mathsf{gnd}_N(\pi)$. For the induction step, $c \in \mathsf{UP}(\mathsf{gnd}(\pi)) \setminus (\mathsf{EQ} \cup \mathsf{gnd}(\pi))$ iff $c \cup [\ell], [\overline{\ell}] \in \mathsf{UP}(\mathsf{gnd}(\pi))$ for some ℓ iff (by induction) $c^{\star} \cup [\ell^{\star}], [\overline{\ell^{\star}}] \in \mathsf{UP}(\mathsf{gnd}_N(\pi))$ where $\star$ swaps the names $n'_{l+1}, \ldots, n'_{l'} \notin N$, $l' \leq v$ that occur in ℓ but not in c with $n_{l+1}, \ldots, n_{l'}$ and is like $*$ otherwise iff $c^{\star} = c^{*} \in \mathsf{UP}(\mathsf{gnd}_N(\pi)) \setminus (\mathsf{EQ} \cup \mathsf{gnd}_N(\pi))$.

(ii) $c \in \mathsf{UP}^{+}(\mathsf{gnd}(\pi))$ iff $c' \in \mathsf{UP}(\mathsf{gnd}(\pi))$ for some $c' \subseteq c$ iff (by (i)) $c'^{*} \in \mathsf{UP}(\mathsf{gnd}_N(\pi))$ for some $c' \subseteq c$ iff $c' \in \mathsf{UP}(\mathsf{gnd}_N(\pi))$ for some $c' \subseteq c^{*}$ iff $c^{*} \in \mathsf{UP}^{+}(\mathsf{gnd}_N(\pi))$.

(iii) $c \in \mathsf{UP}^{-}(\mathsf{gnd}(\pi))$ iff $c \in \mathsf{UP}(\mathsf{gnd}(\pi))$ and for all $c' \subsetneq c$, $c' \notin \mathsf{UP}(\mathsf{gnd}(\pi))$ iff (by (i)) $c^{*} \in \mathsf{UP}(\mathsf{gnd}_N(\pi))$ and for all $c' \subsetneq c$, $c'^{*} \notin \mathsf{UP}(\mathsf{gnd}_N(\pi))$ iff $c^{*} \in \mathsf{UP}(\mathsf{gnd}_N(\pi))$ and for all $c' \subsetneq c^{*}$, $c' \notin \mathsf{UP}(\mathsf{gnd}_N(\pi))$ iff $c^{*} \in \mathsf{UP}^{-}(\mathsf{gnd}_N(\pi))$.

(iv) $\mathsf{XP}(\mathsf{gnd}(\pi))$ only contains unit clauses and perhaps the empty clause. For unit clauses, $[\ell] \in \mathsf{XP}(\mathsf{gnd}(\pi))$ iff $[\ell] \in \mathsf{EQ}$ or for some c', $c' \cup [\ell] \in \mathsf{UP}^{-}(\mathsf{gnd}(\pi)) \setminus \mathsf{EQ}$ iff (by Lemma C.1.15 and (iii)) $[\ell^{\star}] \in \mathsf{EQ}$ or for some c', $c'^{\star} \cup [\ell^{\star}] \in \mathsf{UP}^{-}(\mathsf{gnd}_N(\pi)) \setminus \mathsf{EQ}$ where $\star$ swaps the names $n'_{l+1}, \ldots, n'_{l'} \notin N$, $l' \leq v$ that occur in c' but not in ℓ with $n_{l+1}, \ldots, n_{l'}$ and is like $*$ otherwise iff $[\ell^{\star}] = [\ell^{*}] \in \mathsf{XP}(\mathsf{gnd}_N(\pi))$. For the empty clause, $[] \in \mathsf{XP}(\mathsf{gnd}(\pi))$ iff $[\ell] \in \mathsf{XP}(\mathsf{gnd}(\pi))$ and $[\overline{\ell}] \in \mathsf{XP}(\mathsf{gnd}(\pi))$ for some ℓ iff (since $|\ell|_{\mathrm{w}} \leq v$) $[\ell^{\star}] \in \mathsf{XP}(\mathsf{gnd}_N(\pi))$ and $[\overline{\ell^{\star}}] \in \mathsf{XP}(\mathsf{gnd}_N(\pi))$ where $\star$ swaps the names $n'_1, \ldots, n'_{l'} \notin N$, $l' \leq v$ that occur in ℓ with $n_1, \ldots, n_{l'}$ iff $[] \in \mathsf{XP}(\mathsf{gnd}_N(\pi))$. $\square$

Putting things together: the correctness theorems

Now we put the lemmas together and give two decision procedures for $\mathrel{|\!\approx^{\circ}}$ and $\mathrel{|\!\approx^{\bullet}}$ with respect to proper^{+} knowledge bases.

Theorem 6.8.7 $\mathsf{gnd}(\pi), k \mathrel{|\!\approx^{\circ}} \phi$ *iff* $\mathsf{S}[N, \mathsf{gnd}_N(\pi), k, \phi] = 1$, *where N contains the names from π and ϕ plus $k \cdot v + v$ names for $v \geq |\pi|_{\mathrm{w}}$ and $v \geq |\phi|_{\mathrm{w}}$.*

Proof. By induction on k. For the base case let $k = 0$. Let $n_1, \ldots, n_v \in N$ not occur in π or ϕ. We show by subinduction on the length of ϕ that $\mathsf{gnd}(\pi), 0 \mathrel{|\!\approx^{\circ}} \phi$ iff $\mathsf{S}[N, \mathsf{gnd}_N(\pi), 0, \phi^{*}] = 1$ where $*$ is an arbitrary bijection that swaps the $l \leq v$ names $n'_1, \ldots, n'_l \notin N$ occurring in ϕ with $n_1, \ldots, n_l$ and is the identity otherwise; since $*$ is the identity for all names in the original ϕ, the lemma for $k = 0$ follows. For a clause, $\mathsf{gnd}(\pi), 0 \mathrel{|\!\approx^{\circ}} c$ iff $c \in \mathsf{UP}^{+}(\mathsf{gnd}(\pi))$ iff (by Lemma C.1.16 and since c mentions at most v names that do not occur in π or ϕ) $c^{*} \in \mathsf{UP}^{+}(\mathsf{gnd}_N(\pi))$ where $*$ is the bijection that swaps the names $n'_1, \ldots, n'_l \notin N$ occurring in c with $n_1, \ldots, n_l$ and is the identity otherwise iff $\mathsf{S}[N, \mathsf{gnd}_N(\pi), 0, c^{*}] = 1$. For a non-clausal disjunction, $\mathsf{gnd}(\pi), 0 \mathrel{|\!\approx^{\circ}} (\phi \vee \psi)$ iff $\mathsf{gnd}(\pi), 0 \mathrel{|\!\approx^{\circ}} \phi$ or $\mathsf{gnd}(\pi), 0 \mathrel{|\!\approx^{\circ}} \psi$ iff (by subinduction) $\mathsf{S}[N, \mathsf{gnd}_N(\pi), 0, \phi^{\star}] = 1$ or $\mathsf{S}[N, \mathsf{gnd}_N(\pi), 0, \psi^{\dagger}] = 1$ where without loss of generality (since ϕ and ψ together mention at most v names not in N) $n^{\star} \neq n^{\dagger}$ implies $n^{\star} = n$ or $n^{\dagger} = n$ iff $\mathsf{S}[N, \mathsf{gnd}_N(\pi), 0,$

$(\phi \vee \psi)^*] = 1$ where $n^* = n^\star$ if $n^\star \neq n$, and $n^* = n^\dagger$ otherwise. The case for negated disjunctions is analogous, and the case for double negations is trivial. For an existential, $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} \exists x\, \phi$ iff $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} \phi^x_n$ for some name n iff (by Lemma C.1.14 and since ϕ mentions less than v of $n_1, \ldots, n_v$) $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} \phi^x_n$ for some $n \in N$ iff (by subinduction) $\mathsf{S}[N, \mathsf{gnd}_N(\pi), 0, (\phi^x_n)^\star] = 1$ for some $n \in N$ where $\star$ is the bijection that swaps $n'_1, \ldots, n'_l \notin N$ occurring in ϕ^x_n with $n_1, \ldots, n_l$ iff $\mathsf{S}[N, \mathsf{gnd}_N(\pi), 0, (\exists x\, \phi)^*] = 1$ where $*$ is just like $\star$ except that for all n which n do not occur in ϕ, $n^* = n$ and $n^{\star *} = n^\star$. The case for negated existentials is analogous.

For the main induction step suppose the theorem holds for k. Then $\mathsf{gnd}(\pi), k+1 \mathrel{|\!\!\approx} \phi$ iff (by Lemma C.1.5) $\mathsf{gnd}(\pi) \uplus \ell, k \mathrel{|\!\!\approx} \phi$ and $\mathsf{gnd}(\pi) \uplus \overline{\ell}, k \mathrel{|\!\!\approx} \phi$ for some ℓ whose symbol occurs in π or ϕ and whose names occur in N iff $\mathsf{gnd}(\pi \wedge \ell), k \mathrel{|\!\!\approx} \phi$ and $\mathsf{gnd}(\pi \wedge \overline{\ell}), k \mathrel{|\!\!\approx} \phi$ for some ℓ whose symbol occurs in π or ϕ and whose names occur in N iff (by induction, which is applicable because ℓ mentions at most v names, and thus N still mentions $k \cdot v + v$ names not in $\pi \wedge \ell$ or ϕ) $\mathsf{S}[\mathsf{gnd}_N(\pi \wedge \ell), k, \phi] = \mathsf{S}[\mathsf{gnd}_N(\pi \wedge \overline{\ell}), k, \phi] = 1$ for some ℓ whose symbol occurs in π or ϕ and whose names occur in N iff $\mathsf{S}[\mathsf{gnd}_N(\pi) \uplus \ell, k, \phi] = \mathsf{S}[\mathsf{gnd}_N(\pi) \uplus \overline{\ell}, k, \phi] = 1$ for some ℓ whose symbol occurs in π or ϕ and whose names occur in N iff $\mathsf{S}[N, \mathsf{gnd}_N(\pi), k+1, \phi] = 1$. □

Lemma C.1.17 *Let* $f \in \{\mathsf{UP}, \mathsf{UP}^+, \mathsf{UP}^-\}$.

(i) $\mathsf{S}[N, s \cup s', l, \phi] = \mathsf{S}[N, f(s) \cup s', l, \phi]$;

(ii) $\mathsf{C}[N, s \cup s', l, \phi] = \mathsf{C}[N, f(s) \cup s', l, \phi]$.

Proof. Analogous to the proof of Lemma 6.8.3. □

Theorem 6.8.11 $\mathsf{gnd}(\pi), l \mathrel{|\!\!\approx} \phi$ *iff* $\mathsf{C}[N, \mathsf{gnd}_N(\pi), l, \phi] = 1$, *where N contains the names from π and ϕ plus $l \cdot v + v$ names for $v \geq |\pi|_\mathrm{w}$ and $v \geq |\phi|_\mathrm{w}$.*

Proof. By induction on l. For the base case let $l = 0$. Let $n_1, \ldots, n_v \in N$ not occur in π or ϕ. We show by subinduction on the length of ϕ that $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} \phi$ iff $\mathsf{C}[N, \mathsf{gnd}_N(\pi), 0, \phi^*] = 1$ where $*$ is an arbitrary bijection that swaps the $k \leq v$ names $n'_1, \ldots, n'_k \notin N$ occurring in ϕ with $n_1, \ldots, n_k$ and is the identity otherwise; since $*$ is the identity for all names in the original ϕ, the lemma for $l = 0$ follows. For a negated clause, $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} \neg c$ iff $[] \in \mathsf{XP}(\mathsf{gnd}(\pi))$ or $c \notin \mathsf{UP}^+(\mathsf{gnd}(\pi))$ iff (by Lemma C.1.16 and since c mentions at most v names that do not occur in π or ϕ) $[] \in \mathsf{XP}(\mathsf{gnd}_N(\pi))$ or $c^* \notin \mathsf{UP}^+(\mathsf{gnd}_N(\pi))$ where $*$ is the bijection that swaps the names $n'_1, \ldots, n'_k \notin N$ occurring in c with $n_1, \ldots, n_k$ and is the identity otherwise iff $\mathsf{C}[N, \mathsf{gnd}_N(\pi), 0, \neg c^*] = 1$. For a positive literal, $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} \ell$ iff $\mathsf{gnd}(\pi), 0 \mathrel{|\!\!\approx} \neg\overline{\ell}$ iff (by the case for negated clauses) $\mathsf{C}[N, \mathsf{gnd}_N(\pi), 0, \neg\overline{\ell^*}] = 1$ iff $\mathsf{C}[N, \mathsf{gnd}_N(\pi), 0, \ell^*] = 1$. The subinduction steps are analogous to those from the proof of Theorem 6.8.7.

For the main induction step suppose the theorem holds for l. Then $\mathsf{gnd}(\pi), l+1 \mathrel{|\!\!\approx} \phi$ iff (by Lemma C.1.13) $\mathsf{gnd}(\pi) \otimes \ell, l \mathrel{|\!\!\approx} \phi$ for all ℓ whose symbols occur in π or ϕ and whose names occur in N iff (by Lemmas C.1.9 and 6.8.3) $\mathsf{gnd}(\pi \otimes \ell), l \mathrel{|\!\!\approx} \phi$ for all ℓ whose symbol occurs in π or ϕ and whose names occur in N iff (by induction, which is applicable because ℓ mentions at most v names, and since by Lemma C.1.7 N still mentions $l \cdot v + v$ names not in $\pi \otimes \ell$ or ϕ) $\mathsf{C}[\mathsf{gnd}_N(\pi \otimes \ell), l, \phi] = 1$ for all ℓ whose symbols occur in π or ϕ and whose names occur in N iff (by Lemmas C.1.9 and C.1.17) $\mathsf{C}[\mathsf{gnd}_N(\pi) \otimes \ell, l, \phi] = 1$ for all ℓ whose symbols occur in π or ϕ and whose names occur in N iff $\mathsf{C}[N, \mathsf{gnd}_N(\pi), l+1, \phi] = 1$. $\square$

Complexity analysis

Finally we can prove Theorems 6.8.8 and 6.8.12.

Lemma C.1.18 $\mathsf{C}[N, \mathsf{gnd}_N(\pi), l, \phi]$ *can be computed in time* $O(2^k \cdot (|\pi| + k)^{k+1} \cdot |\phi|^{k+1} \cdot |N|^{(|\pi|_{\mathrm{w}}+|\phi|_{\mathrm{w}})\cdot(k+1)})$.

Proof. Let $f(k)$ denote the complexity of computing $\mathsf{S}[N, \mathsf{gnd}_N(\pi), k, \phi]$.

Splitting creates a tree of height k, whose leaves are setups that consist of $\mathsf{gnd}_N(\pi)$ plus k more unit clauses. Closing such a setup under unit propagation can be done in time linear in the size of the setup $\le |\pi| \cdot |N|^{|\pi|_{\mathrm{w}}} + k$ (Zhang and Stickel 1996). The number of clauses to be checked for subsumption is $\le |\phi| \cdot |N|^{|\phi|_{\mathrm{w}}}$. Hence $f(0) \in O((|\pi| + k) \cdot |\phi| \cdot |N|^{|\pi|_{\mathrm{w}}+|\phi|_{\mathrm{w}}})$.

There are $\le |\pi|$ predicate symbols in π and their maximum arity is $\le |\pi|_{\mathrm{w}}$. Analogously for ϕ. Hence the number of relevant split literals is $\le |\pi| \cdot |N|^{|\pi|_{\mathrm{w}}} + |\phi| \cdot |N|^{|\phi|_{\mathrm{w}}}$. Thus $f(k+1) \in O(2 \cdot (|\pi| \cdot |N|^{|\pi|_{\mathrm{w}}} + |\phi| \cdot |N|^{|\phi|_{\mathrm{w}}}) \cdot f(k))$. Solving the recurrence obtains $f(k) \in O(2^k \cdot (|\pi| \cdot |N|^{|\pi|_{\mathrm{w}}} + |\phi| \cdot |N|^{|\phi|_{\mathrm{w}}})^k \cdot f(0))$.

Thus $f(k) \in O(2^k \cdot (|\pi| + k)^{k+1} \cdot |\phi|^{k+1} \cdot |N|^{(|\pi|_{\mathrm{w}}+|\phi|_{\mathrm{w}})\cdot(k+1)})$. $\square$

Theorem 6.8.8 $\mathsf{gnd}(\pi), k \mathrel{|\!\!\approx} \phi$ *can be decided in time* $O((|\pi| + k)^{k+1} \cdot |\phi|^{k+1} \cdot (\max\{|\pi|_{\mathrm{w}}, |\phi|_{\mathrm{w}}\} \cdot (|\pi| + |\phi| + k + 1))^{(|\pi|_{\mathrm{w}}+|\phi|_{\mathrm{w}})\cdot(k+1)} \cdot 2^k)$.

Proof. Let N be as in Lemma C.1.18. We can estimate $|N| \le \max\{|\pi|_{\mathrm{w}}, |\phi|_{\mathrm{w}}\} \cdot (|\pi| + |\phi| + k + 1)$. By Lemma C.1.18 and Theorem 6.8.7, the theorem follows. $\square$

Lemma C.1.19 $\mathsf{C}[N, \mathsf{gnd}_N(\pi), l, \phi]$ *can be computed in time* $O((|\pi| + l)^{l+1} \cdot |\phi|^{l+1} \cdot (\max\{|\pi|_{\mathrm{w}}, |\phi|_{\mathrm{w}}\} + |N|)^{(\max\{|\pi|_{\mathrm{w}},|\phi|_{\mathrm{w}}\}+|\phi|_{\mathrm{w}})\cdot(l+1)})$.

Proof. Let $f(k)$ denote the complexity of computing $\mathsf{C}[N, \mathsf{gnd}_N(\pi), l, \phi]$.

Unlike splitting, the $\otimes_N$ operation possibly adds more than one unit clause. Hence, at level 0 up to $l \cdot |N|^{\max\{|\pi|_{\mathrm{w}},|\phi|_{\mathrm{w}}\}}$ may have been added to $\mathsf{gnd}_N(\pi)$. Closing the resulting setup under unit propagation can be done in time linear in the size of the setup $\leq (|\pi| + l) \cdot |N|^{\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\}}$ (Zhang and Stickel 1996). The number of clauses to be checked for subsumption is $\leq |\phi| \cdot |N|^{|\phi|_\mathrm{w}}$. Hence $f(0) \in O((|\pi| + l) \cdot |\phi| \cdot |N|^{\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\} + |\phi|_\mathrm{w}})$.

There are $\leq |\pi|$ predicate symbols in π and their maximum arity is $\leq |\pi|_\mathrm{w}$. Analogously for ϕ. Hence the number of relevant add literals is $\leq |\pi| \cdot (|\pi|_\mathrm{w} + |N|)^{|\pi|_\mathrm{w}} + |\phi| \cdot (|\phi|_\mathrm{w} + |N|)^{|\phi|_\mathrm{w}}$. Thus $f(l+1) \in O((|\pi| + |\phi|) \cdot (\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\} + |N|)^{\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\}} \cdot f(l))$. Solving the recurrence yields $O((|\pi| + |\phi|)^l \cdot (\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\} + |N|)^{\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\} \cdot l} \cdot f(0))$.

Thus $f(l) \in O((|\pi| + l)^{l+1} \cdot |\phi|^{l+1} \cdot (\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\} + |N|)^{(\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\} + |\phi|_\mathrm{w}) \cdot (l+1)})$. □

Theorem 6.8.12 $\mathsf{gnd}(\pi), l \mathrel{|\!\approx^{\bullet}} \phi$ *can be decided in time* $O((|\pi| + l)^{l+1} \cdot |\phi|^{l+1} \cdot (\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\} \cdot (|\pi| + |\phi| + l + 2))^{(\max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\} + |\phi|_\mathrm{w}) \cdot (l+1)})$.

Proof. Let N be as in Lemma C.1.19. We can estimate $|N| \leq \max\{|\pi|_\mathrm{w}, |\phi|_\mathrm{w}\} \cdot (|\pi| + |\phi| + l + 1)$. By Lemma C.1.19 and Theorem 6.8.11, the theorem follows. □

C.2 Proof of the normal form

Here we show Theorem 6.9.3, which states that the normal form behaves well with regard to $\mathrel{|\!\approx^{\circ}}$ and $\mathrel{|\!\approx^{\bullet}}$ in the sense converting a formula to normal form retains all proofs (disproofs) in $\mathrel{|\!\approx^{\circ}}$ ($\mathrel{|\!\approx^{\bullet}}$).

Lemma C.2.1 *If ϕ' is the result renaming the variables in ϕ so that no variable bounded by different quantified, then*

(i) $\models \phi \equiv \phi'$*;*

(ii) $s, k \mathrel{|\!\approx^{\circ}} \phi$ *iff* $s, k \mathrel{|\!\approx^{\circ}} \phi'$*;*

(iii) $s, l \mathrel{|\!\approx^{\bullet}} \phi$ *iff* $s, l \mathrel{|\!\approx^{\bullet}} \phi'$*.*

Proof. Follows by trivial inductions on ϕ. □

Lemma C.2.2

(i) If $s, k \mathrel{|\!\approx^{\circ}} \phi$ *or* $s, k \mathrel{|\!\approx^{\circ}} \psi$*, then* $s, k \mathrel{|\!\approx^{\circ}} (\phi \vee \psi)$*.*

(ii) If $s, l \mathrel{|\!\approx^{\bullet}} \neg\phi$ *or* $s, l \mathrel{|\!\approx^{\bullet}} \neg\psi$*, then* $s, l \mathrel{|\!\approx^{\bullet}} \neg(\phi \vee \psi)$*.*

Proof. (i) By induction on k. Let $s, 0 \mathrel{|\!\approx^{\circ}} \phi$ or $s, 0 \mathrel{|\!\approx^{\circ}} \psi$. If $(\phi \vee \psi)$ is not a clause, by Rule $\mathcal{L}^{\circ}3$ $s, 0 \mathrel{|\!\approx^{\circ}} (\phi \vee \psi)$. Otherwise, if c, c' are the two clauses, $c \in \mathsf{UP}^+(s)$ or $c' \in \mathsf{UP}^+(s)$, so $c \cup c' \in \mathsf{UP}^+(s)$, and thus $s, 0 \mathrel{|\!\approx^{\circ}} c \cup c'$, which is $(\phi \vee \psi)$. For the induction step, $s, k+1 \mathrel{|\!\approx^{\circ}} \phi$ or $s, k+1 \mathrel{|\!\approx^{\circ}} \psi$ iff $s \uplus \ell, k \mathrel{|\!\approx^{\circ}} \phi$ and $s \uplus \overline{\ell}, k \mathrel{|\!\approx^{\circ}} \phi$, or $s \uplus \ell, k \mathrel{|\!\approx^{\circ}} \psi$

and $s \uplus \overline{\ell}, k \mathrel{|\!\overset{\circ}{\approx}} \psi$ only if (by induction) $s \uplus \ell, k \mathrel{|\!\overset{\circ}{\approx}} (\phi \vee \psi)$ or $s \uplus \overline{\ell}, k \mathrel{|\!\overset{\circ}{\approx}} (\phi \vee \psi)$ iff $s, k+1 \mathrel{|\!\overset{\circ}{\approx}} (\phi \vee \psi)$.

(ii) By induction on l analogous to (i). If $(\phi \vee \psi)$ is not a clause, by Rule $\mathcal{L}^{\bullet}4$ $s, 0 \mathrel{|\!\overset{\bullet}{\approx}} \neg(\phi \vee \psi)$. Otherwise, if c, c' are the two clauses, either $[] \notin \mathsf{XP}(s)$ and $c \in \mathsf{UP}^{+}(s)$ or $[] \notin \mathsf{XP}(s)$ and $c' \in \mathsf{UP}^{+}(s)$, so $[] \notin \mathsf{XP}(s)$ and $c \cup c' \in \mathsf{UP}^{+}(s)$, and thus $s, 0 \mathrel{|\!\overset{\bullet}{\approx}} \neg(c \cup c')$, which is $\neg(\phi \vee \psi)$. For the induction step, $s, l+1 \mathrel{|\!\overset{\bullet}{\approx}} \neg\phi$ or $s, l+1 \mathrel{|\!\overset{\bullet}{\approx}} \neg\psi$ iff $s \otimes \ell, l \mathrel{|\!\overset{\bullet}{\approx}} \neg\phi$ for some ℓ or $s \otimes \ell, l \mathrel{|\!\overset{\bullet}{\approx}} \neg\psi$ for some ℓ' only if (by induction) $s \otimes \ell, l \mathrel{|\!\overset{\bullet}{\approx}} \neg(\phi \vee \psi)$ for some ℓ iff $s, l+1 \mathrel{|\!\overset{\bullet}{\approx}} \neg(\phi \vee \psi)$. □

Lemma C.2.3 *Let all variables in $\vec{x}_1$, $\vec{x}_2$ be distinct and $\S\vec{x}_i$ be a word over $\{\neg, \exists x_{i1}, \exists x_{i2}, \ldots\}$ with an even number of $\neg$.*

(i) If $s, 0 \mathrel{|\!\overset{\circ}{\approx}} (\S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$, then $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \S\vec{x}_1 \S\vec{x}_2 (c_1 \vee c_2)$.

(ii) If $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg(\S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$, then $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\S\vec{x}_1 \S\vec{x}_2 (c_1 \vee c_2)$.

(iii) If $s, 0 \mathrel{|\!\overset{\bullet}{\approx}} (\S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$, then $s, 0 \mathrel{|\!\overset{\bullet}{\approx}} \S\vec{x}_1 \S\vec{x}_2 (c_1 \vee c_2)$.

(iv) If $s, 0 \mathrel{|\!\overset{\bullet}{\approx}} \neg(\S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$, then $s, 0 \mathrel{|\!\overset{\bullet}{\approx}} \neg\S\vec{x}_1 \S\vec{x}_2 (c_1 \vee c_2)$.

Proof. (i) By induction on the length of $\S\vec{x}_1$. For the base case, we do a subinduction on the length of $\S\vec{x}_2$. The base case of the subinduction is trivial. For the subinduction step for k existentials between two negations, we need subsubinduction on k. The base case is simply $s, 0 \mathrel{|\!\overset{\circ}{\approx}} (c_1 \vee \neg\neg\S\vec{x}_2 c_2)$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} c_1$ or $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\neg\S\vec{x}_2 c_2$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} c_1$ or $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \S\vec{x}_2 c_2$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} (c_1 \vee \S\vec{x}_2 c_2)$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\neg(c_1 \vee \S\vec{x}_2 c_2)$. For the subsubinduction, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} (c_1 \vee \neg\exists x_1 \ldots \exists x_k \neg\S\vec{x}_2 c_2)$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} c_1$ or $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\exists x_1 \ldots \exists x_k \neg\S\vec{x}_2 c_2$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} c_1$ or $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\exists x_2 \ldots \exists x_k \neg\S\vec{x}_2 {c_2}_n^{x_1}$ for all n iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} (c_1 \vee \neg\exists x_2 \ldots \exists x_k \neg\S\vec{x}_2 {c_2}_n^{x_1})$ for all n only if (by subinduction) $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\exists x_2 \ldots \exists x_k \neg\S\vec{x}_2 (c_1 \vee {c_2}_n^{x_1})$ for all n iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\exists x_1 \ldots \exists x_k \neg\S\vec{x}_2 (c_1 \vee c_2)$. We skip the subinduction step for existentials not wrapped by negations, as it is very similar but does not require the subsubinduction. For the induction step for existentials not wrapped by negations, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} (\exists x \S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \exists x \S\vec{x}_1 c_1$ or $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \S\vec{x}_2 c_2$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \S\vec{x}_1 {c_1}_n^{x}$ or $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \S\vec{x}_2 c_2$ for some n iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} (\S\vec{x}_1 {c_1}_n^{x} \vee \S\vec{x}_2 c_2)$ for some n only if (by induction) $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \S\vec{x}_1 \S\vec{x}_2 ({c_1}_n^{x} \vee {c_2}_n^{x})$ for some n iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \exists x \S\vec{x}_1 \S\vec{x}_2 (c_1 \vee c_2)$. We skip the induction step for existentials between two negations; it is very similar to the base case we have shown above and requires the same subsubinduction scheme.

(ii) By induction very similar to (i). For the base case, we do a subinduction on the length of $\S\vec{x}_2$. The base case of the subinduction is trivial. For the subinduction step for existentials not wrapped by negations, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg(c_1 \vee \exists x \S\vec{x}_2 c_2)$ iff

$s,0 \mathrel{|\!\approx} \neg c_1$ and $s,0 \mathrel{|\!\approx} \neg\exists x \S\vec{x}_2 c_2$ iff $s,0 \mathrel{|\!\approx} \neg c_1$ and $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_2 {c_2}^x_n$ for all n iff $s,0 \mathrel{|\!\approx} \neg(c_1 \vee \S\vec{x}_2 {c_2}^x_n)$ for all n only if (by subinduction) $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_2(c_1 \vee {c_2}^x_n)$ for all n iff $s,0 \mathrel{|\!\approx} \neg\exists x\S\vec{x}_2(c_1 \vee c_2)$. We skip the subinduction step for existentials between two negations; like the subinduction from (i) it requires a subsubinduction on the number of existentials. For the induction step for k existentials between two negations we again need a subinduction on k. The base case is simply $s,0 \mathrel{|\!\approx} \neg(\neg\neg\S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$ iff $s,0 \mathrel{|\!\approx} \neg\neg\neg\S\vec{x}_1 c_1$ and $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_2 c_2$ iff $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_1 c_1$ and $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_2 c_2$ iff $s,0 \mathrel{|\!\approx} \neg(\S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$ only if (by induction) $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_1\S\vec{x}_2(c_1 \vee c_2)$ iff $s,0 \mathrel{|\!\approx} \neg\neg\neg\S\vec{x}_1(c_1 \vee \S\vec{x}_2 c_2)$. For the subinduction, $s,0 \mathrel{|\!\approx} \neg(\neg\exists x_1 \dots \exists x_k \neg\S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$ iff $s,0 \mathrel{|\!\approx} \neg\neg\exists x_1 \dots \exists x_k \neg\S\vec{x}_1 c_1$ and $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_2 c_2$ iff $s,0 \mathrel{|\!\approx} \exists x_2 \dots \exists x_k \neg\S\vec{x}_1 {c_1}^{x_1}_n$ and $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_2 c_2$ for some n iff $s,0 \mathrel{|\!\approx} \neg\neg\exists x_2 \dots \exists x_k \neg\S\vec{x}_1 {c_1}^{x_1}_n$ and $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_2 c_2$ for some n iff $s,0 \mathrel{|\!\approx} \neg(\neg\exists x_2 \dots \exists x_k \neg\S\vec{x}_1 {c_1}^{x_1}_n \vee \S\vec{x}_2 c_2)$ for some n only if (by induction) $s,0 \mathrel{|\!\approx} \neg\neg\exists x_2 \dots \exists x_k \neg\S\vec{x}_1\S\vec{x}_2({c_1}^{x_1}_n \vee c_2)$ for some n iff $s,0 \mathrel{|\!\approx} \exists x_1 \dots \exists x_k \neg\S\vec{x}_1\S\vec{x}_2(c_1 \vee c_2)$ iff $s,0 \mathrel{|\!\approx} \neg\neg\exists x_1 \dots \exists x_k \neg\S\vec{x}_1(c_1 \vee \S\vec{x}_2 c_2)$.

(iii) Analogous to (i).

(iv) Analogous to (ii). □

Lemma C.2.4 *Let all variables in $\vec{x}_1$, $\vec{x}_2$ be distinct.*

(i) If $s,0 \mathrel{|\!\approx} \phi$, then $s,0 \mathrel{|\!\approx} \mathsf{NF}[\phi]$.

(ii) If $s,0 \not\mathrel{|\!\approx} \phi$, then $s,0 \not\mathrel{|\!\approx} \mathsf{NF}[\phi]$.

Proof. By induction on the length of ϕ. For the induction steps we additionally need $s,0 \mathrel{|\!\approx} \neg\phi$ implies $s,0 \mathrel{|\!\approx} \neg\mathsf{NF}[\phi]$, which is also shown below. For a clause, $s,0 \mathrel{|\!\approx} c$ iff (since $c = \mathsf{NF}[c]$) $s,0 \mathrel{|\!\approx} \mathsf{NF}[c]$.

Now consider a non-clausal disjunction $(\phi_1 \vee \phi_2)$. When $\mathsf{NF}[\phi_i] = \S\vec{x}_i c_i$, then $s,0 \mathrel{|\!\approx} (\phi_1 \vee \phi_2)$ iff $s,0 \mathrel{|\!\approx} \phi_1$ or $s,0 \mathrel{|\!\approx} \phi_2$ only if (by induction) $s,0 \mathrel{|\!\approx} \mathsf{NF}[\phi_1]$ or $s,0 \mathrel{|\!\approx} \phi_2$ iff $s,0 \mathrel{|\!\approx} \S\vec{x}_1 c_1$ or $s,0 \mathrel{|\!\approx} \S\vec{x}_2 c_2$ only if (by Lemma C.2.2) $s,0 \mathrel{|\!\approx} (\S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$ only if (by Lemma C.2.3) $s,0 \mathrel{|\!\approx} \S\vec{x}_1\S\vec{x}_2(c_1 \vee c_2)$ iff $s,0 \mathrel{|\!\approx} \mathsf{NF}[(\phi_1 \vee \phi_2)]$. The case for $\mathsf{NF}[\phi_i] = \S'\vec{x}_i\, a_i$ for an atom a_i is shown analogously by first showing (by induction on the length of $\S'\vec{x}$) that $s,0 \mathrel{|\!\approx} \S'\vec{x}_i\, a_i$ iff $s,0 \mathrel{|\!\approx} \S\vec{x}_i \neg a_i$, so that the rest of the argument of the previous applies here, too. Otherwise, if $\mathsf{NF}[\phi_i] \neq \S\vec{x}_i c_i$ for some i, $s,0 \mathrel{|\!\approx} (\phi_1 \vee \phi_2)$ iff $s,0 \mathrel{|\!\approx} \phi_1$ or $s,0 \mathrel{|\!\approx} \phi_2$ only if (by induction) $s,0 \mathrel{|\!\approx} \mathsf{NF}[\phi_1]$ or $s,0 \mathrel{|\!\approx} \mathsf{NF}[\phi_2]$ only if (by Lemma C.2.2) $s,0 \mathrel{|\!\approx} (\mathsf{NF}[\phi_1] \vee \mathsf{NF}[\phi_2])$ iff $s,0 \mathrel{|\!\approx} \mathsf{NF}[(\phi_1 \vee \phi_2)]$.

Now consider a negated disjunction $\neg(\phi_1 \vee \phi_2)$. When $\mathsf{NF}[\phi_i] = \S\vec{x}_i c_i$, then $s,0 \mathrel{|\!\approx} \neg(\phi_1 \vee \phi_2)$ iff $s,0 \mathrel{|\!\approx} \neg\phi_1$ and $s,0 \mathrel{|\!\approx} \neg\phi_2$ only if (by induction) $s,0 \mathrel{|\!\approx} \neg\mathsf{NF}[\phi_1]$ and $s,0 \mathrel{|\!\approx} \neg\phi_2$ iff $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_1 c_1$ and $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_2 c_2$ iff $s,0 \mathrel{|\!\approx} \neg(\S\vec{x}_1 c_1 \vee \S\vec{x}_2 c_2)$ only if (by Lemma C.2.3) $s,0 \mathrel{|\!\approx} \neg\S\vec{x}_1\S\vec{x}_2(c_1 \vee c_2)$ iff $s,0 \mathrel{|\!\approx} \neg\mathsf{NF}[(\phi_1 \vee \phi_2)]$ iff $s,0 \mathrel{|\!\approx} \mathsf{NF}[\neg(\phi_1 \vee \phi_2)]$.

The case for $\mathsf{NF}[\phi_i] = \S'\vec{x}_i\, a_i$ for an atom a_i is shown analogously by first showing (by induction on the length of $\S'\,\vec{x}$) that $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\S'\vec{x}_i\, a_i$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\S\vec{x}_i\, \neg a_i$, so that the rest of the argument of the previous applies here, too. Otherwise, if $\mathsf{NF}[\phi_i] \neq \S\vec{x}_i\, c_i$ for some i, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg(\phi_1 \vee \phi_2)$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\phi_1$ and $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\phi_2$ only if (by induction) $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[\phi_1]$ and $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[\phi_2]$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg(\mathsf{NF}[\phi_1] \vee \mathsf{NF}[\phi_2])$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\neg(\phi_1 \vee \phi_2)]$.

For a double negation, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\neg\phi$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \phi$ iff (by induction) $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\phi]$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\neg\neg\phi]$.

For an existential, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \exists x\, \phi$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \phi^x_n$ for some n iff (by induction) $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\phi^x_n]$ for some n iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \exists x\, \mathsf{NF}[\phi]$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\exists x\, \phi]$.

For a negated existential, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\exists x\, \phi$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\phi^x_n$ for all n iff (by induction) $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[\phi^x_n]$ for all n iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\exists x\, \mathsf{NF}[\phi]$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[\exists x\, \phi]$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\neg\exists x\, \phi]$. This also proves that $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\exists x\, \phi$ implies $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[\exists x\, \phi]$.

Now we complete the induction by doing the steps to show that $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\phi$ implies $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[\phi]$. For a clause, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg c$ iff (since $c = \mathsf{NF}[c]$) $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[c]$. The cases for a disjunction has been covered by the case for a negated disjunction above already. For a negated disjunction, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\neg(\phi \vee \psi)$ iff (by induction) $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[(\phi \vee \psi)]$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\neg\mathsf{NF}[(\phi \vee \psi)]$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[\neg(\phi \vee \psi)]$. For a double negation, $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\neg\neg\phi$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\phi$ iff (by induction) $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[\phi]$ iff $s, 0 \mathrel{|\!\overset{\circ}{\approx}} \neg\mathsf{NF}[\neg\neg\phi]$. The cases for an existential has been covered by the case for a negated existential above already.

(ii) Analogous to (i), except that the second to last step for the negated (non-clausal) disjunction is merely an implication (instead of equivalence) justified by Lemma C.2.2 (instead of Rule $\mathcal{L}^{\circ}4$), and similarly the second to last step for non-clausal disjunctions is an equivalence (instead of an implication) justified by Rule $\mathcal{L}^{\bullet}3$ (instead of Lemma C.2.2). □

Theorem 6.9.3

(i) If $s, k \mathrel{|\!\overset{\circ}{\approx}} \phi$*, then* $s, k \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\phi]$.

(ii) If $s, l \mathrel{|\!\overset{\bullet}{\approx}} \phi$*, then* $s, l \mathrel{|\!\overset{\bullet}{\approx}} \mathsf{NF}[\phi]$.

Proof. By Lemma C.2.1 we can assume the ϕ to be preprocessed according to Definition 6.9.1.

(i) By induction on k. The base case follows from Lemma C.2.4. For the induction step, $s, k+1 \mathrel{|\!\overset{\circ}{\approx}} \phi$ iff $s \uplus \ell, k \mathrel{|\!\overset{\circ}{\approx}} \phi$ and $s \uplus \overline{\ell}, k \mathrel{|\!\overset{\circ}{\approx}} \phi$ for some ℓ only if (by induction) $s \uplus \ell, k \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\phi]$ and $s \uplus \overline{\ell}, k \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\phi]$ for some ℓ iff $s, k+1 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\phi]$.

(ii) By induction on l. The base case follows from Lemma C.2.4. For the induction step, $s, l+1 \mathrel{|\!\overset{\bullet}{\approx}} \phi$ iff $s \otimes \ell, l \mathrel{|\!\overset{\bullet}{\approx}} \phi$ for some ℓ only if (by induction) $s \otimes \ell, l \mathrel{|\!\overset{\bullet}{\approx}} \mathsf{NF}[\phi]$ for some ℓ iff $s, l+1 \mathrel{|\!\overset{\bullet}{\approx}} \mathsf{NF}[\phi]$. □

D Long Proofs for $\mathcal{BOL}$

D.1 Proof of the unique-model property

Here we prove the unique-model (modulo UP^+) property of $\mathcal{BOL}$ for proper$^+$ knowledge bases, Theorem 7.4.2. The idea follows the proof of the unique-model property for $\mathcal{BO}$, Theorem 4.5.3. For this section, let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be proper$^+$.

Lemma D.1.1

(i) $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \phi \rfloor \geq p$ *iff* $s_{p'}, k \mathrel{|\!\overset{\circ}{\approx}} \neg\phi$ *for all* $p' < p$*;*

(ii) $\lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \phi \rfloor \geq p$ *iff* $s_{p'}, l \mathrel{|\!\overset{\bullet}{\approx}} \neg\phi$ *for all* $p' < p$*.*

Proof. Follows from the definitions of $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \phi_i \rfloor$ and $\lfloor \vec{s}, l \mathrel{\overset{\bullet}{|}} \phi_i \rfloor$, respectively. □

Lemma D.1.2 *Let* $\vec{s} = \langle s_1, \ldots, s_{m+1} \rangle$ *be such that* $s_p = \mathsf{gnd}(\mathsf{NF}[\bigwedge_{i:\lfloor \vec{s},k \overset{\circ}{|} \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)])$ *for all* $p \in \mathbb{P}$*. Then* $\vec{s}$ *is a well-defined limited epistemic state, and* $\vec{s}|_k^l \mathrel{|\!\approx} \mathbf{O}_k^l \Gamma$.

Proof. We first show by induction on $p \in \mathbb{P}$ that $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \phi_i \rfloor \geq p$ and s_p are well-defined and that if $p > 1$, $\mathsf{UP}^+(s_p) \subseteq \mathsf{UP}^+(s_{p-1})$. Then $\vec{s}$ is a well-defined epistemic state. The base holds trivially.

For the induction step, suppose $\vec{s}$ has been constructed up to $p-1$. Then the expression $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \phi_i \rfloor \geq p$ is well-defined by Lemma D.1.1. Thus s_p is well-defined as well. By Lemma 6.5.2, $\mathsf{UP}^+(s_p) \subseteq \mathsf{UP}^+(s_{p-1})$ for $p > 1$. This completes the induction.

By Theorem 6.8.4, s_p is UP^+-minimal such that $s_p, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s},k \overset{\circ}{|} \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)]$ for every $p \in \mathbb{P}$. Thus by Rule $\mathcal{BOL}$6, $\vec{s}|_k^l \mathrel{|\!\approx} \mathbf{O}_k^l \Gamma$.

Finally we show that $\vec{s} = \langle s_1, \ldots, s_{m+1} \rangle$, that is, that at most the first $m+1$ setups differ (modulo UP^+). To see that, suppose the opposite. Then there is a "hole" in the plausibility ranking, that is, there is some p and i such that $p + 1 = \lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \phi_i \rfloor \neq \infty$, and $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \phi_j \rfloor \neq p$ and for all j. Since $\lfloor \vec{s}, k \mathrel{\overset{\circ}{|}} \phi_i \rfloor = p + 1$, $s_p, k \mathrel{|\!\overset{\circ}{\approx}} \neg\phi_i$, but $s_{p+1}, k \mathrel{\not|\!\overset{\circ}{\approx}} \neg\phi_i$. Then $I_p = I_{p+1}$, and hence $s_p = s_{p+1}$. Contradiction. □

Lemma D.1.3 *For all* p*, let* $s_p, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s},k \overset{\circ}{|} \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)]$ *be* UP^+*-minimal and* $s'_p, 0 \mathrel{|\!\overset{\circ}{\approx}} \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s}',k \overset{\circ}{|} \phi_i \rfloor \geq p} (\phi_i \supset \psi_i)]$ *be* UP^+*-minimal. Then* $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$.

Proof. We show by induction on p that $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$ and that $\lfloor \vec{s}, k \dagger \phi_i \rfloor > p$ iff $\lfloor \vec{s}', k \dagger \phi_i \rfloor > p$ for all i. For the base case consider $p = 1$. By Theorem 6.8.4, $\mathsf{UP}^+(s_1) = \mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_i(\phi_i \supset \psi_i)])) = \mathsf{UP}^+(s'_1)$, and by Lemma 6.8.3, $\lfloor \vec{s}, k \dagger \phi_i \rfloor > 1$ iff $\lfloor \vec{s}', k \dagger \phi_i \rfloor > 1$.

For the induction step suppose the statement holds for $p-1$. By induction, $\lfloor \vec{s} \dagger \phi_i \rfloor \geq p$ iff $\lfloor \vec{s}' \mid \phi_i \rfloor \geq p$ for all i (*). By Theorem 6.8.4,

$$\mathsf{UP}^+(s_p) = \mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\textstyle\bigwedge_{i:\lfloor \vec{s},k\dagger\phi_i \rfloor \geq p}(\phi_i \supset \psi_i)])) \text{ and}$$
$$\mathsf{UP}^+(s'_p) = \mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\textstyle\bigwedge_{i:\lfloor \vec{s}',k\dagger\phi_i \rfloor \geq p}(\phi_i \supset \psi_i)])).$$

By (*), $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$, and by Lemma 6.8.3, $\lfloor \vec{s}, k \dagger \phi_i \rfloor > p$ iff $\lfloor \vec{s}', k \dagger \phi_i \rfloor > p$. □

Lemma D.1.4 *Let $\vec{s} \mathrel{|\!\approx} \mathbf{O}^l_k \Gamma$ and $\vec{s}' \mathrel{|\!\approx} \mathbf{O}^l_k \Gamma$. Then $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$ for all p.*

Proof. Follows immediately from Lemma D.1.3. □

Theorem 7.4.2 *There is an $\vec{s} = \langle s_1, \ldots, s_{m+1} \rangle$ such that $\vec{s} \models \mathbf{O}^l_k \Gamma$, and for all $\vec{s}' \mathrel{|\!\approx} \mathbf{O}^l_k \Gamma$, $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$ for all p.*

Proof. Follows immediately from Lemmas D.1.2 and D.1.4. □

D.2 Proof of the monotonicity theorem

Here we prove Theorem 7.4.3, which claims that for proper$^+$ knowledge bases the effort in limited belief entailments is monotonic. For this section, again let $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be proper$^+$.

Lemma D.2.1 *Let $\vec{s} \mathrel{|\!\approx} \mathbf{B}^l_k(\phi \Rightarrow \psi)$.*

(i) $\vec{s} \mathrel{|\!\approx} \mathbf{B}^l_{k+1}(\phi \Rightarrow \psi)$*;*

(ii) $\vec{s} \mathrel{|\!\approx} \mathbf{B}^{l+1}_k(\phi \Rightarrow \psi)$*.*

Proof. By assumption, for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{s}, l \dagger \phi \rfloor$, then $s_p, k \mathrel{|\!\not\approx} (\phi \supset \psi)$. (i) holds since by Lemma 6.5.4, for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{s}, l \dagger \phi \rfloor$, then $s_p, k+1 \mathrel{|\!\not\approx} (\phi \supset \psi)$. (ii) holds since by Lemmas 7.3.3 and 6.5.3 and the concentricity of $\vec{s}$, for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{s}, l+1 \dagger \phi \rfloor$, then $s_p, k \mathrel{|\!\not\approx} (\phi \supset \psi)$. □

Lemma D.2.2 *Let $\mathbf{O}^l_k \mathrel{|\!\approx} \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$.*

(i) $\mathbf{O}^l_{k+1}\Gamma \mathrel{|\!\approx} \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$*;*

(ii) $\mathbf{O}^{l+1}_k\Gamma \mathrel{|\!\approx} \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$*.*

Proof. (i) Let $\vec{s}'' \mid\!\approx \mathbf{O}^{l}_{k+1}\Gamma$. Then by Rule $\mathcal{BOL}6$, $\vec{s}'' = \vec{s}\,|^{l}_{k+1}$ for some $\vec{s}$ such that for all p, s_p is UP^+-minimal such that $s_p, 0 \mid\!\not\approx \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s},k+1 \dagger \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)]$. By Lemma D.1.2 there is an $\vec{s}'''$ such that $\vec{s}''' \mid\!\approx \mathbf{O}^{l}_{k}\Gamma$. Then by Rule $\mathcal{BOL}6$, $\vec{s}''' = \vec{s}'\,|^{l}_{k}$ for some $\vec{s}'$ such that for all p, s'_p is UP^+-minimal such that $s'_p, 0 \mid\!\not\approx \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s}',k \dagger \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)]$. We first show by induction on p that, if $\vec{s}'$ is ${}^{l}_{k}$-bound-consistent at $1, \ldots, p$, then $\vec{s}$ is ${}^{l}_{k+1}$-bound-consistent at $1, \ldots, p$ and $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$, and otherwise $\mathsf{UP}^+(s_p) \supseteq \mathsf{UP}^+(s'_p)$. Afterwards we prove that if $\vec{s}'\,|^{l}_{k} \mid\!\approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$, which holds by assumption, then also $\vec{s}\,|^{l}_{k+1} \mid\!\approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$.

The base case is trivial. For the induction step suppose the claim holds for $p-1$. First suppose $\vec{s}'$ is ${}^{l}_{k}$-bound-consistent at $1, \ldots, p$. Then by induction and Lemmas 7.3.3, D.1.1, and 6.8.3 and ${}^{l}_{k}$-bound-consistency of $\vec{s}'$, $\lfloor \vec{s}, k+1 \dagger \phi_i \rfloor \geq p$ iff $\lfloor \vec{s}', k \dagger \phi_i \rfloor \geq p$. Moreover, by induction and Lemmas D.1.1 and 6.8.3, $\lfloor \vec{s}, l \ddagger \phi_i \rfloor \geq p$ iff $\lfloor \vec{s}', l \ddagger \phi_i \rfloor \geq p$. Thus and by induction, $\vec{s}$ is ${}^{l}_{k+1}$-bound-consistent at $1, \ldots, p$, and by Theorem 6.8.4, $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$. Now suppose $\vec{s}'$ is not ${}^{l}_{k}$-bound-consistent at some $p' \leq p$. Then by induction and Lemmas 7.3.3, D.1.1, and 6.8.3, $\lfloor \vec{s}, k+1 \dagger \phi_i \rfloor \geq p$ if $\lfloor \vec{s}', k \dagger \phi_i \rfloor \geq p$. Thus by Theorem 6.8.4, $\mathsf{UP}^+(s_p) \supseteq \mathsf{UP}^+(s'_p)$.

Now we show the entailment $\vec{s}\,|^{l}_{k+1} \mid\!\approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$. Let j be maximal such that $\vec{s}'$ is ${}^{l}_{k}$-bound-consistent at $1, \ldots, j$. Observe that by the above induction, $\mathsf{UP}^+((\vec{s}\,|^{l}_{k+1})_p) = \mathsf{UP}^+((\vec{s}'\,|^{l}_{k})_p)$ for all $p \leq j$ (*), and $\mathsf{UP}^+((\vec{s}\,|^{l}_{k+1})_p) \supseteq \mathsf{UP}^+((\vec{s}'\,|^{l}_{k})_{p'})$ for all $p > j$ and $p' > j$ (**).

Now by assumption, $\vec{s}'\,|^{l}_{k} \mid\!\approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$. Then for all p, if $p \leq \lfloor \vec{s}'\,|^{l}_{k}, l' \ddagger \phi \rfloor$, then $(\vec{s}'\,|^{l}_{k})_p, k' \mid\!\not\approx (\phi \supset \psi)$. If $\lfloor \vec{s}'\,|^{l}_{k}, l' \ddagger \phi \rfloor \leq j$, then by (*) and Lemma 6.8.3, for all p, if $p \leq \lfloor \vec{s}\,|^{l}_{k+1}, l' \ddagger \phi \rfloor$, then $(\vec{s}\,|^{l}_{k+1})_p, k' \mid\!\not\approx (\phi \supset \psi)$. Otherwise, by (**) and Lemma 6.5.3 and 6.8.3, for all p, $(\vec{s}\,|^{l}_{k+1})_p, k' \mid\!\not\approx (\phi \supset \psi)$. Thus in either case, $\vec{s}\,|^{l}_{k+1} \mid\!\approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$.

(ii) Let $\vec{s}'' \mid\!\approx \mathbf{O}^{l+1}_{k}\Gamma$. Then by Rule $\mathcal{BOL}6$, $\vec{s}'' = \vec{s}\,|^{l+1}_{k}$ for some $\vec{s}$ such that for all p, s_p is UP^+-minimal such that $s_p, 0 \mid\!\not\approx \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s},k \dagger \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)]$. By Lemma D.1.2 there is an $\vec{s}'''$ such that $\vec{s}''' \mid\!\approx \mathbf{O}^{l}_{k}\Gamma$. Then by Rule $\mathcal{BOL}6$, $\vec{s}''' = \vec{s}'\,|^{l}_{k}$ for some $\vec{s}'$ such that for all p, s'_p is UP^+-minimal such that $s'_p, 0 \mid\!\not\approx \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s}',k \dagger \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)]$. We first show by induction on p that, if $\vec{s}'$ is ${}^{l}_{k}$-bound-consistent at $1, \ldots, p$, then $\vec{s}$ is ${}^{l+1}_{k}$-bound-consistent at $1, \ldots, p$ and $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$, and otherwise $\mathsf{UP}^+(s_p) \supseteq \mathsf{UP}^+(s'_p)$. Then we prove that if $\vec{s}'\,|^{l}_{k} \mid\!\approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$, which holds by assumption, then also $\vec{s}\,|^{l+1}_{k} \mid\!\approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$.

The base case is trivial. For the induction step suppose the claim holds for $p-1$. First suppose $\vec{s}'$ is ${}^{l}_{k}$-bound-consistent at $1, \ldots, p$. Then by induction and Lemmas D.1.1 and 6.8.3, $\lfloor \vec{s}, k \dagger \phi_i \rfloor \geq p$ iff $\lfloor \vec{s}', k \dagger \phi_i \rfloor \geq p$. Moreover by induction and Lemmas 7.3.3, D.1.1, and 6.8.3 and ${}^{l}_{k}$-bound-consistency of $\vec{s}'$, $\lfloor \vec{s}, l+1 \ddagger \phi_i \rfloor \geq p$ iff $\lfloor \vec{s}', l \ddagger \phi_i \rfloor \geq p$. Thus and by induction, $\vec{s}$ is ${}^{l+1}_{k}$-bound-consistent at $1, \ldots, p$, and by Theorem 6.8.4,

$\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$. Now suppose $\vec{s}'$ is not ${}^{l}_{k}$-bound-consistent at some $p' \leq p$. Then by induction and Lemmas D.1.1 and 6.8.3, $\lfloor \vec{s}, k \mathrel{\overset{\circ}{\mid}} \phi_i \rfloor \geq p$ if $\lfloor \vec{s}', k \mathrel{\overset{\circ}{\mid}} \phi_i \rfloor \geq p$. Thus by Theorem 6.8.4, $\mathsf{UP}^+(s_p) \supseteq \mathsf{UP}^+(s'_p)$.

Let j be maximal such that $\vec{s}'$ is ${}^{l}_{k}$-bound-consistent at $1, \ldots, j$. Observe that by the above induction, $\mathsf{UP}^+((\vec{s}|^{l+1}_{k})_p) = \mathsf{UP}^+((\vec{s}'|^{l}_{k})_p)$ for all $p \leq j$, and $\mathsf{UP}^+((\vec{s}|^{l+1}_{k})_p) \supseteq \mathsf{UP}^+((\vec{s}'|^{l}_{k})_{p'})$ for all $p > j$ and $p' > j$ (**).

Now by assumption, $\vec{s}'|^{l}_{k} \mathrel{|\!\!\approx} \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$. Then for all p, if $p \leq \lfloor \vec{s}'|^{l}_{k}, l' \mathrel{\overset{\bullet}{\mid}} \phi \rfloor$, then $(\vec{s}'|^{l}_{k})_p, k' \mathrel{|\!\!\overset{\circ}{\approx}} (\phi \supset \psi)$. If $\lfloor \vec{s}'|^{l}_{k}, l' \mathrel{\overset{\bullet}{\mid}} \phi \rfloor \leq j$, then by (*) and Lemma 6.8.3, for all p, if $p \leq \lfloor \vec{s}|^{l}_{k}, l' \mathrel{\overset{\bullet}{\mid}} \phi \rfloor$, then $(\vec{s}|^{l}_{k})_p, k' \mathrel{|\!\!\overset{\circ}{\approx}} (\phi \supset \psi)$. Otherwise, by (**) and Lemma 6.5.3 and 6.8.3, for all p, $(\vec{s}|^{l}_{k})_p, k' \mathrel{|\!\!\overset{\circ}{\approx}} (\phi \supset \psi)$. Thus in either case, $\vec{s}|^{l}_{k} \mathrel{|\!\!\approx} \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$. □

Theorem 7.4.3 *If $\mathbf{O}^{l}_{k}\Gamma \mathrel{|\!\!\approx} \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$, then $\mathbf{O}^{\tilde{l}}_{\tilde{k}}\Gamma \mathrel{|\!\!\approx} \mathbf{B}^{\tilde{l}'}_{\tilde{k}'}(\phi \Rightarrow \psi)$ for all $\tilde{k} \geq k$, $\tilde{l} \geq l$, $\tilde{k}' \geq k'$, $\tilde{l}' \geq l$.*

Proof. Follows from Lemmas D.2.1 and D.2.2 by an easy induction on k, l, k', l'. □

D.3 Proof of the soundness theorem

In this appendix we prove that belief entailments with proper$^+$ knowledge bases in $\mathcal{BOL}$ are sound with respect to $\mathcal{BO}$, Theorem 7.4.4. For this section, let π and $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be proper$^+$.

Lemma D.3.1 $\pi \models \phi$ *iff* $\mathsf{UP}^+(\mathsf{gnd}(\pi)) \models \phi$.

Proof. Suppose $\pi = \bigwedge \forall \vec{x}_j c_j$. Then $w \models \pi$ iff $w \models {c_j}^{\vec{x}_j}_{\vec{n}}$ for all $\vec{n}$ for all j iff $w \models c$ for all $c \in \mathsf{gnd}(\pi)$ iff (by Lemma 6.3.5) $w \models c$ for all $c \in \mathsf{UP}^+(\mathsf{gnd}(\pi))$. □

Lemma D.3.2 *Suppose $\vec{e}$ is such that $w \in e_p$ iff $w \models \bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)$. Suppose $\vec{s}$ is such that s_p is UP^+-minimal such that $s_p, 0 \mathrel{|\!\!\overset{\circ}{\approx}} \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s}, k \mathrel{\overset{\circ}{\mid}} \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)]$. Then*

(i) $\lfloor \vec{s}, k' \mathrel{\overset{\circ}{\mid}} \phi \rfloor \leq \lfloor \vec{e} \mid \phi \rfloor$;

(ii) if $s_p, k' \mathrel{|\!\!\overset{\circ}{\approx}} \phi$ for all $p \in \mathbb{P}$, then $w \models \phi$ for all $w \in e_p$ and $p \in \mathbb{P}$.

Proof. (i) We show by induction on p that $\lfloor \vec{s}, k' \mathrel{\overset{\circ}{\mid}} \phi \rfloor \geq p$ implies $\lfloor \vec{e} \mid \phi \rfloor \geq p$. The base case is trivial. For the induction step suppose $\lfloor \vec{s}, k' \mathrel{\overset{\circ}{\mid}} \phi \rfloor \geq p$. By Lemma D.1.1, $s_{p'}, k' \mathrel{|\!\!\overset{\circ}{\approx}} \neg\phi$ for all $p' < p$. By Theorem 6.8.4, $\mathsf{UP}^+(s_{p'}) = \mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_{i:\lfloor \vec{s}, k \mathrel{\overset{\circ}{\mid}} \phi_i \rfloor \geq p'}(\phi_i \supset \psi_i)]))$ for all $p' < p$. By Lemma 6.8.3 and Theorem 6.5.1, $\mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_{i:\lfloor \vec{s}, k \mathrel{\overset{\circ}{\mid}} \phi_i \rfloor \geq p'}(\phi_i \supset \psi_i)])) \models \neg\phi$ for all $p' < p$. By Lemma D.3.1 and Theorem 6.9.2, $\bigwedge_{i:\lfloor \vec{s}, k \mathrel{\overset{\circ}{\mid}} \phi_i \rfloor \geq p'}(\phi_i \supset \psi_i) \models \neg\phi$ for all $p' < p$. By induction, $\lfloor \vec{s}, k \mathrel{\overset{\circ}{\mid}} \phi_i \rfloor \geq p'$ implies $\lfloor \vec{e} \mid \phi_i \rfloor \geq p'$ for all $p' < p$. Thus $\bigwedge_{i:\lfloor \vec{e} \mid \phi_i \rfloor \geq p'}(\phi_i \supset \psi_i) \models \neg\phi$ for all $p' < p$. Thus $\lfloor \vec{e} \mid \phi \rfloor \geq p$.

(ii) Suppose $s_p, k' \mathrel{|\!\approx} \phi$ for all $p \in \mathbb{P}$. In particular, $s_{m+1}, k' \mathrel{|\!\approx} \phi$. By Theorem 6.8.4, $\mathsf{UP}^+(s_{m+1}) = \mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_{i:\lfloor\vec{s},k\dagger\phi_i\rfloor \geq m+1}(\phi_i \supset \psi_i)]))$. By Lemma 6.8.3 and Theorem 6.5.1, $\mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_{i:\lfloor\vec{s},k\dagger\phi_i\rfloor \geq m+1}(\phi_i \supset \psi_i)])) \models \phi$. By Lemma D.3.1 and Theorem 6.9.2, $\bigwedge_{i:\lfloor\vec{s},k\dagger\phi_i\rfloor \geq m+1}(\phi_i \supset \psi_i) \models \phi$. By (i), $\lfloor\vec{s}, k \dagger \phi_i\rfloor \geq m+1$ implies $\lfloor\vec{e} \mid \phi_i\rfloor \geq m+1$. Thus $\bigwedge_{i:\lfloor\vec{e} \mid \phi_i\rfloor \geq m+1}(\phi_i \supset \psi_i) \models \phi$. Thus $w \models \phi$ for all $w \in e_{m+1}$. Since $e_p \subseteq e_{m+1}$ for all $p \in \mathbb{P}$ by Theorem 4.5.3, $w \models \phi$ for all $p \in \mathbb{P}$ and $w \in e_p$. □

Lemma D.3.3 *Suppose $\vec{e}$ is such that $w \in e_p$ iff $w \models \bigwedge_{i:\lfloor\vec{e} \mid \phi_i\rfloor \geq p}(\phi_i \supset \psi_i)$. Suppose $\vec{s}$ is such that s_p is UP^+-minimal such that $s_p, 0 \mathrel{|\!\approx} \mathsf{NF}[\bigwedge_{i:\lfloor\vec{s},k\dagger\phi_i\rfloor \geq p}(\phi_i \supset \psi_i)]$. Suppose $\vec{s}$ is $\frac{l}{k}$-bound-consistent at $1, \ldots, p$. Then*

(i) $\lfloor\vec{s}, l' \dagger \phi\rfloor \geq \lfloor\vec{e} \mid \phi\rfloor$;

(ii) if $s_p, k' \mathrel{|\!\approx} \phi$, then $w \models \phi$ for all $w \in e_p$;

(iii) if $s_p, l' \mathrel{|\!\not\approx} \phi$, then $w \not\models \phi$ for some $w \in e_p$.

Proof. Note that to show (i) it suffices to establish that $\lfloor\vec{e} \mid \phi\rfloor \geq p$ implies $\lfloor\vec{s}, l' \dagger \phi\rfloor \geq p$ for all $p \in \mathbb{P}$. We prove this along with (ii) and (iii) by induction on p. The base case for (i) is trivially true. As for (ii) and (iii), s_1 is UP^+-minimal such that $s_1, 0 \mathrel{|\!\approx} \mathsf{NF}[\bigwedge_i(\phi_i \supset \psi_i)]$. By Theorem 6.8.4, $\mathsf{UP}^+(s_1) = \mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_i(\phi_i \supset \psi_i)]))$. If $s_1, k' \mathrel{|\!\approx} \phi$, then by Lemma 6.8.3 and Theorem 6.5.1, $\mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_i(\phi_i \supset \psi_i)])) \models \phi$, and by Lemma D.3.1 and Theorem 6.9.2, $\bigwedge_i(\phi_i \supset \psi_i) \models \phi$, and thus $w \models \phi$ for all $w \in e_1$; hence (ii) holds. If $s_1, l' \mathrel{|\!\not\approx} \phi$, then by Lemma 6.8.3 and Theorem 6.7.1, $\mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_i(\phi_i \supset \psi_i)])) \not\models \phi$, and by Lemma D.3.1 and Theorem 6.9.2, $\bigwedge_i(\phi_i \supset \psi_i) \not\models \phi$, and thus $w \not\models \phi$ for some $w \in e_1$; hence (iii) holds.

For the induction step, suppose the lemma holds for $p - 1$. Suppose $\lfloor\vec{s}, l' \dagger \phi\rfloor < p$. Then $s_{p'}, l' \mathrel{|\!\not\approx} \neg\phi$ for some $p' < p$. By induction and (iii), $w \not\models \neg\phi$ for some $w \in e_{p'}$ and for some $p' < p$. Thus $\lfloor\vec{e} \mid \phi\rfloor < p$; so (i) holds. As for (ii) and (iii), s_p is UP^+-minimal such that $s_p, 0 \mathrel{|\!\approx} \mathsf{NF}[\bigwedge_{i:\lfloor\vec{s},k\dagger\phi_i\rfloor \geq p}(\phi_i \supset \psi_i)]$. For one thing, $\lfloor\vec{s}, k \dagger \phi_i\rfloor \geq p$ only if (Lemma D.3.2) $\lfloor\vec{e} \mid \phi_i\rfloor \geq p$. For another, $\lfloor\vec{s}, k \dagger \phi_i\rfloor < p$ (by assumption) $\lfloor\vec{s}, l \dagger \phi_i\rfloor < p$ only if (by (i)) $\lfloor\vec{e} \mid \phi_i\rfloor < p$. Hence, $\lfloor\vec{s}, k \dagger \phi_i\rfloor \geq p$ iff $\lfloor\vec{e} \mid \phi_i\rfloor \geq p$. Thus s_p is UP^+-minimal such that $s_p, 0 \mathrel{|\!\approx} \mathsf{NF}[\bigwedge_{i:\lfloor\vec{e} \mid \phi_i\rfloor \geq p}(\phi_i \supset \psi_i)]$. By Theorem 6.8.4, $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_{i:\lfloor\vec{e}\dagger\phi_i\rfloor \geq p}(\phi_i \supset \psi_i)]))$. If $s_p, k' \mathrel{|\!\approx} \phi$, then by Lemma 6.8.3 and Theorem 6.5.1, $\mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_{i:\lfloor\vec{e}\dagger\phi_i\rfloor \geq p}(\phi_i \supset \psi_i)])) \models \phi$, and by Lemma D.3.1 and Theorem 6.9.2, $\bigwedge_{i:\lfloor\vec{e}\dagger\phi_i\rfloor \geq p}(\phi_i \supset \psi_i) \models \phi$, and thus $w \models \phi$ for all $w \in e_p$; hence (ii) holds. If $s_p, l' \mathrel{|\!\not\approx} \phi$, then by Lemma 6.8.3 and Theorem 6.7.1, $\mathsf{UP}^+(\mathsf{gnd}(\mathsf{NF}[\bigwedge_{i:\lfloor\vec{e}\dagger\phi_i\rfloor \geq p}(\phi_i \supset \psi_i)])) \not\models \phi$, and by Lemma D.3.1 and Theorem 6.9.2, $\bigwedge_{i:\lfloor\vec{e}\dagger\phi_i\rfloor \geq p}(\phi_i \supset \psi_i) \not\models \phi$, and thus $w \not\models \phi$ for some $w \in e_p$; hence (iii) holds. □

Theorem 7.4.4 *If* $\mathbf{O}^{l}_{k}\Gamma \approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$, *then* $\mathbf{O}\Gamma \models \mathbf{B}(\phi \Rightarrow \psi)$.

Proof. Let $\mathbf{O}^{l}_{k}\Gamma \approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$, $\vec{e} \models \mathbf{O}\Gamma$. By Lemma D.1.2, there is an $\vec{s} \approx \mathbf{O}^{l}_{k}\Gamma$. By Rule $\mathcal{BOL}$6, $\vec{s} = \vec{s}'|^{l}_{k}$ for some $\vec{s}'$ whose setups s'_p are UP^+-minimal such that $s'_p, 0 \not\approx \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s}', k \dagger \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)]$.

First suppose $\vec{s}'$ is ${}^{l}_{k}$-bound-consistent at $1, \ldots, \lfloor \vec{s}, l' \dagger \phi \rfloor$. Then $\mathbf{O}^{l}_{k}\Gamma \approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$ iff (by assumption) for all $p \in \mathbb{P}$ with $p \leq \lfloor \vec{s}, l' \dagger \phi \rfloor$, $s_p, k' \approx (\phi \supset \psi)$ iff (by Lemma 6.8.3 and since $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(s'_p)$ for all $p \leq \lfloor \vec{s}, l' \dagger \phi \rfloor$) for all $p \in \mathbb{P}$ with $p \leq \lfloor \vec{s}', l' \dagger \phi \rfloor$, $s'_p, k' \approx (\phi \supset \psi)$ only if (by Lemma D.3.3) for all $p \in \mathbb{P}$ with $p \leq \lfloor \vec{e} \mid \phi \rfloor$ and $w \in e_p$, $w \models (\phi \supset \psi)$ iff $\vec{e} \models \mathbf{B}(\phi \Rightarrow \psi)$.

Now suppose $\vec{s}'$ is not ${}^{l}_{k}$-bound-consistent at some $p^* \leq \lfloor \vec{s}, l' \dagger \phi \rfloor$. Note that if $\vec{s} = \langle s_1, \ldots, s_{p^*}, \ldots, s_j \rangle$ and $\vec{s}' = \langle s'_1, \ldots, s'_{j'} \rangle$, then $\mathsf{UP}^+(s_{p^*}) = \mathsf{UP}^+(s_{p^*+1}) = \ldots = \mathsf{UP}^+(s_j) = \mathsf{UP}^+(s'_{j'})$ (*). Then $\mathbf{O}^{l}_{k}\Gamma \approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$ iff (by assumption) for all $p \in \mathbb{P}$ with $p \leq \lfloor \vec{s}, l' \dagger \phi \rfloor$, $s_p, k' \approx (\phi \supset \psi)$ iff (by (*) and Lemmas 6.8.3 and 6.5.3) $s'_p, k' \approx (\phi \supset \psi)$ for all $p \in \mathbb{P}$ only if (by Lemma D.3.2) $w \models (\phi \supset \psi)$ for all $w \in e_p$ and for all $p \in \mathbb{P}$ only if $\vec{e} \models \mathbf{B}(\phi \Rightarrow \psi)$. $\square$

D.4 Proof of the decidability theorem

In this appendix we prove the correctness of our decision procedure for belief entailments in $\mathcal{BOL}$, Theorem 7.5.4, and the complexity result Theorem 7.5.5. For this section, let π and $\Gamma = \{\phi_1 \Rightarrow \psi_1, \ldots, \phi_m \Rightarrow \psi_m\}$ be proper$^+$.

Lemma D.4.1 *Let* N *contain the names from* Γ *plus* $\max\{k, l\} \cdot v + v$ *names for* $v \geq |\Gamma|_{\mathrm{w}}$. *Then a sequence of proper*$^+$ $\pi_1, \ldots, \pi_j$ *with names names from* Γ *and* $v \geq |\pi_p|_{\mathrm{w}}$ *exists such that* $\langle \mathsf{gnd}(\pi_1), \ldots, \mathsf{gnd}(\pi_j) \rangle \approx \mathbf{O}^{l}_{k}\Gamma$ *and* $\langle \mathsf{gnd}_N(\pi_1), \ldots, \mathsf{gnd}_N(\pi_j) \rangle = \mathsf{MOD}[N, k, l, \Gamma]$.

Proof. Let $\pi_p = \mathsf{NF}[\bigwedge_{i:\lfloor \vec{s}, k \dagger \phi_i \rfloor \geq p}(\phi_i \supset \psi_i)]$ where $\vec{s}$ is such that $s_p = \mathsf{gnd}(\pi_p)$ for every $p \in \mathbb{P}$. Then $\vec{s}$ and $\pi_1, \ldots, \pi_j$ are well-defined and $\vec{s}|^{l}_{k} \approx \mathbf{O}^{l}_{k}\Gamma$ by Lemma D.1.2. Let $\vec{s}' = \langle \mathsf{gnd}_N(\pi_1), \ldots, \mathsf{gnd}_N(\pi_j) \rangle$. We show by induction on $p \in \mathbb{P}$ that

(i) $\lfloor \vec{s}, k \dagger \phi_i \rfloor \geq p$ iff $p = 1$ or $\mathsf{S}[N, s'_{p-1}, k, \neg\phi_i] = 1$;

(ii) $\vec{s}$ is ${}^{l}_{k}$-bound-consistent at p iff for all i,
$\max\{\mathsf{S}[N, \mathsf{gnd}_N(\pi_{p'}), k, \neg\phi_i] \mid p' < p\} = \max\{\mathsf{C}[N, \mathsf{gnd}_N(\pi_{p'}), l, \neg\phi_i] \mid p' < p\}$.

By Lemma D.1.2, only the first $m + 1$ spheres of $\vec{s}$ differ (modulo UP^+), and hence the same holds for $\vec{s}'$. Thus and by (i), $\vec{s}'$ is precisely the $\vec{s}'$ from Definition 7.5.1. Moreover, by (ii), the maximal p such that $\vec{s}$ is ${}^{l}_{k}$-bound-consistent in the definition of $\vec{s}|^l_k$ (Definition 7.3.2) matches $p^\star$ in $\mathsf{MOD}[N,k,l,\Gamma]$ (Definition 7.5.1). Thus, since $s_p = \mathsf{gnd}(\pi_p)$ iff $s'_p = \mathsf{gnd}_N(\pi_p)$, the lemma follows once (i) and (ii) are proved.

The base case of the induction holds trivially. Now suppose the claim holds for p. For the induction step for (i), $\lfloor \vec{s}, k \dagger \phi_i \rfloor \geq p$ iff (by Lemma D.1.1) $s_{p'}, k \mathrel{|\!\!\approx} \neg\phi_i$ for all $p' < p$ iff (by Lemmas 6.5.3 and 6.8.3) $p = 1$ or $s_{p-1}, k \mathrel{|\!\!\approx} \neg\phi_i$ iff (by Theorem 6.8.7) $p = 1$ or $\mathsf{S}[N, s'_{p-1}, k, \neg\phi_i] = 1$.

For the induction step for (ii), $\vec{s}$ is ${}^l_k$-bound-consistent at p iff $\{i \mid \lfloor \vec{s}, k \dagger \phi_i \rfloor \geq p\} = \{i \mid \lfloor \vec{s}, l \dagger \phi_i \rfloor \geq p\}$ iff (by Lemma D.1.1) $\{i \mid s_{p'}, k \mathrel{|\!\!\approx} \neg\phi_i$ for all $p' < p\} = \{i \mid s_{p'}, l \mathrel{|\!\!\overset{\circ}{\approx}} \neg\phi_i$ for all $p' < p\}$ iff (by construction) $\{i \mid \mathsf{gnd}(\pi_{p'}), l \mathrel{|\!\!\overset{\circ}{\approx}} \neg\phi_i$ for all $p' < p\} = \{i \mid \mathsf{gnd}(\pi_{p'}), k \mathrel{|\!\!\approx} \neg\phi_i$ for all $p' < p\}$ iff (by Theorems 6.8.7 and 6.8.11) for all i, $\max\{\mathsf{S}[N, \mathsf{gnd}_N(\pi_{p'}), k, \neg\phi_i] \mid p' < p\} = \max\{\mathsf{C}[N, \mathsf{gnd}_N(\pi_{p'}), l, \neg\phi_i] \mid p' < p\}$. □

Theorem 7.5.4 *Let N contain the names from Γ, ϕ, ψ plus $\max\{k, l, k', l'\} \cdot v + v$ names for $v \geq |\Gamma|_\mathrm{w}$ and $v \geq |\phi|_\mathrm{w}$ and $v \geq |\psi|_\mathrm{w}$.*
Then $\mathbf{O}^l_k\Gamma \mathrel{|\!\!\approx} \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$ iff $\mathsf{BEL}[N,k,l,k',l',\Gamma,\phi,\psi] = 1$.

Proof. Suppose $\vec{s} = \langle s_1, \ldots, s_j \rangle$ such that $\vec{s} \mathrel{|\!\!\approx} \mathbf{O}^l_k\Gamma$, which exists by Lemma D.1.2, and $\vec{s}' = \langle s'_1, \ldots, s'_{j'} \rangle = \mathsf{MOD}[N,k,l,\Gamma]$. Then $\mathbf{O}^l_k\Gamma \mathrel{|\!\!\approx} \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$ iff (since $\vec{s}$ is unique modulo UP^+ by Lemma D.1.4, and by Lemma 6.8.3) for all $p \in \mathbb{P}$, if $p \leq \lfloor \vec{s}, l' \dagger \phi \rfloor$, then $s_p, k' \mathrel{|\!\!\approx} (\phi \supset \psi)$ iff (by Lemma 6.8.3) $s_{p'}, k' \mathrel{|\!\!\approx} (\phi \supset \psi)$ for all $p' \leq \min\{p \mid s_p, l' \mathrel{|\!\!\overset{\circ}{\approx}} \neg\phi$ or $p = j\}$ iff (by Lemma 6.5.3) $s_{p'}, k' \mathrel{|\!\!\approx} (\phi \supset \psi)$ for $p' = \min\{p \mid s_p, l' \mathrel{|\!\!\overset{\circ}{\approx}} \neg\phi$ or $p = j\}$ iff (by Lemmas D.4.1 and 6.8.3) a sequence of proper$^+$ $\pi_1, \ldots, \pi_j$ exists with names from Γ and $v \geq |\pi_p|_\mathrm{w}$ such that $\mathsf{UP}^+(s_p) = \mathsf{UP}^+(\mathsf{gnd}(\pi))$ and $s'_p = \mathsf{gnd}_N(\pi)$ for all $p \in \mathbb{P}$, and $\mathsf{gnd}(\pi_{p'}), k' \mathrel{|\!\!\approx} (\phi \supset \psi)$ for $p' = \min\{p \mid \mathsf{gnd}(\pi_p), l' \mathrel{|\!\!\overset{\circ}{\approx}} \neg\phi$ or $p = j\}$ iff (by Theorems 6.8.7 and 6.8.11 and Lemma C.1.17) $\mathsf{S}[N, s'_{p'}, k', (\phi \supset \psi)] = 1$ for $p' = \min\{p \mid \mathsf{C}[N, s'_p, l', \neg\phi] = 1$ or $p = j'\}$ iff $\mathsf{BEL}[N,k,l,k',l',\Gamma,\phi,\psi] = 1$. □

Next we turn to the complexity.

Lemma D.4.2 $\mathsf{MOD}[N,k,l,\Gamma]$ *can be computed in time*
$O(m^2 \cdot (\|\Gamma\| + \max\{k,l\})^{2\cdot(\max\{k,l\}+1)} \cdot (\|\Gamma\| + |N|)^{2\cdot|\Gamma|_\mathrm{w}\cdot(\max\{k,l\}+1)} \cdot 2^k)$.

Proof. To compute $\vec{s}'$ in Definition 7.5.1, $O(m^2)$ instances of $\mathsf{S}[N, s'_p, k, \neg\phi_i]$ need to be computed. Then to determine $\vec{s}$ we need to compute another $O(m^2)$ instances of $\mathsf{S}[N, s'_p, k, \neg\phi_i]$ as well as $\mathsf{C}[N, s'_p, l, \neg\phi_i]$. Since $|\neg\phi_i| \leq \|\Gamma\|$ and by Lemmas C.1.18 and C.1.19, the lemma follows. □

Lemma D.4.3 $\mathsf{BEL}[N,k,l,k',l',\Gamma,\phi,\psi]$ *can be computed in time*

$O(f(N,k,l,\Gamma) + g(N,\Gamma,l',\phi) + h(N,\Gamma,k',\phi,\psi))$ *where*

$$\begin{aligned}
f(N,k,l,\Gamma) &= m^2 \cdot (\|\Gamma\| + \max\{k,l\})^{2\cdot(\max\{k,l\}+1)} \cdot (\|\Gamma\| + |N|)^{2\cdot|\Gamma|_{\mathrm{w}}\cdot(\max\{k,l\}+1)} \cdot 2^k;\\
g(N,\Gamma,l',\phi) &= m \cdot (\|\Gamma\| + l')^{l'+1} \cdot |\neg\phi|^{l'+1} \cdot \\
&\qquad (\max\{|\Gamma|_{\mathrm{w}}, |\neg\phi|_{\mathrm{w}}\} + |N|)^{(\max\{|\Gamma|_{\mathrm{w}},|\neg\phi|_{\mathrm{w}}\}+|\neg\phi|_{\mathrm{w}})\cdot(l'+1)};\\
h(N,\Gamma,k',\phi,\psi) &= (\|\Gamma\| + k')^{k'+1} \cdot |(\phi \supset \psi)|^{k'+1} \cdot |N|^{(|\Gamma|_{\mathrm{w}}+|(\phi\supset\psi)|_{\mathrm{w}})\cdot(k'+1)} \cdot 2^{k'}.
\end{aligned}$$

Proof. Every setup of $\mathsf{MOD}[N,k,l,\Gamma]$ is $\mathsf{gnd}_N(\pi)$ for some π with $|\pi| \in O(\|\Gamma\|)$. Hence the complexities follow by Lemmas D.4.2, C.1.18, and C.1.19. □

Theorem 7.5.5 *Suppose* $\|\Gamma\| \geq |(\phi \supset \psi)|$. *Let* $j = \max\{k,l\}$ *and* $j' = \{k',l'\}$ *and* $i = \max\{j,j'\}$. *Then* $\mathbf{O}^l_k\Gamma \approx \mathbf{B}^{l'}_{k'}(\phi \Rightarrow \psi)$ *can be determined in*

$$\begin{aligned}
&O(m^2 \cdot (\|\Gamma\| + j)^{2\cdot(j+1)} \cdot ((|\Gamma|_{\mathrm{w}} + |(\phi \supset \psi)|_{\mathrm{w}} + 1) \cdot (\|\Gamma\| + i + 1))^{2\cdot|\Gamma|_{\mathrm{w}}\cdot(j+1)} \cdot 2^k +\\
&\quad m \cdot (\|\Gamma\| + j')^{j'+1} \cdot |(\phi \supset \psi)|^{j'+1} \cdot\\
&\qquad ((|\Gamma|_{\mathrm{w}} + |(\phi \supset \psi)|_{\mathrm{w}}) \cdot (\|\Gamma\| + i + 2))^{(\max\{|\Gamma|_{\mathrm{w}},|\Gamma|_{\mathrm{w}}\}+|(\phi\supset\psi)|_{\mathrm{w}})\cdot(j'+2)}).
\end{aligned}$$

Proof. Let N contain the names from Γ and $(\phi \supset \psi)$ plus $(i+1)\cdot\max\{|\Gamma|_{\mathrm{w}}, |(\phi \supset \psi)|_{\mathrm{w}}\}$ additional names. We can estimate $|N| \leq |\Gamma|_{\mathrm{w}} \cdot \|\Gamma\| + |\phi \supset \psi|_{\mathrm{w}} \cdot |(\phi \supset \psi)| + (i+1) \cdot \max\{|\Gamma|_{\mathrm{w}}, |(\phi \supset \psi)|_{\mathrm{w}}\} \leq |\Gamma|_{\mathrm{w}} \cdot (\|\Gamma\| + i + 1) + |\phi \supset \psi|_{\mathrm{w}} \cdot (|(\phi \supset \psi)| + i + 1) \leq (|\Gamma|_{\mathrm{w}} + |(\phi \supset \psi)|_{\mathrm{w}}) \cdot (\|\Gamma\| + i + 1)$. Then with Lemma D.4.3 the theorem follows. □

Bibliography

Adams, Ernest W. (1965). The Logic of Conditionals. In: *Inquiry* 8.1-4, pages 166–197 (cited on page 11).

Alchourrón, Carlos E., Peter Gärdenfors, and David Makinson (1985). On the logic of theory change: Partial meet contraction and revision functions. In: *Journal of Symbolic Logic* 50.2, pages 510–530 (cited on pages 14, 15, 104).

Anderson, Alan R. and Nuel D. Belnap (1975). *Entailment: The Logic of Relevance and Necessity*. Princeton University Press (cited on page 29).

Aucher, Guillaume (2005). A combined system for update logic and belief revision. In: *Proceedings of the Seventh Pacific Rim International Workshop on Multi-Agents*. Springer, pages 1–17 (cited on page 26).

Baader, Franz and Werner Nutt (2003). Basic Description Logics. In: *The Description Logic Handbook: Theory, Implementation, and Applications*. Edited by Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, pages 47–100 (cited on page 28).

Bacchus, Fahiem, Joseph Y. Halpern, and Hector J. Levesque (1999). Reasoning about Noisy Sensors and Effectors in the Situation Calculus. In: *Artificial Intelligence* 111.1–2, pages 171–208 (cited on pages 26, 73, 109).

Baltag, Alexandru and Sonja Smets (2008). A qualitative theory of dynamic interactive belief revision. In: *Texts in logic and games* 3, pages 9–58 (cited on page 26).

Belle, Vaishak, Gerhard Lakemeyer, and Hector J. Levesque (2016). A first-order logic of probability and only knowing in unbounded domains. In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI)* (cited on page 68).

Belnap, Nuel D. (1977). A useful four-valued logic. In: *Modern uses of multiple-valued logic*. Springer, pages 5–37 (cited on page 29).

Ben-David, Shai and Yael Gafni (1989). All we believe fails in impossible worlds. In: *Bar-Ilan Symposium on Artificial Intelligence* (cited on pages 9, 10).

Börger, Egon, Erich Grädel, and Yuri Gurevich (1997). *The Classical Decision Problem*. Springer (cited on pages 27, 28).

Boutilier, Craig (1991). Inaccessible Worlds and Irrelevance: Preliminary Report. In: *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 413–418 (cited on page 68).

Boutilier, Craig (1993). Revision Sequences and Nested Conditionals. In: *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 519–525 (cited on pages 17, 67, 72, 111).

Boutilier, Craig (1996). Iterated revision and minimal change of conditional beliefs. In: *Journal of Philosophical Logic* 25.3, pages 263–305 (cited on pages 17, 72).

Chappell, Timothy (2013). Plato on Knowledge in the Theaetetus. In: *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta (cited on page 7).

Church, Alonzo (1936a). A note on the Entscheidungsproblem. In: *Journal of Symbolic Logic* 1, pages 40–41 (cited on pages 10, 26, 37).

Church, Alonzo (1936b). An unsolvable problem of elementary number theory. In: *American Journal of Mathematics* 58, pages 345–363 (cited on pages 10, 26, 37).

Clark, Keith L. (1978). Negation as failure. In: *Logic and data bases*. Springer, pages 293–322 (cited on page 23).

Claßen, Jens (2013). Planning and Verification in the Agent Language Golog. PhD thesis. RWTH Aachen University (cited on page 45).

Claßen, Jens and Gerhard Lakemeyer (2006). Foundations for Knowledge-Based Programs using ES. In: *Proceedings of the Tenth Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 318–328 (cited on page 111).

Levesque, Hector J., Ray Reiter, Yves Lespérance, Fangzhen Lin, and Richard Scherl (1997). GOLOG: A Logic Programming Language for Dynamic Domains. In: *The Journal of Logic Programming* 31, pages 59–84 (cited on page 23).

Darwiche, Adnan and Judea Pearl (1997). On the logic of iterated belief revision. In: *Artificial Intelligence* 89.1, pages 1–29 (cited on pages 15–17, 104, 106).

Davis, Martin, George Logemann, and Donald Loveland (1962). A Machine Program for Theorem Proving. In: *Communications of the ACM* 5.7, pages 394–397 (cited on page 118).

Davis, Martin and Hilary Putnam (1960). A Computing Procedure for Quantification Theory. In: *Journal of the ACM* 7.3, pages 201–215 (cited on page 118).

De Giacomo, Giuseppe, Yves Lespérance, and Fabio Patrizi (2016). Bounded situation calculus action theories. In: *Artificial Intelligence* 237, pages 172–203 (cited on page 29).

Delgrande, James P. (1995). A framework for logics of explicit belief. In: *Computational Intelligence* 11.1, pages 47–88 (cited on pages 30, 31).

Delgrande, James P. and Yi Jin (2012). Parallel belief revision: Revising by sets of formulas. In: *Artificial Intelligence* 176.1, pages 2223–2245 (cited on page 17).

Delgrande, James P. and Hector J. Levesque (2012). Belief Revision with Sensing and Fallible Actions. In: *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 148–157 (cited on pages 25, 26).

Demolombe, Robert and Maria del Pilar Pozos Parra (2006). Belief Revision in the Situation Calculus Without Plausibility Levels. In: *Proceedings of the Sixteenth International Symposium on Methodologies for Intelligent Systems (ISMIS)*, pages 504–513 (cited on pages 25, 26).

Fagin, Ronald and Joseph Y. Halpern (1987). Belief, awareness, and limited reasoning. In: *Artificial intelligence* 34.1, pages 39–76 (cited on page 31).

Fagin, Ronald, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi (1995). *Reasoning about knowledge*. Volume 4. MIT press (cited on pages 7, 9, 40, 54).

Fang, Liangda and Yongmei Liu (2013). Multiagent Knowledge and Belief Change in the Situation Calculus. In: *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI)*, pages 304–312 (cited on pages 25, 26).

Fang, Liangda, Yongmei Liu, and Ximing Wen (2015). On the Progression of Knowledge and Belief for Nondeterministic Actions in the Situation Calculus. In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2955–2963 (cited on pages 25, 26).

Fikes, Richard E. and Nils J. Nilsson (1972). STRIPS: A new approach to the application of theorem proving to problem solving. In: *Artificial Intelligence* 2.3, pages 189–208 (cited on pages 19, 22).

Fikes, Richard E. and Nils J. Nilsson (1993). STRIPS, a retrospective. In: *Artificial Intelligence* 59.1, pages 227–232 (cited on page 19).

Finger, Joseph J. (1987). Exploiting constraints in design synthesis. PhD thesis. Stanford University (cited on page 20).

Finger, Marcelo and Renata Wassermann (2007). Anytime approximations of classical logic from above. In: *Journal of Logic and Computation* 17.1, pages 53–82 (cited on page 32).

Gärdenfors, Peter (1978). Conditionals and Changes of Belief. In: *Acta Philosophica Fennica* 30, pages 381–404 (cited on page 11).

Gärdenfors, Peter (1988). *Knowledge in flux: Modeling the dynamics of epistemic states*. The MIT press (cited on page 104).

Gärdenfors, Peter and David Makinson (1988). Revisions of knowledge systems using epistemic entrenchment. In: *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge (TARK)*, pages 83–95 (cited on page 15).

Gelfond, Michael and Vladimir Lifschitz (1993). Representing action and change by logic programs. In: *The Journal of Logic Programming* 17.2, pages 301–321 (cited on pages 19, 25).

Gelfond, Michael and Vladimir Lifschitz (1998). Action languages. In: *Linköping Electronic Articles in Computer and Information Science* 3.16 (cited on pages 19, 25).

Gödel, Kurth (1929). Über die Vollständigkeit des Logikkalküls. PhD thesis. University of Vienna (cited on page 37).

Goldszmidt, Moisés and Judea Pearl (1996). Qualitative Probabilities for Default Reasoning, Belief Revision and Causal Modelling. In: *Artificial Intelligence* 84, pages 57–112 (cited on page 11).

Grädel, Erich, Phokion G. Kolaitis, and Moshe Y. Vardi (1997). On the decision problem for two-variable first-order logic. In: *Bulletin of symbolic logic* 3.01, pages 53–69 (cited on pages 27, 28).

Grove, Adam (1988). Two modellings for theory change. In: *Journal of Philosophical Logic* 17.2, pages 157–170 (cited on pages 11, 15, 50).

Haas, Andrew R. (1987). The case for domain-specific frame axioms. In: *The Frame Problem in Artificial Intelligence*. Edited by Frank M. Brown, pages 343–348 (cited on pages 19, 23).

Halpern, Joseph Y. (1999). On the adequacy of modal logic. In: *Electronic News Journal on Reasoning about Actions and Change* 3.3 (cited on page 7).

Halpern, Joseph Y. (2015). A modification of the Halpern-Pearl definition of causality. In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3022–3033 (cited on page 11).

Halpern, Joseph Y. and Gerhard Lakemeyer (1995). Levesque's axiomatization of only knowing is incomplete. In: *Artificial Intelligence* 74.2, pages 381–387 (cited on page 10).

Halpern, Joseph Y., Yoram Moses, and Moshe Y. Vardi (1994). Algorithmic knowledge. In: *Proceedings of the Fifth Conference on Theoretical Aspects of Reasoning about Knowledge (TARK)*, pages 255–266 (cited on page 31).

Halpern, Joseph Y. and Judea Pearl (2005). Causes and explanations: A structural-model approach. Part I: Causes. In: *British Journal for Philosophy of Science* 56.4, pages 843–887 (cited on page 11).

Helmert, Malte (2006). The Fast Downward Planning System. In: *Journal of Artificial Intelligence Research* 26 (cited on page 19).

Hintikka, Jaakko (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press (cited on pages 7, 38).

Hintikka, Jaakko (1975). Impossible possible worlds vindicated. In: *Journal of Philosophical Logic* 4.4, pages 475–484 (cited on pages 29, 133).

Hoffmann, Jörg and Bernhard Nebel (2001). The FF planning system: Fast plan generation through heuristic search. In: *Journal of Artificial Intelligence Research* 14 (cited on page 19).

Humberstone, Lloyd (1987). The modal logic of 'all and only'. In: *Notre Dame Journal of Formal Logic* 28.2, pages 177–188 (cited on page 9).

Jin, Yi and Michael Thielscher (2004). Representing Beliefs in the Fluent Calculus. In: *Proceedings of the Sixteenth European Conference on Artificial Intelligence (ECAI)*, pages 823–827 (cited on page 24).

Jin, Yi and Michael Thielscher (2007). Iterated belief revision, revised. In: *Artificial Intelligence* 171.1, pages 1–18 (cited on page 16).

Kaplan, Aaron N. and Lenhart K. Schubert (2000). A computational model of belief. In: *Artificial Intelligence* 120.1, pages 119–160 (cited on page 31).

Katsuno, Hirofumi and Alberto O. Mendelzon (1991). Propositional knowledge base revision and minimal change. In: *Artificial Intelligence* 52.3, pages 263–294 (cited on page 18).

Kern-Isberner, Gabriele (2001). *Conditionals in Nonmonotonic Reasoning and Belief Revision*. Springer (cited on pages 11, 112).

Klassen, Toryn Q., Sheila A. McIlraith, and Hector J. Levesque (2015). Towards Tractable Inference for Resource-Bounded Agents. In: *Proceedings of the Twelfth International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense)* (cited on page 31).

Kleene, Stephen C. (2002). *Mathematical logic*. Courier Corporation (cited on pages 9, 34, 37).

Konolige, Kurt (1986). A Deduction Model of Belief. In: *Research notes in Artificial Intelligence* (cited on page 31).

Kowalski, Robert (1992). Database updates in the event calculus. In: *The Journal of Logic Programming* 12.1, pages 121–146 (cited on page 24).

Kowalski, Robert and Marek Sergot (1989). A logic-based calculus of events. In: *Foundations of knowledge base management*, pages 23–55 (cited on pages 19, 24).

Kripke, Saul (1959). A Completeness Theorem in Modal Logic. In: *Journal of Symbolic Logic* 24.1 (cited on pages 8, 38).

Kripke, Saul A. (1976). Is There a Problem About Substitutional Quantification? In: *Truth and Meaning*. Edited by Gareth Evans and John McDowell. Oxford University Press, pages 324–419 (cited on page 34).

Lakemeyer, Gerhard (1994). Limited reasoning in first-order knowledge bases. In: *Artificial Intelligence* 71.2, pages 213–255 (cited on page 30).

Lakemeyer, Gerhard (1996). Limited reasoning in first-order knowledge bases with full introspection. In: *Artificial Intelligence* 84.1, pages 209–255 (cited on page 30).

Lakemeyer, Gerhard and Hector J. Levesque (1998). $\mathcal{AOL}$: a logic of acting, sensing, knowing, and only knowing. In: *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 316–327 (cited on page 42).

Lakemeyer, Gerhard and Hector J. Levesque (2002). Evaluation-based reasoning with disjunctive information in first-order knowledge bases. In: *Proceedings of the Eighth Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 73–81 (cited on pages 30, 113, 131, 133, 139, 142).

Lakemeyer, Gerhard and Hector J. Levesque (2004). Situations, Si! Situation Terms, No! In: *Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 516–526 (cited on pages 42, 45, 46).

Lakemeyer, Gerhard and Hector J. Levesque (2005). Only-knowing: Taking it beyond autoepistemic reasoning. In: *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI)*, pages 633–638 (cited on pages 39, 67).

Lakemeyer, Gerhard and Hector J. Levesque (2009). A Semantical Account of Progression in the Presence of Defaults. In: *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 842–847 (cited on pages 22, 46, 101).

Lakemeyer, Gerhard and Hector J. Levesque (2011). A semantic characterization of a useful fragment of the situation calculus with knowledge. In: *Artificial Intelligence* 175.1, pages 142–164 (cited on pages 19, 22, 23, 33, 42–46, 71, 80, 87, 101, 108, 110, 111, 158).

Lakemeyer, Gerhard and Hector J. Levesque (2013). Decidable Reasoning in a Logic of Limited Belief with Introspection and Unknown Individuals. In: *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)* (cited on pages 30, 113, 114, 126, 131, 133, 142, 144).

Lakemeyer, Gerhard and Hector J. Levesque (2014). Decidable Reasoning in a Fragment of the Epistemic Situation Calculus. In: *Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 468–477 (cited on pages 30, 31, 113, 114, 119, 121, 131, 133, 142–144).

Lakemeyer, Gerhard and Hector J. Levesque (2016). Decidable Reasoning in a Logic of Limited Belief with Function Symbols. In: *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning (KR)*. To appear (cited on pages 30, 31, 113, 114, 131, 133, 142, 144).

Levesque, Hector J. (1984a). A logic of implicit and explicit belief. In: *Proceedings of the Fourth National Conference on Artificial Intelligence*, pages 198–202 (cited on pages 29–32).

Levesque, Hector J. (1984b). Foundations of a functional approach to knowledge representation. In: *Artificial Intelligence* 23.2, pages 155–212 (cited on pages 8, 12, 33, 63, 65, 68, 98).

Levesque, Hector J. (1990). All I know: a study in autoepistemic logic. In: *Artificial Intelligence* 42.2, pages 263–309 (cited on pages 9, 10, 13, 39, 67).

Levesque, Hector J. and Gerhard Lakemeyer (2001). *The Logic of Knowledge Bases*. MIT Press (cited on pages 8, 10, 12, 13, 33–37, 39–41, 63, 65, 67, 68, 174, 176).

Levesque, Hector J. and Gerhard Lakemeyer (2008). Cognitive Robotics. In: *Handbook of Knowledge Representation*. Edited by Frank Van Harmelen, Vladimir Lifschitz, and Bruce Porter. Elsevier, pages 869–886 (cited on page 2).

Levesque, Hector J. and Ray Reiter (1998). High-level robotic control: Beyond planning. In: *AAAI Spring Symposium on Integrating Robotics Research*. Volume 37 (cited on page 2).

Levi, Isaac (1988). Iteration of Conditionals and the Ramsey Test. In: *Synthese* 76.1, pages 49–81 (cited on page 67).

Lewis, David (1973). *Counterfactuals*. John Wiley & Sons (cited on pages 11, 15, 50, 67).

Lewis, David (1979). Counterfactual Dependence and Time's Arrow. In: *Noûs* 13, pages 455–476 (cited on page 11).

Libkin, Leonid (2013). *Elements of finite model theory* (cited on page 28).

Lifschitz, Vladimir, Leora Morgenstern, and David Plaisted (2008). Knowledge Representation and Classical Logic. In: *Handbook of Knowledge Representation*. Edited by Frank Van Harmelen, Vladimir Lifschitz, and Bruce Porter. Elsevier, pages 3–88 (cited on page 113).

Lin, Fangzhen (1995). Embracing Causality in Specifying the Indirect Effects of Actions. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1985–1991 (cited on page 20).

Lin, Fangzhen (2008). Situation Calculus. In: *Handbook of Knowledge Representation*. Edited by Frank Van Harmelen, Vladimir Lifschitz, and Bruce Porter. Elsevier, pages 649–669 (cited on page 19).

Lin, Fangzhen and Ray Reiter (1994). Forget It! In: *Proceedings of the AAAI Fall Symposium on Relevance*, pages 154–159 (cited on pages 82, 90, 110).

Lin, Fangzhen and Ray Reiter (1997). How to progress a database. In: *Artificial Intelligence* 92.1, pages 131–167 (cited on pages 22, 81, 82, 90, 91, 112).

Liu, Yongmei (2006). Tractable Reasoning in Incomplete First-order Knowledge Bases. PhD thesis. University of Toronto (cited on pages 30, 113, 114, 128, 131, 133, 143).

Liu, Yongmei and Gerhard Lakemeyer (2009). On First-order Definability and Computability of Progression for Local-effect Actions and Beyond. In: *Proceedings of the Twenty-First International Joint Conference on Artifical Intelligence (IJCAI)*, pages 860–866 (cited on pages 22, 112).

Liu, Yongmei, Gerhard Lakemeyer, and Hector J. Levesque (2004). A Logic of Limited Belief for Reasoning with Disjunctive Information. In: *Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR)* (cited on pages 30, 113–115, 128, 131, 133, 143).

Lynch, Christopher (2004). Unsound theorem proving. In: *Proceedings of the Workshop on Disproving: Non-Theorems, Non-Validity, Non-Provability*, pages 1–12 (cited on page 32).

McCarthy, John (1959). Programs with common sense. In: *Proceedings of the Symposium on Mechanization of Thought Processes*. Edited by Marvin Minsky. Her Majesty's Stationary Office (cited on pages 7, 12, 19).

McCarthy, John (1963). *Situations, Actions, and Causal Laws*. Technical Report AI Memo 2. AI Lab, Stanford University (cited on pages 19, 22, 41).

McCarthy, John (1977). Epistemological problems of artificial intelligence. In: *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 1038–1044 (cited on page 21).

McCarthy, John (1980). Circumscription—A Form of Non-Monotonic Reasoning. In: *Artificial Intelligence* 13, pages 27–39 (cited on pages 20, 25).

McCarthy, John (1986). Applications of circumscription to formalizing common-sense knowledge. In: *Artificial Intelligence* 28.1, pages 89–116 (cited on page 20).

McCarthy, John (1997). Modality, si! Modal logic, no! In: *Studia Logica* 59.1, pages 29–32 (cited on page 7).

McCarthy, John (2007). What Is Artificial Intelligence? Unpublished manuscript (cited on page 1).

McCarthy, John and Patrick J. Hayes (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In: *Machine Intelligence*. Volume 4. Edinburgh University Press, pages 463–502 (cited on pages 20, 41).

McCarthy, John, Marvin Minsky, Nathan Rochester, and Claude Shannon (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (cited on page 1).

McDermott, Drew, Malik Ghallab, Adele Howe, Craig Knoblock, Ashwin Ram, Manuela Veloso, Daniel Weld, and David Wilkins (1998). *PDDL—The Planning Domain Definition Language*. Technical report (cited on page 19).

McIlraith, Sheila A. (2000). Integrating Actions and State Constraints: A Closed-Form Solution to the Ramification Problem (Sometimes). In: *Artificial Intelligence* 116.1-2, pages 87–121 (cited on page 20).

Miller, Rob, Leora Morgenstern, and Theodore Patkos (2013). Reasoning About Knowledge and Action in an Epistemic Event Calculus. In: *Proceedings of the Eleventh International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense)* (cited on page 25).

Moore, Robert C. (1985). Semantical considerations on nonmonotonic logic. In: *Artificial Intelligence* 25.1, pages 75–94 (cited on pages 39, 67).

Mortimer, Michael (1975). On languages with two variables. In: *Mathematical Logic Quarterly* 21.1, pages 135–140 (cited on page 28).

Nardi, Daniele and Ronald J. Brachman (2003). An Introduction to Description Logics. In: *The Description Logic Handbook: Theory, Implementation, and Applications*. Edited by Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, pages 5–44 (cited on page 28).

Nayak, Abhaya C. (1994). Iterated belief change based on epistemic entrenchment. In: *Erkenntnis* 41.3, pages 353–390 (cited on pages 17, 72).

Nayak, Abhaya C., Maurice Pagnucco, and Pavlos Peppas (2003). Dynamic Belief Revision Operators. In: *Artificial Intelligence* 146, pages 193–228 (cited on pages 15–17, 72, 104, 107).

Neisser, Ulric, Gwyneth Boodoo, Thomas J. Bouchard Jr., A. Wade Boykin, Nathan Brody, Stephen J. Ceci, Diane F. Halpern, John C. Loehlin, Robert Perloff, Robert

J. Sternberg, and Susana Urbina (1996). Intelligence: knowns and unknowns. In: *American psychologist* 51.2, page 77 (cited on page 1).

Patel-Schneider, Peter F. (1990). A decidable first-order logic for knowledge representation. In: *Journal of automated reasoning* 6.4, pages 361–388 (cited on page 30).

Pearl, Judea (1990). System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In: *Proceedings of the Third Conference on Theoretical Aspects of Reasoning about Knowledge (TARK)*, pages 121–135 (cited on pages 11, 49, 61–63, 68, 151).

Pearl, Judea (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann (cited on page 68).

Pednault, Edwin P. D. (1989). ADL: Exploring the middle ground between STRIPS and the situation calculus. In: *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning (KR)*. Morgan Kaufmann, pages 324–332 (cited on pages 19, 23).

Peppas, Pavlos (2008). Belief Revision. In: *Handbook of Knowledge Representation*. Edited by Frank Van Harmelen, Vladimir Lifschitz, and Bruce Porter. Elsevier, pages 317–359 (cited on pages 13, 14, 16, 18).

Reiter, Ray (1980). A logic for default reasoning. In: *Artificial Intelligence* 13.1, pages 81–132 (cited on pages 39, 67).

Reiter, Ray (1991). The Frame Problem in the Situation Calculus: A Simple Solution (sometimes) and a Completeness Result for Goal Regression. In: *Artificial Intelligence and Mathematical Theory of Computation*. Edited by Vladimir Lifschitz. Academic Press, pages 359–380 (cited on pages 19–23, 41, 80, 81, 86).

Reiter, Ray (2001). *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. The MIT Press (cited on pages 2, 19, 21, 23, 41, 45, 73, 80, 81, 86).

Rott, Hans (2009). Shifting Priorities: Simple Representations for Twenty-Seven Iterated Theory Change Operators. In: *Towards Mathematical Philosophy*. Edited by David Makinson, Jacek Malinowski, and Heinrich Wansing. Volume 28. Springer, pages 269–296 (cited on page 17).

Sattler, Ulrike, Diego Calvanese, and Ralf Molitor (2003). Relationships with other Formalisms. In: *The Description Logic Handbook: Theory, Implementation, and Applications*. Edited by Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, pages 142–183 (cited on page 28).

Schaerf, Marco and Marco Cadoli (1995). Tractable reasoning via approximation. In: *Artificial Intelligence* 74.2, pages 249–310 (cited on pages 30, 32).

Scherl, Richard and Hector J. Levesque (2003). Knowledge, Action, and the Frame Problem. In: *Artificial Intelligence* 144.1–2, pages 1–39 (cited on pages 22, 23, 41, 72, 109, 110).

Schubert, Lenhart K. (1989). Monotonic solution of the frame problem in the situation calculus: an efficient method for worlds with fully specified actions. In: *Knowledge Representation and Defeasible Reasoning*. Edited by Henry E. Kyburg Jr., Ronald P. Loui, and Greg N. Carlson. Kluwer, pages 23–67 (cited on pages 19, 23).

Schwering, Christoph and Gerhard Lakemeyer (2014). A Semantic Account of Iterated Belief Revision in the Situation Calculus. In: *Proceedings of the Twenty-First European Conference on Artificial Intelligence (ECAI)*, pages 801–806 (cited on pages 49, 71, 217).

Schwering, Christoph and Gerhard Lakemeyer (2015). Projection in the Epistemic Situation Calculus with Belief Conditionals. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, pages 1583–1589 (cited on pages 49, 71, 217).

Schwering, Christoph and Gerhard Lakemeyer (2016). Decidable Reasoning in a First-Order Logic of Limited Conditional Belief. In: *Proceedings of the Twenty-First European Conference on Artificial Intelligence (ECAI).* To appear (cited on pages 113, 133, 217).

Schwering, Christoph, Gerhard Lakemeyer, and Maurice Pagnucco (2015). Belief Revision and Progression of Knowledge Bases in the Epistemic Situation Calculus. In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3214–3220 (cited on pages 67, 71, 112, 217).

Scott, Dana (1962). A decision method for validity of sentences in two variables. In: *Journal of Symbolic Logic* 27.377, page 74 (cited on page 28).

Shapiro, Steven, Maurice Pagnucco, Yves Lespérance, and Hector J. Levesque (2011). Iterated belief change in the situation calculus. In: *Artificial Intelligence* 175.1, pages 165–192 (cited on pages 25, 26, 104, 109, 110).

Spohn, Wolfgang (1988). Ordinal Conditional Functions: A Dynamic Theory of Epistemic States. In: *Causation in Decision, Belief Change, and Statistics*. Edited by William L. Harper and Brian Skyrms, pages 105–134 (cited on pages 17, 26, 112).

Stalnaker, Robert C. (1968). A theory of conditionals. In: *Studies in Logical Theory*. Oxford University Press, pages 98–112 (cited on pages 11, 67).

Tarski, Alfred (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. In: *Studia Philosophica* 1, pages 261–405 (cited on page 34).

Tarski, Alfred (1944). The Semantic Conception of Truth and the Foundations of Semantics. In: *Philosophy and Phenomenological Research* 4.3, pages 341–376 (cited on page 34).

Thielscher, Michael (1997). Ramifications and causality. In: *Artificial Intelligence* 89.1, pages 317–364 (cited on page 21).

Thielscher, Michael (1998). Introduction to the fluent calculus. In: *Linköping Electronic Articles in Computer and Information Science* 3.14 (cited on pages 19, 24).

Thielscher, Michael (1999). From situation calculus to fluent calculus: State update axioms as a solution to the inferential frame problem. In: *Artificial Intelligence* 111.1, pages 277–299 (cited on page 24).

Thielscher, Michael (2000). Representing the Knowledge of a Robot. In: *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 109–120 (cited on page 24).

Thielscher, Michael (2005). FLUX: A Logic Programming Method for Reasoning Agents. In: *Theory and Practice of Logic Programming* 5.4–5, pages 533–565 (cited on page 24).

Turing, Alan (1936). On computable numbers, with an application to the Entscheidungsproblem. In: *Proceedings of the London Mathematical Society* 42, pages 230–265 (cited on pages 10, 26, 37).

Van Benthem, Johan (2007). Dynamic logic for belief revision. In: *Journal of Applied Non-Classical Logics* 17.2, pages 129–155 (cited on pages 26, 111).

Van Ditmarsch, Hans (2005). Prolegomena to dynamic logic for belief revision. In: *Synthese* 147.2, pages 229–275 (cited on page 26).

Van Ditmarsch, Hans, Wiebe van der Hoek, and Barteld Pieter Kooi (2007). *Dynamic epistemic logic*. Springer (cited on page 26).

Vardi, Moshe Y. (1986). On epistemic logic and logical omniscience. In: *Proceedings of the First Conference on Theoretical Aspects of Reasoning about Knowledge (TARK)*, pages 293–305 (cited on page 31).

Vassos, Stavros, Gerhard Lakemeyer, and Hector J. Levesque (2008). First-Order Strong Progression for Local-Effect Basic Action Theories. In: *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 662–672 (cited on pages 22, 112).

Vassos, Stavros and Hector J. Levesque (2013). How to progress a database III. In: *Artificial Intelligence* 195, pages 203–221 (cited on pages 22, 81).

Waldinger, Richard (1981). Achieving several goals simultaneously. In: *Readings in artificial intelligence*. Edited by Bonnie Lynn Webber and Nils J. Nilsson. Morgan Kaufmann, pages 250–271 (cited on page 21).

Zhang, Hantao and Mark E. Stickel (1996). An efficient algorithm for unit propagation. In: *Proceedings of the Fourth International Symposium on Artificial Intelligence and Mathematics* (cited on pages 189, 190).

Statement of Originality

Parts of this thesis have been published in the following papers.

Christoph Schwering and Gerhard Lakemeyer (2014). A Semantic Account of Iterated Belief Revision in the Situation Calculus. In: *Proceedings of the Twenty-First European Conference on Artificial Intelligence (ECAI)*, pages 801–806.

Christoph Schwering and Gerhard Lakemeyer (2015). Projection in the Epistemic Situation Calculus with Belief Conditionals. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, pages 1583–1589.

Christoph Schwering, Gerhard Lakemeyer, and Maurice Pagnucco (2015). Belief Revision and Progression of Knowledge Bases in the Epistemic Situation Calculus. In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3214–3220.

Christoph Schwering and Gerhard Lakemeyer (2016). Decidable Reasoning in a First-Order Logic of Limited Conditional Belief. In: *Proceedings of the Twenty-First European Conference on Artificial Intelligence (ECAI)*. To appear.

I carried out this research under the supervision and guidance of my co-authors, Gerhard Lakemeyer and Maurice Pagnucco. To each paper I contributed the question under investigation, drafted the solution, and carried out most of the technical work and of the presentation.