On logics and semantics for interpretability

Mikec, Luka

PhD Program in Mathematics and Computer Science,
University of Barcelona Graduate School
PhD Program in Mathematics, Department of Mathematics,
Faculty of Science, University of Zagreb

Luka Mikec

ON LOGICS AND SEMANTICS FOR INTERPRETABILITY

COTUTELLE DOCTORAL THESIS

Barcelona, 2021

PhD Program in Mathematics and Computer Science,
University of Barcelona Graduate School
PhD Program in Mathematics, Department of Mathematics,
Faculty of Science, University of Zagreb

Luka Mikec

ON LOGICS AND SEMANTICS FOR INTERPRETABILITY

COTUTELLE DOCTORAL THESIS

Supervisors:
Dr. Joost J. Joosten
Prof. dr. sc. Mladen Vuković
Tutor:
Dr. Ramon Jansana Ferrer

Barcelona, 2021

Acknowledgments

I thank Joost J. Joosten and Mladen Vuković for the substantial amount of time, knowledge and enthusiasm they invested in our work, resulting in this thesis. While I preferred exploring mathematics for the sake of mathematics, and did not focus too much on the precise structure and content of the thesis, writing a thesis has the benefit of having mentors; in my case, two people who were ready to help, consistently, both in scientific matters and otherwise.

This thesis contains joint work with my coauthors, who greatly affected the quality of the text: Marta Bílková, Evan Goris, Joost J. Joosten, Fedor Pakhomov, Tin Perkov, Albert Visser and Mladen Vuković.

I thank members of the seminar groups on foundation of mathematics in Barcelona and Zagreb; in particular Vedran Čačić for his many insightful questions and comments over the past years.

I am grateful to Jan Mas Rovira, Aleix Solé Sánchez, and the employees of *Facultat de Matemàtiques i Informàtica (Barcelona)* and *Matematički odsjek (Zagreb)* for their time, help and understanding in resolving the administrative work created by the cotutelle arrangement.

Last but not least, thanks to my friends, family, and co-workers in Zagreb.

This work has been supported in part by various EU, Catalan, Croatian and Spanish grants, and institutional funding, in particular:

- HRZZ-UIP-05-2017-9219 (Croatian Science Foundation);
- HRZZ-IP-2018-01-7459 (Croatian Science Foundation);
- Department of Mathematics, Faculty of Science, Zagreb.

Abstract

Keywords: modal logic; metamathematics; formalised interpretability; interpretability logics; generalised Veltman semantics

In this thesis we will study various properties of formalised relativised interpretability. In the central part of this thesis we study for different interpretability logics the following aspects: completeness for modal semantics, decidability and algorithmic complexity.

In particular, we will study two basic types of relational semantics for interpretability logics. One is the Veltman semantics, which we shall refer to as the regular or ordinary semantics; the other is called generalised Veltman semantics. In the recent years and especially during the writing of this thesis, generalised Veltman semantics was shown to be particularly well-suited as a relational semantics for interpretability logics. In particular, modal completeness results are easier to obtain in some cases; and decidability can be proven via filtration in all known cases. We prove various new and reprove some old completeness results with respect to the generalised semantics. We use the method of filtration to obtain the finite model property for various logics.

Apart from results concerning semantics in its own right, we also apply methods from semantics to determine decidability (implied by the finite model property) and complexity of provability (and consistency) problems for certain interpretability logics.

From the arithmetical standpoint, we explore three different series of interpretability principles. For two of them, for which arithmetical and modal soundness was already known, we give a new proof of arithmetical soundness. The third series results from our modal considerations. We prove it arithmetically sound and also characterise frame conditions w.r.t. ordinary Veltman semantics. We also prove results concerning the new series and generalised Veltman semantics.

Resum (extended abstract in Catalan)

El tema d'aquesta tesis són les lògiques d'interpretabilitat les quals descriuen el comportament del predicat d'interpretabilitat. Per tal de discutir la interpretabilitat entre teories matemàtiques, permeteu que primer diguem unes paraules sobre interpretacions. Hi ha diferents nocions d'una interpretació en ús, però una cosa que tenen en comú és que totes involucren una traducció que preserva l'estructura; aquesta traducció transforma formules de la teoria interpretada a formules de la teoria interpretadora. Aquest mapa cal que preservi la demostrabilitat fins a un cert punt, i.e. si A és un teorem de la teoria interpretada, llavors la imatge de A ha de ser demostrable en la teoria interpretadora. Que aquest mapa *preservi l'estructura* significa que almenys commuta amb les connectives lògiques. Les fórmules quantificades poden ser modificades lleugerament quan són interpretades; específicament hom pot fitar el domini de (totes) les fórmules quantificades fent servir un predicat fixat anomenat *domini especificador* (i.e. estem interessats en interpretabilitat relativitzada). Això ens permet construir una interpretació de, per exemple, una teoria de nombres en una teoria de conjunts, on (per la construcció habitual) només alguns conjunts es fan servir per representar nombres. Podem requerir els axiomes de la teoria interpretada de ser demostrables en la teoria interpretadora, però també podem requerir que això també es compleixi per tots els teoremes de la teoria interpretada (aquesta diferència és rellevant només quan hom treballa en una metateoria dèbil).

Les lògiques d'interpretabilitat descriuen el comportament d'un tipus específic d'interpretabilitat. Per començar, limitem el nostre interès en teories de primer ordre. En segon lloc, només ens concentrem en interpretabilitat entre extensions finites d'una teoria fixada T. En tercer lloc, estem interessats en interpretabilitat *formalitzada*, i.e, no estudiem el problema de si $T + A$ interpreta $T + B$, sinó el problema de si T pot *demostrar* que $T + A$ interpreta $T + B$. En quart lloc, no estem interessats per quines A i B tenim que $T + A$ interpreta $T + B$, sinó que estem interessats en aquelles propietats que són estructurals en el sentit que es compleixen per qualsevol tria de A i B. Optem per teoremes d'interpretabilitat en aquesta tesi; i.e. per tal que $T + A$ interpreti $T + B$ hem requerit que la traducció de qualsevol *teorema* de $T + B$ sigui demostrable en $T + A$. La teoria T hauria de ser suficientment forta; i.e. seqüencial. Si tal teoria és axiomatitzable, té un predicat $\mathsf{Int}_\mathsf{T}(\cdot, \cdot)$, definit d'una manera natural, expressant el fet que el primer argument del predicat interpreta el segon argument. La lògica d'interpretabilitat de T és definida d'una manera molt semblant a la lògica de demostrabilitat de T, però amb

un operador binari: $\rhd$; la interpretació corresponent d'aquest operador és $\mathsf{Int_T}$. Així, la lògica d'interpretabilitat d'una teoria T és el conjunt de totes les fórmules lògiques modals en el llenguatge de lògiques d'interpretabilitat que són demostrables per qualsevol lectura aritmètica que es doni a les variables proposicionals i prenent els operadors modals a les seves aritmetitzacions corresponents. A diferència del que pot ser el cas en lògiques de demostrabilitat, la lògica d'interpretabilitat de T realment depèn de T.

Per exemple, la lògica d'interpretabilitat de la teoria de conjunts de Gödel-Bernays (que és la lògica denominada **IL**P), i la lògica d'interpretabilitat de l'Aritmètica de Peano (que és la lògica denominada **IL**M), difereixen.

Donada una teoria seqüencial T, hi ha una certa quantitat de contingut, normalment denominat **IL**(All), que la lògica d'interpretabilitat de T inevitablement ha de tenir. Els continguts exactes de **IL**(All) no són coneguts; de fet, millorar la fita inferior és la pregunta que motiva la major part de les investigació en aquest camp. Una simple fita és la *lògica d'interpretabilitat bàsica*, denominada **IL**. Aquesta és una extensió de la lògica de demostrabilitat i conté cinc esquemes d'axioma addicionals que en la literatura són coneguts com J1-J5.

Tornant a la qüestió de **IL**(All), hi ha una manera interessant i sorprenent de millorar les millors fites inferiors conegudes, i.e. de trobar nous principis d'interpretabilitat aritmèticament vàlids. L'enfocament és estudiar semàntiques relacionals modals (semblant a Kripke). Nous principis aritmèticament vàlids han sorgit prenent les condicions de marc de principis ja coneguts, modificant-les, i llavors obtenint la fórmula modal que caracteritza la condició de marc modificada. Això, efectivament, no garanteix la validesa aritmètica de la fórmula modal obtinguda de tal forma, però noves fórmules aritmèticament vàlides s'han descobert talment. Un altre enfocament relacionat és intentar establir completesa d'una certa extensió de **IL**. Si la demostració de completesa modal falla per a alguna extensió concreta, estendre l'extensió més enllà, fins que sigui modalment completa, pot produir noves fórmules aritmèticament vàlides (aquest intent serà seguit en el capítol final de la tesi).

Hi ha dos tipus de semàntiques modals per lògiques d'interpretabilitat. Una és coneguda com *semàntica regular Veltman* (o *semàntica ordinària Veltman*, o només *Veltman semantics* quan no hi ha risc d'ambigüitat). L'altra és coneguda com *semàntica generalitzada Veltman*, introduïda per Verbrugge, que combina una semàntica en l'estil de Kripke amb una semàntica de veïnat. La semàntica regular Veltman pot ser usada per demostrar completesa per moltes lògiques d'interpretabilitat. Tanmateix, per lògiques més complexes, la semàntica generalitzada Veltman es poden emprar per donar demostracions de completesa més simples i fàcils d'entendre. En els darrers anys i especialment durant la redacció d'aquesta tesi, la semàntica generalitzada Veltman ha sigut provada de ser particularment ben adequada com a semàntica relacional per lògiques d'interpretabilitat. En particular, resultats sobre completesa modal són més fàcils d'obtenir en alguns casos;

i decidibilitat pot ser demostrada via filtració en tots els casos coneguts. Demostrem diversos nous i redemostrem alguns resultats coneguts respecte la semàntica generalitzada. En alguns casos, només sabem que una lògica és completa respecte la semàntica generalitzada Veltman. També hi ha exemples de lògiques completes respecte semàntica generalitzada Veltman però incompletes respecte semàntica regular Veltman. Tots els resultats de complexitat (la majoria dels quals són establerts en aquesta tesi) estan basats en semàntica regular Veltman. Pel que fa a decidibilitat, sembla que la semàntica generalitzada Veltman és una eina més apropiada, ja que permet un mètode uniforme per obtenir la propietat de model finit.

En aquesta tesi estudiarem diverses propietats d'interpretabilitat relativitzada formalitzada.

En la part central d'aquesta tesi estudiem per diferents lògiques d'interpretabilitat els següents aspectes: completesa per semàntiques modal, decidibilitat i complexitat algorísmica.

A banda de resultats al voltant de les semàntiques en el seu si, també apliquem mètodes de semàntiques per determinar la complexitat de problemes de demostrabilitat (i de consistència) per certes lògiques d'interpretabilitat.

Des del punt de vista aritmètic, explorem tres sèries diferents de principis d'interpretabilitat. Per dos d'ells, pels quals la solidesa aritmètica i modal ja era coneguda, donem una nova demostració de solidesa aritmètica. La tercera sèrie resulta de les nostres consideracions modals. Demostrem que és sòlida aritmèticament i que també caracteritza condicions de marc respecte semàntica regular Veltman. A més, donem una demostració de completesa per certes lògiques relacionades amb la tercera sèrie (les lògiques $\mathbf{IL}\mathsf{WR}$ i $\mathbf{IL}\mathsf{W}_\omega$).

Permeteu que descrivim l'estructura de la tesi.

En el Capítol 1 donem una introducció informal del tema general de la tesi. En el Capítol 2 donem una introducció més formal, definicions bàsiques i presentem alguns resultats senzills.

En els dos capítols subseqüents explorem completesa modal. Primer introduïm l'eina clau: etiquetes asseguradores. Aquí presentem la teoria general d'etiquetes asseguradores, incloent la noció d'etiquetes asseguradores Γ-completes. Desenvolupem la teoria usada posteriorment en la tesi, però també demostrem resultats interessants per si sols (com la caracterització de Γ-completesa).

En el Capítol 4 fem servir etiquetes asseguradores per tal d'obtenir diversos resultats de completes respecte la semàntica generalitzada Veltman. Definim $\mathbf{IL}\mathsf{X}$-estructures per $\mathsf{X} \subseteq \{\mathsf{M}, \mathsf{P}, \mathsf{M_0}, \mathsf{P_0}, \mathsf{R}\}$ i $\mathsf{X} \subseteq \{\mathsf{W}, \mathsf{W}^*\}$ i demostrem que la les lògiques $\mathbf{IL}\mathsf{X}$ corresponents són completes respecte la seva classe de marcs característica. En particular obtenim que $\mathbf{IL}\mathsf{P_0}$ i $\mathbf{IL}\mathsf{R}$ són completes, els quals són resultats nous. També definim el *problema d'iteració d'etiqueta* i introduïm un tipus especial d'estructures, $\mathbf{IL}\mathsf{WP}$-estructures, que poden ser usades per solucionar aquest problema en el cas simple de la lògica **IL**P. La motivació d'això és que el problema d'iteració d'etiqueta reapareix en lògiques més complexes com **IL**WR, on la solució encara és desconeguda. Sospitem que la mateixa solució pot ser aplicable fins i tot en lògiques més complexes, però hi ha altres problemes que encara no s'han solucionat en aquest cas. Tornem al tema de completesa en el capítol final de la tesi on entre altres resultats donem una demostració condicional de la completesa de **IL**WR.

En el Capítol 5 apliquem resultats de completesa i obtenim resultats de decidibilitat. Aquest és una aplicació, i potser la més útil, de la semàntica generalitzada: l'habilitat de definir filtracions amb bon comportament.

El Capítol 6 tracta la complexitat; demostrem que **IL**, **IL**W i **IL**P són PSPACE-completes.

En el Capítol 7 treballem amb l'aspecte aritmètic de les lògiques d'interpretabilitat. Concretament, donem una nova demostració de solidesa per dues sèries de principis recentment descobertes.

En el capítol final, Capítol 8, introduïm una altra sèrie de principis, demostrem que és aritmèticament sòlida i la hi donem semàntica ordinària Veltman. Com ja hem mencionat abans, també donem demostracions condicionals de completesa per lògiques relacionades amb aquesta sèrie nova.

Prošireni sažetak (extended abstract in Croatian)

Tema ove disertacije su logike interpretabilnosti, koje opisuju ponašanje predikata interpretabilnosti. Prvo ćemo nešto reći o interpretacijama između teorija. Postoji nekoliko verzija interpretacija u upotrebi, ali ono što je zajedničko svima jest da se radi o preslikavanju koje čuva strukturalna svojstva. To preslikavanje preslikava formule interpretirane teorije u formule interpretirajuće teorije. Za ovo se preslikavanje zahtijeva da u nekoj mjeri očuva dokazivost, tj. ako je A aksiom interpretirane teorije, onda slika formule A mora biti dokaziva u interpretirajućoj teoriji. Zahtjev da preslikavanje čuva strukturalna svojstva znači da komutira s propozicionalnim veznicima. Za kvantificirane se formule dopušta malo odstupanje prilikom interpretacije; konkretno, moguće je ograničiti domenu (svih) kvantificiranih formula koristeći unaprijed određen predikat kojeg zovemo *specifikator domene* (tj. zanima nas *relativizirana* interpretabilnost). Ovo nam omogućuje izgradnju, primjerice, teorije brojeva u teoriji skupova, gdje (u uobičajenoj konstrukciji) samo neki skupovi predstavljaju brojeve. Možemo zahtijevati da su aksiomi interpretirane teorije dokazivi u interpretirajućoj teoriji, ali možemo to zahtijevati i za sve teoreme uopće interpretirane teorije (razlika je bitna samo kada se radi u slaboj metateoriji).

Logike interpretabilnosti opisuju ponašanje određene verzije interpretabilnosti. Za početak, ograničavamo se na teorije prvoga reda. Drugo, zanima nas samo interpretabilnost između konačnih proširenja neke unaprijed određene teorije T. Treće, zanima nas *formalizirana* interpretabilnost, tj. ne proučavamo problem vrijedi li da $T + A$ interpretira $T + B$, već se bavimo problemom može li T *dokazati* da $T + A$ interpretira $T + B$. Četvrto, ne zanima nas za koje pojedinačne formule A i B vrijedi da $T + A$ interpretira $T + B$, već nas zanimaju ona svojstva koja su strukturalna u smislu da vrijede za bilo kakav odabir formula A i B. U ovoj se disertaciji odlučujemo za interpretabilnost teorema, tj. za to da $T + A$ interpretira $T + B$ zahtijevamo da svaki prijevod *teorema* od $T + B$ bude dokaziv u $T + A$. Teorija T pritom mora biti dovoljno snažna, primjerice sekvencijalna. Ako je takva teorija aksiomatizabilna, onda postoji i predikat $\mathsf{Int}_T(\cdot, \cdot)$, definiran na prirodan način, koji izražava svojstvo da je prvi argument predikata interpretira drugi argument. Logika interpretabilnosti teorije T definira se na sličan način kao i logika dokazivosti teorije T, ali s dodanim binarnim modalnim operatorom $\rhd$ čija je intendirana interpretacija Int_T. Dakle, logika interpretabilnosti teorije T jest skup svih modalnih formula u jeziku logike interpretabilnosti koje su dokazive za koju god aritmetičku interpretaciju propozicionalnih varijabli te preslikavajući modalne operatore u njihove intendirane aritmetizacije. Logika interpretabilnosti teorije T ovisi o teoriji T.

Primjerice, logika interpretabilnosti Gödel–Bernaysove teorije skupova (što je logika koju označavamo kao $\mathbf{IL}\mathsf{P}$), i logika interpretabilnosti Peanove Aritmetike (što je logika koju označavamo kao $\mathbf{IL}\mathsf{M}$), razlikuju se.

Za sekvencijalnu teoriju T postoji određeni skup formula, koji se obično označava kao $\mathbf{IL}(\mathrm{All})$, kojeg logika interpretabilnosti bilo koje teorije T mora sadržavati. Točan sadržaj skupa $\mathbf{IL}(\mathrm{All})$ nije poznat; u stvari, popravljanje donje granice je pitanje koje motivira većinu istraživanja u ovom području. Jedna jednostavna donja granica je *osnovna logika interpretabilnosti*, koju označavamo s $\mathbf{IL}$. To je proširenje logike dokazivosti koje sadrži pet dodatnih shema aksioma koje se u literaturi označava sa $\mathsf{J1}$–$\mathsf{J5}$.

Vraćajući se na pitanje sadržaja skupa $\mathbf{IL}(\mathrm{All})$, postoji zanimljiv i iznenađujuć pristup podizanju najbolje poznate donje granice; drugim riječima, traženja novih aritmetički valjanih principa interpretabilnosti. Pristup o kojem je riječ jest proučavanje modalne relacijske semantike. Novi su aritmetički valjani principi otkriveni promatrajući karakteristična svojstva već poznatih principa, modificirajući ih, te određujući modalne formule koje karakteriziraju tako dobivena relacijska svojstva. Ovakav postupak naravno ne garantira aritmetičku valjanost tako otkrivenih modalnih formula, ali neki aritmetički valjani principi doista jesu pronađeni na ovaj način. Još jedan sličan pristup je pokušati dokazati modalnu potpunost određenih proširenja logike $\mathbf{IL}$. Ako dokaz modalne potpunosti ne uspije, daljnje proširivanje spomenutog proširenja, dok god se ne ustanovi modalna potpunost, može rezultirati novim aritmetički valjanim formulama (ovaj će se pristup primijeniti u posljednjem poglavlju ove disertacije).

Postoje dva osnovna tipa modalne semantike korištena za logike interpretabilnosti. Jedan je *regularna Veltmanova semantika* (ili *obična Veltmanova semantika*, ili samo *Veltmanova semantika* kad ne postoji mogućnost zabune). Druga je *generalizirana Veltmanova semantika* koju je uvela Verbrugge, a koja osim osobina relacijske semantike ima i osobine okolinske semantike. Regularnu Veltmanovu semantiku moguće je koristiti za dokaz potpunosti brojnih logika interpretabilnosti. Međutim, za kompleksnije logike, generalizirana semantika može se iskoristiti za dati jednostavnije i razumljivije dokaze potpunosti. Posljednjih godina i posebno tijekom pisanja ove disertacije, generalizirana se semantika pokazala kao posebno dobro primjenjiva relacijska semantika za logike interpretabilnosti. Preciznije, jednostavnije je doći do rezultata o modalnoj potpunosti, a odlučivost se može dokazati koristeći filtracije u svim poznatim slučajevima. Mi dokazujemo različite nove rezultate, a uz to dajemo i nove dokaze starih rezultata, vezane uz potpunost u odnosu na generaliziranu semantiku. U nekim slučajevima znamo samo da je određena logika potpuna u odnosu na generaliziranu semantiku. Štoviše, postoje primjeri logika koje su potpune u odnosu na generaliziranu semantiku, a za koje znamo da su nepotpune u odnosu na regularnu Veltmanovu semantiku. Svi poznati rezultati o složenosti (od kojih se većina dokazuje upravo u ovoj disertaciji) koriste regularnu Veltmanovu semantiku. Što se tiče odlučivosti, čini se da je generalizirana semantika ponovno pogodniji alat, jer omogućava uniformnu metodu dokazivanja svojstva konačnih modela.

U ovoj se disertaciji bavimo različitim svojstvima formalizirane relativizirane interpretabilnosti.

U središnjem dijelu disertacije bavimo se različitim logikama interpretabilnosti i sljedećim njihovim aspektima: potpunost u odnosu na modalnu semantiku, odlučivost te algoritamska složenost.

Osim rezultata koji se tiču same semantike, koristimo semantičke metode kako bismo odredili algoritamsku složenost problema dokazivosti (i konzistentnosti) za različite logike interpretabilnosti.

Što se tiče aritmetičkog aspekta, proučavamo tri niza principa interpretabilnosti. Za dva među njima, za koja su aritmetička i modalna adekvatnost već poznati, dajemo nove dokaze aritmetičke adekvatnosti. Treći je niz rezultat naših modalnih razmatranja. Dokazujemo da je aritmetički adekvatan i karakteriziramo klasu okvira u odnosu na običnu Veltmanovu semantiku. Osim toga, razmatramo potpunost nekih logika vezanih uz treći niz (radi se o logikama $\mathbf{IL}\mathsf{WR}$ i $\mathbf{IL}\mathsf{W}_\omega$).

Sad ćemo dati pregled strukture disertacije.

U prvom poglavlju dajemo neformalan uvod u okvirno područje kojem pripada ova disertacija.

U drugom poglavlju dajemo formalniji uvod, osnovne definicije te dokazujemo neke jednostavnije rezultate.

U iduća dva poglavlja istražujemo modalnu potpunost. Prvo uvodimo ključni alat: osiguravajuće oznake. Razvijamo općenitu teoriju osiguravajućih oznaka, uključujući koncept Γ-punih osiguravajućih oznaka. Razvijamo i teoriju koja se koristi kasnije u disertaciji, ali dokazujemo i rezultate zanimljive same po sebi (poput karakterizacije Γ-punih skupova).

U četvrtom poglavlju koristimo osiguravajuće oznake kako bismo dokazali potpunost različitih logika interpretabilnosti u odnosu na generaliziranu Veltmanovu semantiku. Definiramo $\mathbf{IL}\mathsf{X}$-strukture za $\mathsf{X} \subseteq \{\mathsf{M}, \mathsf{P}, \mathsf{M_0}, \mathsf{P_0}, \mathsf{R}\}$ te $\mathsf{X} \subseteq \{\mathsf{W}, \mathsf{W^*}\}$ i dokazujemo da je pripadna logika $\mathbf{IL}\mathsf{X}$ potpuna u odnosu na svoju karakterističnu klasu okvira. Posebno, dokazujemo da su logike $\mathbf{IL}\mathsf{P_0}$ i $\mathbf{IL}\mathsf{R}$ potpune, što su novi rezultati. Također definiramo *problem iteracije oznaka* i uvodimo specijalan tip struktura, $\mathbf{IL}\mathsf{WP}$-strukture, koje rješavaju ovaj problem u relativno jednostavnom slučaju logike $\mathbf{IL}\mathsf{P}$. Motivacija za ovo istraživanje jest što se problem iteracije oznaka javlja u složenijim logikama poput logike $\mathbf{IL}\mathsf{WR}$, gdje je potpuno rješenje zasad nepoznato. Mi vjerujemo da bi isto rješenje moglo biti iskoristivo i za druge složenije logike, ali za to ustvrditi potrebno je riješiti i neke druge probleme. Ponovno se bavimo potpunošću u finalnom poglavlju ove disertacije gdje se, između ostalih rezultata, bavimo dokazom potpunosti logike **IL**WR pod pretpostavkom postojanja odgovora na neke druge otvorene probleme.

U petom poglavlju koristimo dobivene rezultate o potpunosti i dokazujemo odlučivost različitih logika. Ovo je još jedna, i možda najkorisnija, primjena generalizirane semantike: mogućnost uniformne definicije filtracija.

Šesto poglavlje bavi se algoritamskom složenošću; dokazujemo da **IL**, **IL**W i **IL**P pripadaju klasi složenosti PSPACE-potpunih problema.

Sedmo poglavlje bavi se aritmetičkim aspektima logika interpretabilnosti. Dajemo nov dokaz aritmetičke adekvatnosti nedavno otkrivenih nizova principa interpretabilnosti.

U posljednjem poglavlju uvodimo još jedan niz principa interpretabilnosti, dokazujemo mu aritmetičku adekvatnost, i pronalazimo mu Veltmanovu semantiku. Kao što smo već najavili, dajemo i uvjetne dokaze potpunosti za logike koje se tiču novog niza principa interpretabilnosti.

Contents

Chapter 1

Introduction

This chapter is envisioned as a short and, to the extent that is attainable given the matter at hand, gentle introduction to the topics of this thesis. In the second chapter, *Preliminaries*, we give a technical introduction and lay the groundwork for the remainder of the thesis.

1.1 Gödel's theorems and Provability Logic

While the results we obtain in this thesis are not particularly shaped by Gödel's theorems and the phenomenon of *formal incompleteness*, the methods and tools involved in obtaining these results are still central in the field of formalised interpretability. So let us say a few words on Gödel's theorems. In 1920s the mathematician David Hilbert initiated what is now known as Hilbert's Program (proposed in [31]). This was a two-fold program; it called for the axiomatisation of all mathematics, and moreover this axiomatisation should be such that the resulting theory of mathematics is *provably* consistent. Thus, the result of the program would be a *theory* of mathematics (a combination of axioms and rules of inference using which one can prove mathematical results), but also a rigorous *proof* that this theory is consistent. In this context, *consistency* means one cannot, starting only with axioms and inferring their consequences step-by-step using the given rules of inference, prove a logical contradiction.

The general sentiment among logicians and involved mathematicians is that the goal of Hilbert's Program has been shown unattainable by *Gödel's incompleteness theorems*. The first incompleteness theorem, proved by Kurt Gödel in 1930 and published in 1931 ([24]), is best formulated as the result on *incompletability*: any reasonable theory is not and, more importantly, cannot be extended to, a reasonable and formally complete theory—a theory that proves either A or the negation of A for every possible sentence A. By *reasonable* we mean consistent, sufficiently strong (sequential) and presentable in a mechanical way (recursively enumerable). We will be interested only in reasonable theories, as these cover the majority of those used and studied in mathematics. From this point on we assume that all the theories we mention or quantify over are reasonable, even when we are not explicit about this.

Gödel proved the second incompleteness theorem in 1930 too. This theorem states that every reasonable theory is too weak to prove its own consistency. This does not mean that one cannot prove the consistency of a theory in some other theory. Trivially, one can prove consistency of a reasonable theory T in a theory T enriched with the axiom asserting that T is consistent. It should be noted that Gerhard Gentzen showed in 1936 that in the case of the first-order theory Peano Arithmetic (PA) one can provide a sufficiently strong consistency-proving extension in a more natural way, by using a strong enough version of the principle of transfinite induction ([23]). But the sole fact that one can prove consistency of T in a theory stronger than T, even if the axioms of the stronger theory are regarded as reasonable or natural, is not surprising in itself. What Hilbert's Program aimed for is a consistency proof in a theory that is *at most* as strong as T. And by Gödel's incompleteness theorems, such a proof is not possible. Methodologically, one may of course doubt about the purpose if it were possible to prove consistency of T in T itself. Either T is (a) inconsistent and then it will prove its own consistency or it is (b) consistent but proving consistency will, because of the possibility of (a), not be a particularly convincing evidence of the consistency.

An intriguing feature of Gödel's results is the method behind the proofs. The method Gödel used is also where the connection with this thesis becomes tighter. In order to prove formal incompleteness of any reasonable theory T, Gödel proved that for every such theory there exists a sentence, usually denoted by G_T, such that neither it nor its negation—in case T moreover only proves true statements—are provable in T. The amusing feature of G_T is its interpretation. It can be seen as being equivalent to the sentence "This sentence is not provable in T". If G_T is true, it is true but unprovable, and if it is false, it is provable but false; neither option being particularly attractive (and the first option being the lesser evil). This is reminiscent of the liar paradox, but while the liar paradox can be considered to be, at least in its basic form, a consequence of poor understanding of the way natural language works and/or unreasonable expectations of the concept of truth, the 'paradox' that G_T provides has rock-solid formal foundations. Thus, G_T is a mathematical sentence with some degree of self-referentiality.

In what sense can a mathematical sentence, mathematical sentences being sentences such as "There is no largest prime number."[1], become self-referential? The first step in constructing self-referential sentences is to code objects we wish to refer to with mathematical objects. In the case of Peano Arithmetic, we code symbols, formulas, and sequences

[1] In the standard language of first-order arithmetic,

$$(\forall n)(\exists m)(\forall k)\big((\exists \ell)k\ell = n+m+1 \rightarrow k = 1 \vee k = n+m+1\big).$$

of formulas, with numbers. The properties these syntactical objects may have are coded as predicates with free variables, with the standard interpretation in the form of various number-theoretic properties. For example, the property of a formula that it is a theorem of PA (i.e. that it is provable inside PA) might, had things been (wildly) simpler, correspond to the property of the number coding this formula to be divisible by 1931. A large part of Gödel's original proof deals with establishing one such (albeit more complex) possible connection between number theory and a *theory of syntax* (a theory of symbols, formulas and finite sequences of formulas).

In particular, the property that a formula is provable is one of the key ingredients of Gödel's original proof of his two celebrated incompleteness theorems. A way to define this property is to first define a formula with two free variables $\mathsf{Proof}_T(p, A)$ formalising the fact that (the finite sequence of formulas coded by the number) p is a proof of (the formula coded by the number) A. In a certain technical sense, this property really is not that much more complex than the property of being divisible by 1931. One can then define the property of being provable, i.e. being a theorem, by letting $\mathsf{Prov}_T(A) := \exists p \mathsf{Proof}_T(p, A)$. It is more convenient to speak of formulas and proofs, rather than their codes. In the remainder of the introduction we will write $\ulcorner p \urcorner$ to refer to the code of a proof p, and $\ulcorner A \urcorner$ to refer to the code of a formula A. Usually it is clear from the context whether a symbol represents a formula, a proof, or a similar object, or if we are actually referring to the code of this object. For example, if A is a sentence, we might write $\mathsf{Prov}_T(A)$. The intended reading is that we plug in *the code* of A, or actually its syntactical representation (so-called numerals[2]), into the open formula $\mathsf{Prov}_T(\cdot)$. Now, one can wonder (and people have, indeed, wondered) what sort of formulas concerning Prov_T are provable in T. For example, by propositional logic we know $\mathsf{Prov}_T(\ulcorner A \urcorner) \vee \neg \mathsf{Prov}_T(\ulcorner A \urcorner)$ must be provable for every sentence A.

By Gödel's first incompleteness theorem, we know that for some A the sentence $\mathsf{Prov}_T(\ulcorner A \urcorner) \vee \mathsf{Prov}_T(\ulcorner \neg A \urcorner)$ is *not* provable in T. By Gödels second incompleteness theorem, we know that $\neg \mathsf{Prov}_T(\ulcorner A \wedge \neg A \urcorner)$ is not provable in T for any A, either. Let us introduce a shorthand for $\mathsf{Prov}_T(\ulcorner A \urcorner)$: $\Box A$. What does the set of T-provable formulas built using only formula-placeholders $A, B, C, \ldots$, propositional logical connectives and $\Box$ look like? Let us call this set $\mathbf{PL}(T)$. Since the sentence G_T, whose existence can be thought of as being responsible for the incompleteness theorems, provably satisfies the property $G_T \leftrightarrow \Box \neg G_T$, one might expect $\mathbf{PL}(T)$ to behave highly erratically. Surprisingly, not only does $\mathbf{PL}(T)$ have a simple axiomatisation (for almost any theory T), but $\mathbf{PL}(T)$ is also decidable. That is, there is an algorithm that, given any formula, tells us

[2]We often need to refer to concrete numbers in formulas. The expressions we do this with are called numerals. There is of course more than one way to do this; one simple expression that will be interpreted as the number n (for any fixed n) is the expression $S(S(\ldots S(0)))$, with n occurrences of "S". The symbol "S" stands for the direct successor of a given number (in the standard model). See Chapter 7 for the definition of a better variant, *efficient numerals*.

whether this formula is contained in $\mathbf{PL}(T)$ or not.

Given a theory T, the set $\mathbf{PL}(T)$ defined in the way just described is called *the provability logic of* T. Perhaps surprisingly, almost all theories T have the same provability logic, known as the Gödel–Löb Provability Logic **GL**. This set of formulas has many properties one might expect of a logic, and in fact it is a simple example of a relatively well-behaved modal logic.

Before moving forward towards interpretability logics, which extend provability logics in a natural way, let us first say a few words on comparing theories and interpretations.

1.2 Comparing theories

If we have two theories T and U, a natural way to compare them would be to ask whether $T \subseteq U$ or $U \subseteq T$. Unfortunately, providing answers to these questions is not very informative with regard to the question of how different these theories really are. Suppose U and T are essentially the same theory, but differ in the fact that T uses the symbol $+$ in all the places where U uses the symbol $\times'$ and analogously for $\times$ in T and $+'$ in U. Clearly $T \nsubseteq U$ and $U \nsubseteq T$, even though T and U are the 'same' theory.

Taking into account other such differences between theories, which we may consider to be only superficial differences, we can obtain some notion of an *interpretation* of one theory in another theory. There are different notions of an interpretation in use, but what they all share is that they involve a structure-preserving mapping; mapping formulas of the interpreted theory to formulas of the interpreting theory. This mapping is required to preserve provability to some extent, i.e. if A is a theorem of the interpreted theory, then the image of A must be provable in the interpreting theory. That the mapping is *structure-preserving* means that it at least commutes with logical connectives. Quantified formulas are allowed to be modified slightly when interpreted; specifically one can bound the domain of (all) quantified formulas using a fixed predicate called the *domain specifier* (i.e. we are interested in *relativised* interpretability). This enables us to build an interpretation of, e.g., a number theory in a set theory, where (by the usual construction) only some sets are used to represent numbers. Finally, relational symbols of the interpreted theory are allowed to become any formula of the same arity in the language of the interpreting theory. We can ask for axioms of the interpreted theory to be provable in the interpreting theory, but we can also require this to hold for all theorems of the interpreted theory (this difference matters only when one works in a weak metatheory).

Before interpretability was studied as a subject in its own right, interpretations have been used to establish results across mathematics. Let us give a few examples. Perhaps the best-known examples are models of non-Euclidean geometries (for example, hyperbolic geometry [1]) in Euclidean geometry. To a logician, an even more famous example might be the interpretation of number theory in set theory. For example, the interpretation of first-order Peano Arithmetic in a modification of **ZF** where the axiom of infinity is replaced with its negation (in fact, interpretability goes both ways between these two theories, see [41]). Another example, one we already discussed, is the interpretability of the *theory of syntax* in any sufficiently strong theory such as Peano Arithmetic (and indeed in much weaker theories such as S^1_2 and $\mathrm{I}\Delta_0 + \Omega_1$).

These and other interpretations might be fascinating in their own right. But they have been put to use to provide more palpable results too. A famous result is another result by Gödel. In [30] Gödel constructed an interpretation of **ZF** extended with the continuum hypothesis (**CH**), in pure **ZF**. This implies that if **ZF** is consistent, then **ZF** + **CH** is consistent. Another example is relative undecidability. In general, if T interprets U and U is undecidable, T need not be undecidable (unlike the analogous situation with an inconsistent theory U). However, if U is *essentially* undecidable, then T is undecidable (this result is due to Tarski, see [61]).

In the next subsection we shall specify the aspects of interpretability that we are interested in this thesis. Before doing so, let me embark on a short personal digression to close this subsection. I first heard of modal logic, and of the provability logic **GL** in particular, in a course I attended during my undergraduate study (in Rijeka). My undergraduate thesis was about an automated search for interesting theorems of **GL**. The notion of interestingness was determined by various heuristics, such as "if A is a theorem, then $(B) \to (A)$ is not interesting". The result of this automated search was a certain number of theorems of **GL** that the algorithm deemed sufficiently interesting. I wanted to present this collection in some organised form, ideally a graph, and was faced with having to decide what would determine whether there is an arrow between any two given nodes (where nodes represent interesting theorems of **GL**). Some options I explored is to let A point to B if $\mathbf{K} \vdash (A) \to (B)$; if $\mathbf{K4} \vdash (A) \to (B)$ ($\mathbf{GL} \vdash (A) \to (B)$ is not a good choice since $\mathbf{GL} \vdash B$); or if B is a substitution instance of A. In retrospect, some version of the notion of interpretability is what I was after. At that point I did not know of interpretability, and in particular that interpretability logics are probably the best studied extension of provability logics. (An additional curiosity is that when I started working on my undergraduate thesis, I was not aware that one of the very few places where interpretability logics are studied is in the same country, in Zagreb, where I later moved for my master's degree.)

1.3 Interpretability logics

Interpretability logics describe the behaviour of a specific kind of interpretability. For a start, we limit our interest to first-order theories. Second, we only concern ourselves with the interpretability between finite extensions of some fixed theory T. Third, we are interested in *formalised* interpretability, i.e. we do not study the problem of whether $T+A$ interprets $T+B$, but rather the problem of whether T can *prove* that $T+A$ interprets $T+B$. Fourth, we are not interested in for which particular A and B we have that $T+A$ interprets $T+B$, rather we are interested in those properties that are structural in that they hold for any choices of A and B. We opt for theorems interpretability in this thesis; i.e. for $T+A$ to interpret $T+B$ we require that the translation of any *theorem* of $T+B$ is provable in $T+A$. The theory T should be sufficiently strong; i.e. sequential. If such a theory is axiomatisable, it has a predicate $\mathsf{Int}_T(\cdot,\cdot)$, defined in a natural way, expressing the fact that the extension of T by the first argument of the predicate interprets the extension of T by the second argument.

The interpretability logic of T is defined in much the same way as the provability logic of T, but with an additional binary modal operator $\rhd$ whose intended interpretation is Int_T. Thus, the interpretability logic of a theory T is the set of all modal logical formulas in the modal interpretability logic language that are provable for whatever arithmetical reading is given to the propositional variables and taking the modal operators to their intended arithmetisations.

Unlike what might be the case with provability logics, the interpretability logic of T really depends on T. For example, the interpretability logic of Gödel–Bernays set theory (which is the logic denoted by $\mathbf{IL}\mathsf{P}$), and the interpretability logic of Peano Arithmetic (which is the logic denoted by $\mathbf{IL}\mathsf{M}$), differ.

Given a sequential theory T, there is a certain amount of content, usually denoted by $\mathbf{IL}(\text{All})$, the interpretability logic of T is bound to have. The exact contents of $\mathbf{IL}(\text{All})$ is not known; in fact, improving the lower bound is the question that motivates most of the research in the field. A simple lower bound is the *basic interpretability logic*, denoted by $\mathbf{IL}$. This is an extension of the provability logic $\mathbf{GL}$ (in sequential theories, provability of a formula A can be shown to be equivalent to stating that the negation of A interprets a contradiction) and contains five additional axioms schemas which in the literature are referred to as $\mathsf{J1}$–$\mathsf{J5}$. Let us present and give a short of description of each of these schemas. In the next section on preliminaries we will introduce a few conventions that will enable us to drop most of the parentheses used in this introduction.

The principle $\mathsf{J1}$ is $\Box(A \to B) \to (A \rhd B)$. This principle implies that if an extension $T+A$ is stronger (in the traditional sense) than another extension $T+B$, then $T+A$ interprets $T+B$. If this were not the case, clearly the notion of the interpretability we picked would have been too weak. However, the fact that this schema is provable means that not only does it express a true fact, but that the base theory T is in fact 'aware' that this schema holds. There are some true facts expressible in this language, most notably $\Box A \to A$, that are not provable, so one has to be careful even with simple properties.

The principle $\mathsf{J2}$ is $((A \rhd B) \wedge (B \rhd C)) \to (A \rhd C)$. Essentially this principle expresses the transitivity of interpretability, and is proved by composing particular interpretations witnessing the facts that $A \rhd B$ and $B \rhd C$.

The principle J3 is $((A \rhd C) \wedge (B \rhd C)) \to ((A \vee B) \rhd C)$. This principle reflects the fact that given two interpretations we can construct a third interpretation such that it translates formulas just like the first interpretation if a certain formula (A) holds, and just like the second interpretation otherwise.

The principle J4 is $(A \rhd B) \to (\Diamond A \to \Diamond B)$. This principle is a formalised version of the fact that interpretability implies relative consistency.

The principle J5 is $(\Diamond A) \rhd A$ and arithmetically, this is the most complex axiom schema of **IL**. Essentially this principle claims that from a model of the consistency of A we can extract a model of A itself. The interpretation is built with the help of the formalised Henkin construction ([67]).

Getting back to the question of **IL**(All), there is an interesting and surprising way of improving the best known lower bounds, i.e. of finding new arithmetically valid principles of interpretability. The approach is to study modal (Kripke-like) relational semantics for the logic **IL** and the extensions thereof. To every extension of **IL** we can associate a class of relational frames, the *characteristic class*. This is the class of those, and only those, frames which validate all the theorems of the extension in question (we will define frame validity later in the thesis). The distinguishing property satisfied exactly by the frames within the characteristic class is called the *frame condition*. New arithmetically valid principles of interpretability have been found by taking the frame conditions of already known principles, modifying them, and then obtaining the modal formula that characterises the modified frame condition. This does not, of course, guarantee the arithmetical validity of the thus obtained modal formula, but new arithmetically valid formulas have been found this way. If for some extension the validity of a formula on all the frames in the characteristic class implies its provability in the aforementioned extension, we speak of *modal completeness*. Another related approach of finding new principles of interpretability is to try and establish modal completeness of a certain extension of **IL**. If proving modal completeness fails for a given extension, extending the extension further, until it becomes modally complete, might yield new arithmetically valid formulas (this attempt will be followed in the final chapter of this thesis).

There are two related kinds of modal semantics used for interpretability logics. One is known as *regular Veltman semantics* (or *ordinary Veltman semantics*, or just *Veltman semantics* when there is no risk of ambiguity), introduced by Veltman. The other is *generalised Veltman semantics*, introduced by Verbrugge, which combines Kripke-like semantics with neighbourhood semantics. Regular Veltman semantics can be used to provide complete semantics to many interpretability logics. However, for more complex logics, generalised Veltman semantics can be used to provide simpler and much easier to understand proofs of completeness (this is one of the results of this thesis). In some cases, we only know that a logic is complete w.r.t. generalised Veltman semantics. There are also examples of logics complete w.r.t. generalised Veltman semantics, but incomplete w.r.t. regular Veltman semantics. All known complexity results (most of which are established in this thesis) rely on regular Veltman semantics. As for decidability, it seems that generalised Veltman semantics is a more appropriate tool, as it allows for a uniform method of obtaining the finite model property.

In the next chapter, Preliminaries, we give precise definitions of the concepts we use throughout the thesis. In Section 2.7 we give an overview of the thesis structure.

Chapter 2

Preliminaries

This chapter has a few different purposes. The first is to give a more formal introduction than the one provided in Chapter 1, including additional relevant historical references. The second purpose is to define the key notions that we use throughout the thesis, such as the principles of interpretability and Veltman semantics. Finally, this chapter contains some observations that did not fit other chapters.

2.1 Provability logics

This thesis is concerned with interpretability logics, a class of modal logics that extend the provability logic **GL**. Let us first say a few words on provability. We assume the reader has a certain amount of arithmetical background and refer for further details to [10].

Sufficiently strong formal theories T can reason about their own provability. The usual way to do this is through a certain Σ_1-predicate that formalises provability, usually denoted by Prov_T. For example, the following is provable in T:

$$\mathsf{Prov}_T\left(\overline{\ulcorner\neg\mathsf{Prov}_T\left(\overline{\ulcorner\bot\urcorner}\right)\urcorner}\right) \to \mathsf{Prov}_T\left(\overline{\ulcorner\bot\urcorner}\right),$$

that is, (the formalised version of) Gödel's second incompleteness theorem. Gödel noticed that provability can be viewed as a modal operator (this was briefly mentioned in [25]): if we let $\Box$ stand for Prov_T, Gödel's second incompleteness theorem can be expressed more succinctly:

$$\Box\neg\Box\bot \to \Box\bot.$$

In the modal language the uses of $\overline{\ulcorner\cdot\urcorner}$ are implicit. Examples of other properties of Prov_T expressible in a modal language are $\Box(A \to B) \to (\Box A \to \Box B)$ and $\Box A \to \Box\Box A$ (where A and B are arbitrary sentences).

The provability logic **GL** (Gödel, Löb) is a modal propositional logic with the single unary modal operator $\Box$. The axioms of the system **GL** are all propositional tautologies (in the new language), and all instances of the schemas K: $\Box(A \to B) \to (\Box A \to \Box B)$, and L: $\Box(\Box A \to A) \to \Box A$. The inference rules of **GL** are modus ponens and necessitation

$A/\Box A$.

Solovay [57] proved the arithmetical completeness theorem for **GL**. This theorem holds for all Σ_1-sound extensions of $\mathsf{I}\Delta_0+\mathsf{EXP}$, where EXP is the sentence formalising the totality of exponentiation (see [17]). This theorem shows that the language of provability logic **GL** is too weak to distinguish between most of the theories that are usually considered.[1] For example, whether a theory is finitely axiomatisable does not affect the theory's provability logic.

Other formal properties, beside provability, have been explored through modal or semi-modal systems; in particular, interpretability logics. All interpretability logics we consider are extensions, both in terms of their language and their theoremhood, of **GL**.

2.2 Interpretability logics

We consider the usual modal treatment of interpretability: interpretability logics. Let us briefly describe what is usually meant by "interpretability" in the context of interpretability logics.

Let T_1 and T_2 be some first-order theories of finite signatures σ_1 and σ_2, respectively. For convenience, we may further assume there are no constants and function symbols. An interpretation of T_2 in T_1 is a pair (f, U) where U is a formula in the language of T_1:

- f maps n-ary relational symbols $R \in \sigma_2$ to formulas with n free variables in the language of T_1;
- $f(A \to B) = f(A) \to f(B)$, similarly for other logical connectives;
- $f(\forall x F) = \forall x(U(x) \to f(F))$ and similarly for $(\exists x)F$;
- $T_1 \vdash (\exists x)U(x)$;
- for all sentences F in the language of T_2:

$$T_2 \vdash F \;\Rightarrow\; T_1 \vdash f(F).$$

The last requirement, that translations of the theorems of T_2 are provable in T_1, is sometimes modified. The variant we employ is known as *theorems interpretability*. See e.g. [37] and [68] for alternative possibilities, and further details concerning interpretations in general.

In a sufficiently strong formal theory T in the language $\mathcal{L}_T$, one can construct a binary interpretability predicate Int_T. This predicate expresses that an extension of T interprets another extension of T. Both the interpreting and the interpreted theory are assumed

[1] However, it is possible that the provability logics of theories below $\mathsf{I}\Delta_0 + \mathsf{EXP}$ differ from **GL** (see [3]).

to extend T with a finite number of sentences or, equivalently, with a single sentence. Interpretability was first explored in [61], in a non-modal setting.

Modal logics for interpretability were first studied by Hájek in 1981 [33] and Švejdar in 1983 [59]. Visser introduced the modal logic **IL** (interpretability logic), a modal logic with a binary modal operator $\rhd$ representing interpretability, in 1990 [66]. This operator is the only addition to the language of propositional logic; i.e. the language of interpretability logics is given by

$$A ::= \bot \mid p \mid A \to A \mid A \rhd A,$$

where p ranges over a countable set of propositional variables. Other Boolean connectives are defined as abbreviations, as usual. In particular, we let $\top$ abbreviate $\bot \to \bot$. Since $\Box B$ too can be defined (over **IL**) as an abbreviation (expanded to $\neg B \rhd \bot$), we do not formally include $\Box$ in the language. Similarly, we do not include $\Diamond$ in the language, where $\Diamond B$ stands for $\neg\Box\neg B$. We stress that we still wish to use $\Box$ and $\Diamond$ in our presentation, but they are to be understood as abbreviations. We treat $\rhd$ as having higher priority than $\to$, but lower than other logical connectives. For example, $A \rhd B \to \neg(A \rhd \neg C) \rhd \Box C \wedge B$ is to be understood as $(A \rhd B) \to (\neg(A \rhd \neg C) \rhd ((\Box C) \wedge B))$.

Let T be a sufficiently strong formal theory in the language of arithmetic. Any mapping $A \mapsto A^*$, with A a modal formula and $A^* \in \mathcal{L}_T$, such that:

- if p is a propositional variable, p^* is a sentence;
- it commutes with logical connectives;
- $\bot^*$ is $0 = 1$,
- $(A \rhd B)^* = \mathsf{Int}_T\left(\overline{\ulcorner A^* \urcorner}, \overline{\ulcorner B^* \urcorner}\right)$, where $\overline{\ulcorner X \urcorner}$ is the numeral of the Gödel number of X;

is called *an arithmetical realisation.*

The *interpretability logic of a theory* T, denoted by **IL**(T), is the set of all modal formulas A such that $T \vdash A^*$ for all arithmetical realisations. While there are open questions regarding interpretability logics of certain theories, it is known that they all extend the basic system **IL**. In fact, we know of much better lower bounds, but **IL** has simple well-behaved semantics and is traditionally taken as the starting building block.

Definition 2.1 The interpretability logic **IL** is axiomatised by the following axiom schemas:

- classical tautologies (in the new language);

(K) $\Box(A \to B) \to (\Box A \to \Box B)$;

(L) $\Box(\Box A \to A) \to \Box A$;

(J1) $\Box(A \to B) \to A \rhd B$;

(J2) $(A \rhd B) \wedge (B \rhd C) \to A \rhd C$;

(J3) $(A \rhd C) \wedge (B \rhd C) \to A \vee B \rhd C$;

(J4) $A \rhd B \to (\Diamond A \to \Diamond B)$;

(J5) $\Diamond A \rhd A$.

Rules of inference are modus ponens and necessitation $A/\Box A$.

We say that a modal formula A is *valid* in a formal theory T if $T \vdash A^*$ for every arithmetical realisation $*$. A modal theory S is arithmetically sound w.r.t. T if all its theorems are valid in T. The modal theory S is arithmetically complete w.r.t. T if it proves exactly those formulas that are valid in T. Sometimes we omit "arithmetically" from "arithmetically sound (complete)", if there is no risk of confusion with other notions of soundness and completeness. For the proof that the system **IL** is sound w.r.t. any reasonable formal theory, see [66].

The system **IL** is, unlike **GL**, arithmetically incomplete w.r.t. any reasonable theory. For example, **IL** does not prove all instances of $A \rhd B \to A \rhd B \wedge \Box\neg A$, which are all valid in every reasonable theory (see e.g. [66]). To achieve arithmetical completeness, we have to study extensions of the basic system **IL**. Extensions are built by adding new axiom schemas, the so-called principles of interpretability. Two principles and the corresponding extensions of **IL** are of particular interest because these extensions are *the interpretability logics* of many interesting theories.

Montagna's principle M: $A \rhd B \to A \wedge \Box C \rhd B \wedge \Box C$ is valid in theories proving full induction. We denote by $\mathbf{IL}\mathsf{M}$ the system obtained by adding all instances of the principle M to the system **IL** as new axioms. Berarducci [2] and Shavrukov [56] independently proved that $\mathbf{IL}(\mathrm{T}) = \mathbf{IL}\mathsf{M}$, if T is Σ_1-sound and proves full induction. The persistence principle P: $A \rhd B \to \Box(A \rhd B)$ is valid in finitely axiomatisable theories. Visser [66] proved the arithmetical completeness of $\mathbf{IL}\mathsf{P}$ w.r.t. any finitely axiomatisable Σ_1-sound theory containing $\mathsf{I}\Delta_0 + \mathsf{SUPEXP}$, where SUPEXP asserts the totality of superexponentiation: $n \mapsto 2^n_n$ where $2^n_0 = n$ and $2^n_{m+1} = 2^{(2^n_m)}$. Thus, the interpretability logic $\mathbf{IL}\mathsf{M}$ of first-order Peano Arithmetic differs from the interpretability logic $\mathbf{IL}\mathsf{P}$ of Gödel-Bernays set theory. It is still an open problem what is the interpretability logic of weaker theories like $\mathsf{I}\Delta_0 + \mathsf{EXP}$, $\mathsf{I}\Delta_0 + \Omega_1$ and PRA. For $\mathsf{I}\Delta_0 + \Omega_1$, this question depends on what the provability logic of $\mathsf{I}\Delta_0 + \Omega_1$ is, which in turn may depend on very hard problems in computational complexity—see [3]. For results regarding interpretability in PRA we refer to [12] and [34].

In particular, one can ask what is the set of principles valid in all reasonable theories.[2] This set is usually denoted by $\mathbf{IL}(\mathrm{All})$. Note that this does not mean that there has to be

[2]"Reasonable" usually means "an extension of S^1_2 or $\mathsf{I}\Delta_0 + \Omega_1$", or "as weak as possible under the condition that $\mathbf{IL}(\mathrm{All})$ remains elegantly axiomatisable, if any".

a theory T that attains **IL**(All) as its interpretability logic, i.e. **IL**(T) = **IL**(All). Clearly **IL** $\subseteq$ **IL**(All) $\subseteq$ **ILP** $\cap$ **ILM**. In fact, we know that both inclusions are proper. The ongoing search for **IL**(All) is a main motivation behind studying extensions of **IL** today. Studying modal properties of lower bounds of **IL**(All) turns out to be useful for finding new principles within **IL**(All). For example, the principle R (we will define this and other principles later) was discovered while trying to prove modal completeness of $\mathbf{IL}\mathsf{P_0W}$ [28]. See the most recent development [29] for an overview of the progress that has been made in the search for **IL**(All). In this thesis (Chapter 8) we define a new series of principles for which it is an open question whether it extends the currently best known lower bound of **IL**(All).

Definitions and further details regarding interpretability and interpretability logics in general can be found in e.g. [66].

2.3 Semantics for interpretability logics

The most commonly used semantics for the interpretability logic **IL** and its extensions is *Veltman semantics* (or *ordinary Veltman semantics*).

Definition 2.2 ([18], Definition 1.2) A Veltman frame $\mathfrak{F}$ is a structure $(W, R, \{S_w : w \in W\})$, where W is a non-empty set, R is a transitive and converse well-founded binary relation on W and for all $w \in W$ we have:

a) $S_w \subseteq R[w]^2$, where $R[w] = \{x \in W : wRx\}$;

b) S_w is reflexive on $R[w]$;

c) S_w is transitive;

d) if $wRuRv$ then uS_wv.

The standard logic of (formalised) provability, the logic **GL**, is complete w.r.t. the semantics based on the so-called **GL**-frames: pairs (W, R) where W is non-empty and R is a transitive and converse well-founded binary relation on W. In the context of **GL**-frames we have $w \Vdash \Box A$ if and only if: wRx implies $x \Vdash A$. The aforementioned completeness result concerning **GL** was first proved by Segerberg in [54]. All interpretability logics that we study here conservatively extend the logic of provability. So, it should not be surprising that (W, R) in the preceding definition is precisely a **GL**-frame. For reasons already explained earlier, we will usually work as if the symbol $\Box$ is not in the language.

A *Veltman model* is a quadruple $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$, where the first three components form a Veltman frame. The forcing relation $\Vdash$ is extended as usual in Boolean cases, and $w \Vdash A \rhd B$ holds if and only if for all u such that wRu and $u \Vdash A$ there exists v such that uS_wv and $v \Vdash B$.

In this thesis we will mainly use a different semantics, which we will refer to as *generalised Veltman semantics*. R. Verbrugge [62] defined this specific generalisation of Veltman semantics. The main purpose of its introduction, and until recently the only use, was to show that certain extensions of **IL** are independent, by Verbrugge [62], Švejdar [60], Visser [67], Vuković [72] and Goris and Joosten [28].

Definition 2.3 A generalised Veltman frame $\mathfrak{F}$ is a structure $(W, R, \{S_w : w \in W\})$, where W is a non-empty set, R is a transitive and converse well-founded binary relation on W and for all $w \in W$ we have:

a) $S_w \subseteq R[w] \times (\mathcal{P}(R[w]) \setminus \{\emptyset\})$;

b) S_w is quasi-reflexive: wRu implies $uS_w\{u\}$;

c) S_w is quasi-transitive: if uS_wV and vS_wZ_v for all $v \in V$, then $uS_w(\bigcup_{v\in V} Z_v)$;

d) if $wRuRv$, then $uS_w\{v\}$;

e) monotonicity: if uS_wV and $V \subseteq Z \subseteq R[w]$, then uS_wZ.

A *generalised Veltman model* is a quadruple $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$, where the first three components form a generalised Veltman frame. Now $w \Vdash A \rhd B$ holds if and only if for all u such that wRu and $u \Vdash A$ there exists V such that uS_wV and $V \Vdash B$. By $V \Vdash B$ we mean that $v \Vdash B$ for all $v \in V$.

Given an ordinary or a generalised model $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$, we write $\mathfrak{M} \Vdash A$ if $w \Vdash A$ for all $w \in W$. Similarly, given an ordinary or a generalised frame $\mathfrak{F} = (W, R, \{S_w : w \in W\})$, we write $\mathfrak{F} \Vdash A$ if $\mathfrak{M} \Vdash A$ for all models $\mathfrak{M}$ based on $\mathfrak{F}$ (i.e. where $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ for some $\Vdash$).

2.4 Variations of generalised semantics

In a recent collaboration with Jan Mas Rovira and Joost J. Joosten we established some basic results regarding generalised Veltman semantics and the possible variations thereof. See [38] or [44] for a longer discussion on possible choices for the condition of quasi-transitivity. In a sense we show that our choice (Definition 2.3) is the most general one (i.e. as little restrictive as possible). In this section we will only quote the main results, and for the full proofs of all results we recommend [44].

Let us give a list of quasi-transitivity conditions considered so far in the literature:

Nr.	Semantic requirement for quasi-transitivity	First mention
(1)	$uS_xY \Rightarrow \forall \{Y_y\}_{y\in Y}\Big((\forall\, y \in Y\ yS_xY_y) \Rightarrow \exists Z \subseteq \bigcup_{y\in Y} Y_y \ \&\ uS_xZ\Big)$	[38]
(2)	$uS_xY \Rightarrow \forall \{Y_y\}_{y\in Y}\Big((\forall\, y \in Y\ yS_xY_y) \Rightarrow uS_x \bigcup_{y\in Y} Y_y\Big)$	[62]
(3)	$uS_xY \Rightarrow \exists\, y \in Y\, \forall Y'(yS_xY' \Rightarrow \exists\, Y''{\subseteq}Y' \ \&\ uS_xY'')$	[38]
(4)	$uS_xY \Rightarrow \exists\, y \in Y\, \forall Y'(yS_xY' \Rightarrow uS_xY')$	[36]
(5)	$uS_xY \Rightarrow \forall\, y \in Y\, \forall Y'(yS_xY' \Rightarrow \exists\, Y''{\subseteq}Y' \ \&\ uS_xY'')$	[38]
(6)	$uS_xY \Rightarrow \forall\, y \in Y\, \forall Y'(yS_xY' \Rightarrow uS_xY')$	[62]
(7)	$uS_xY \Rightarrow \forall\, y \in Y\, \forall Y'(yS_xY' \ \&\ y \notin Y' \Rightarrow \exists\, Y''{\subseteq}Y'\ uS_xY'')$	[38]
(8)	$uS_xY \Rightarrow \forall\, y \in Y\, \forall Y'(yS_xY' \ \&\ y \notin Y' \Rightarrow uS_xY')$	[28]

Note that the monotonicity condition (Definition 2.3) does not affect the definition of truth, so we could work without it too. More importantly, we can also always perform the closure under monotonicity and end up with a model whose truth values are preserved.

Proposition 2.4 Let $\mathfrak{F} = (W, R, S)$ be a generalised Veltman frame with quasi-transitivity (i) for some $i \in \{1, \ldots, 8\}$. Let $\mathfrak{F}' = (W, R, S')$ where S' is the monotonic closure of S:

$$S' = \{(w, x, Y') : (w, x, Y) \in S, Y \subseteq Y' \subseteq R[w]\}.$$

Then $\mathfrak{F}'$ is a generalised Veltman frame satisfying quasi-transitivity Condition (2). Furthermore, let V be an arbitrary valuation and A an arbitrary formula. Let $\mathfrak{M} = (\mathfrak{F}, V)$ and $\mathfrak{M}' = (\mathfrak{F}', V)$. We have that for every world w:

$$\mathfrak{M}, w \Vdash A \text{ if and only if } \mathfrak{M}', w \Vdash A.$$

The preceding proposition tells us that Notion (2) of quasi-transitivity is, in a sense, the most general one: any formula satisfiable with another notion of quasi-transitivity (out of the notions mentioned here) must be satisfiable in some model in our selected notion of a model.

It is well known that any ordinary Veltman model corresponds to a generalised Veltman model, as the following proposition shows.

Proposition 2.5 Let $\mathfrak{M} = (W, R, S, V)$ be an ordinary Veltman model. Let

$$S' = \{(w, u, V) : (\exists v \in W) uS_wv \in V \subseteq R[w]\}.$$

Then $\mathfrak{M}' = (W, R, S', V)$ is a generalised Veltman model for every notion of quasi-transitivity (i) with $i \in \{1, \ldots, 8\}$. Furthermore, for every world w and formula A:

$$\mathfrak{M}, w \Vdash A \text{ if and only if } \mathfrak{M}', w \Vdash A.$$

See e.g. [70], [44] or [38] for a proof. The other direction, transforming a generalised model to an appropriate ordinary model, is harder, and here we state one transformation that works for some notions of quasi-transitivity. The first such transformation was constructed by Verbrugge in [62] (Theorem 2), and these two transformations can both be used (only) for the following notions of quasi-transitivity: (3), (4), (5), (6).

For a generalised frame $\mathfrak{F} = (W, R, S)$ we let $S_w[u] = \{V : uS_wV\}$. We collect this and similar conventions in Section 2.8.

Proposition 2.6 Let $\mathfrak{M} = (W, R, S, V)$ be a generalised Veltman model. Let

$$S' = \{(w, u, v) : (\exists V \in S_w[u])v \in V\}.$$

Then $\mathfrak{M}' = (W, R, S', V)$ is an ordinary Veltman model where the notion of quasi-transitivity is (3), (4), (5) or (6). Furthermore, for every world w and formula A:

$$\mathfrak{M}, w \Vdash A \text{ if and only if } \mathfrak{M}', w \Vdash A.$$

One might wonder what is the relation between the different notions of quasi-transitivity, apart from the fact that they can all be seen as strengthenings of (2).

Proposition 2.7 ([44]) Let $\mathfrak{F}$ be a generalised Veltman frame. Let M stand for the monotonicity condition. The following implications hold.[3]

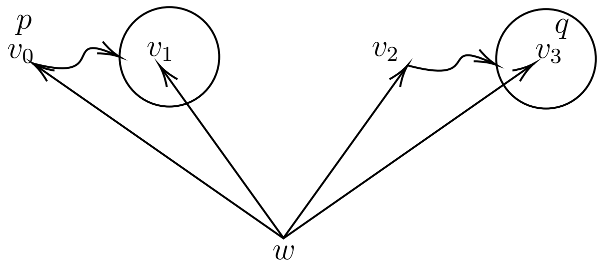
1. M & (1) $\Rightarrow$ (2)
2. (2) $\Rightarrow$ (1)
3. M & (3) $\Rightarrow$ (4)
4. (4) $\Rightarrow$ (3)
5. (5) $\Rightarrow$ (1)
6. M & (5) $\Rightarrow$ (2)
7. (5) $\Rightarrow$ (3)
8. M & (5) $\Rightarrow$ (4)
9. M & (5) $\Rightarrow$ (6)
10. (5) $\Rightarrow$ (7)
11. M & (5) $\Rightarrow$ (8)
12. (6) $\Rightarrow$ (1)
13. M & (6) $\Rightarrow$ (2)
14. (6) $\Rightarrow$ (3)
15. (6) $\Rightarrow$ (4)
16. (6) $\Rightarrow$ (5)
17. (6) $\Rightarrow$ (7)
18. (6) $\Rightarrow$ (8)
19. M & (7) $\Rightarrow$ (8)
20. (8) $\Rightarrow$ (7)

Remark 2.8 Suppose we are working with a definition of a Veltman frame which does not include monotonicity. As we have seen, closing this frame under monotonicity will not change truth values in any model based on this frame. This fact on its own does not imply we can safely assume to have monotonicity in the definition. In fact, requiring monotonicity *by definition* might change truth values. The problem lies in the fact that

[3] Note that M & $(i) \Rightarrow (j)$ means literally "if a frame satisfies monotonicity and quasi-transitivity in the sense (i), then it satisfies quasi-transitivity in the sense (j) too". This is very different from the situation described in Proposition 2.4 where we started with a frame that is quasi-transitive in the sense (i), closed it under monotonicity and obtained a frame that is quasi-transitive in the sense (j). See Remark 2.8. Apart from the results already discussed, we did not check what other 'implications' arise when performing a closure.

the closure under monotonicity might invalidate the selected notion of quasi-transitivity. If we now perform the closure under quasi-transitivity, this might change the truth values.

Consider the following frame where the selected notion of quasi-transitivity is (8): $wRv_0, wRv_1, wRv_2, wRv_3$, $v_0 S_w \{v_1\}$, $v_2 S_w \{v_3\}$.



Consider the model based on this frame where p is true exactly at v_0 and q is true exactly at v_3. Clearly $w \Vdash \neg(p \rhd q)$. Once we take the monotonic closure of S we get $v_0 S_w \{v_1, v_2\}$ and by the closure under quasi-transitivity (8) we get $v_0 S_w \{v_3\}$. Now $w \Vdash p \rhd q$, i.e. the truth values are not preserved.

In the remainder of the thesis we will always use Notion (2) of quasi-transitivity, and we will always assume monotonicity.

2.5 Extensions of IL

When we need to refer to an extension of **IL** by a single modal formula or a set of modal formulas X, we will write **IL**X.

Let (X) (resp. $(\mathsf{X})_{\text{gen}}$) denote a formula of first-order or higher-order logic such that for all ordinary (resp. generalised) Veltman frames $\mathfrak{F}$ the following holds:

$$\mathfrak{F} \Vdash \mathsf{X} \text{ if and only if } \mathfrak{F} \models (\mathsf{X}) \text{ (resp. } \mathfrak{F} \models (\mathsf{X})_{\text{gen}}).$$

The formulas (X) and $(\mathsf{X})_{\text{gen}}$ are called the characteristic properties (or frame conditions) of the given logic **IL**X. The class of all ordinary (resp. generalised) Veltman frames $\mathfrak{F}$ such that $\mathfrak{F} \models (\mathsf{X})$ (resp. $\mathfrak{F} \models (\mathsf{X})_{\text{gen}}$) is called the characteristic class of (resp. generalised) frames for **IL**X. If $\mathfrak{F} \models (\mathsf{X})_{\text{gen}}$ we also say that the frame $\mathfrak{F}$ possesses the property $(\mathsf{X})_{\text{gen}}$. We say that an ordinary (resp. generalised) Veltman model $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ is an **IL**X-model (resp. $\mathbf{IL}_{\text{set}}\mathsf{X}$-model), or that model $\mathfrak{M}$ possesses the property (X) (resp. $(\mathsf{X})_{\text{gen}}$), if the frame $(W, R, \{S_w : w \in W\})$ possesses the property (X) (resp. $(\mathsf{X})_{\text{gen}}$). A logic **IL**$\mathsf{X}$ will be said to be complete with respect to ordinary (resp. generalised) semantics if for all modal formulas A we have that validity of A over all **IL**X-frames (resp. all $\mathbf{IL}_{\text{set}}\mathsf{X}$-frames) implies $\mathbf{IL}\mathsf{X} \vdash A$.

We say that **IL**X has the finite model property (FMP) w.r.t. ordinary (resp. generalised) semantics if for each formula A satisfiable in some **IL**X-model (resp. $\mathbf{IL}_{\text{set}}\mathsf{X}$-model), A is also satisfiable in some finite **IL**X-model (resp. $\mathbf{IL}_{\text{set}}\mathsf{X}$-model).

The following table displays the current state of research regarding the principles we discuss throughout the thesis. Here, o stands for ordinary Veltman semantics, and g for generalised Veltman semantics. When a logic is complete w.r.t. ordinary semantics, it is also complete w.r.t. generalised semantics (see Proposition 2.6), and similarly for the FMP. For the results implied by this fact we do not cite any particular source in the table below.

	principle	compl. (o)	compl. (g)	FMP (o)	FMP (g)
M	$A \rhd B \to A \wedge \Box C \rhd B \wedge \Box C$	+ [18]	+	+ [18]	+
M_0	$A \rhd B \to \Diamond A \wedge \Box C \rhd B \wedge \Box C$	+ [27]	+	?	+ [53]
P	$A \rhd B \to \Box(A \rhd B)$	+ [18]	+	+ [18]	+
P_0	$A \rhd \Diamond B \to \Box(A \rhd B)$	− [28]	+ [50]	?	+ [50]
R	$A \rhd B \to \neg(A \rhd \neg C) \rhd B \wedge \Box C$	?	+ [50]	?	+ [50]
W	$A \rhd B \to A \rhd B \wedge \Box\neg A$	+ [19]	+	+ [19]	+
F	$A \rhd \Diamond A \to \Box\neg A$	− [62]	?	+ [19]	+
W^*	$A \rhd B \to B \wedge \Box C \rhd B \wedge \Box C \wedge \Box\neg A$	+ [27]	+	?	+ [49]

De Jongh and Veltman proved the completeness of the logics **IL**, **IL**M and **IL**P w.r.t. their characteristic classes of ordinary (and finite) Veltman frames in [18]. Goris and Joosten [27, 28] proved the completeness of **IL**M_0 and **IL**W^* w.r.t. ordinary semantics. Mikec and Vuković [50] proved completeness of **IL**R and **IL**P_0 w.r.t. generalised Veltman semantics. A more thorough introduction concerning each topic (completeness, decidability, completeness) will be given in the relevant chapter of this thesis.

In addition to these principles, we also discuss three series of principles, namely (R_n), (R^n) and (W_n). Their definitions are a bit lengthy, so we will postpone defining them until we discuss them (in the final two chapters of the thesis).

The two series (R_n) and (R^n) have been introduced recently [29] and not much is known regarding their semantics. See Chapter 7 for the definition. Their frame conditions w.r.t. ordinary Veltman semantics are defined in [29]. The corresponding conditions w.r.t. generalised Veltman semantics have been explored by Jan Mas Rovira, Joost J. Joosten, and the author. The preliminary results were presented in [45] and proofs with more detail are available in [44]. We quote these results (without proofs) in Chapter 7, as we define these two series only then. Note that this work on the two series concerns just the frame conditions; the work on completeness has not started yet. Similarly, the problem of the finite model property is open for both series.

The third series (W_n) is introduced in the final chapter of this thesis. Some basic results were already discussed in [46]. Their frame conditions w.r.t. ordinary Veltman semantics coincide with the conjunction of $(\mathsf{W})_{\text{gen}}$ and $(\mathsf{R})_{\text{gen}}$. Their generalised semantics is discussed in Chapter 8.

Let us also mention that recently Kurahashi and Okawa obtained results concerning certain natural sublogics of **IL**, specifically the logics between **IL** and a logic called **IL**$^-$

which is the logic of the class of structures called *Veltman prestructures* [42]. The authors prove that twelve of these logics are complete w.r.t. their characteristic classes of Veltman prestructures, and that further eight are incomplete in this sense, yet complete w.r.t. their characteristic classes of a generalised form of Veltman prestructures. All twenty logics they study are proven complete w.r.t. finite structures, and thus decidable.

In the next section we look at the frame conditions (X) and $(\mathsf{X})_{\text{gen}}$ of the aforementioned logics and make some additional observations. We mainly focus on the principle W whose semantics is more contrived than is the case with other principles.

2.6 Frame conditions

We start this section with some simple observations regarding Veltman semantics. For a given ordinary or generalised Veltman model $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$, a world $w \in W$ and a formula A we define:

$$[A]_w = \{x : wRx \;\&\; x \Vdash A\}.$$

Definition 2.9 Let $\mathcal{R}$ be a binary relation and V an arbitrary set. We say that w is $\mathcal{R}$-maximal in V if $w \in V$ and for all x such that $w\mathcal{R}x$ we have $x \notin V$.

Let $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ be an ordinary or a generalised Veltman model. Let $V \subseteq W$ be an arbitrary non-empty set of worlds. Clearly, there has to be an R-maximal v in V (due to the converse well-foundedness of R). We often use this fact. In particular, we often use the following lemma.

Lemma 2.10 Let $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ be an ordinary or a generalised Veltman model. If $\mathfrak{M}, w \Vdash \neg(A \rhd B)$ then there is a world x that is R-maximal in $[A]_w$ (i.e. $x \Vdash A$ and there are no worlds z such that xRz and $z \Vdash A$) and for all y: if xS_wy then $y \nVdash B$.[4]

Proof. Let $V = \{x \in R[w] : x \Vdash A \;\&\; (\forall y)(xS_wy \Rightarrow y \nVdash B)\}$. Then V is non-empty and there is a maximal $x \in V$. It remains to see x is R-maximal in $[A]_w$. Assume for a contradiction there is z such that xRz and $z \Vdash A$. Note that for all y we have that zS_wy implies xS_wy. Thus, for all y we have that zS_wy implies $y \nVdash B$ (otherwise there is y with $xS_wy \Vdash B$, contrary to the definition of V). However, this contradicts the definition of x; i.e. x is not maximal in V. ⊣

Recall that for a principle X, we denote by (X) and $(\mathsf{X})_{\text{gen}}$ the properties of Veltman frames, or generalised Veltman frames, respectively, such that the following holds: X is

[4]The variable y in the statement of this corollary quantifies over worlds in the context of ordinary Veltman semantics, and otherwise y quantifies over sets of worlds.

valid on a (generalised) Veltman frame $\mathfrak{F}$ if and only if $\mathfrak{F}$ has the property (X), or $(\mathsf{X})_{\text{gen}}$, respectively.

In the following table we summarise the frame conditions of some principles. In particular, we tried to include all the principles that are well known, have unusual properties, or are important in the search for $\mathbf{IL}(\mathsf{All})$. Note that the notion of generalised Veltman semantics that was mainly used in [62] substantially differs from the one used nowadays (which was also first defined in [62], but explored in more detail only later). However, the definition of truth has the same form in both cases, the characteristic properties are the same, and the proofs of characterisation are similar. Thus [62] is cited as the first proof of $(\mathsf{M})_{\text{gen}}$ and $(\mathsf{P})_{\text{gen}}$ being the characteristic properties for $\mathbf{IL}\mathsf{M}$ and $\mathbf{IL}\mathsf{P}$. That these are the characteristic properties with respect to the other notion was verified in [72].

	Property	cf.
(M)	$uS_wvRz \Rightarrow uRz$	[18]
$(\mathsf{M})_{\text{gen}}$	$uS_wV \Rightarrow (\exists V' \subseteq V)(uS_wV' \ \&\ R[V'] \subseteq R[u])$	[62]
$(\mathsf{M_0})$	$wRuRxS_wvRz \Rightarrow uRz$	[27]
$(\mathsf{M_0})_{\text{gen}}$	$wRuRxS_wV \Rightarrow (\exists V' \subseteq V)(uS_wV' \ \&\ R[V'] \subseteq R[u])$	[71]
(P)	$wRw'RuS_wv \Rightarrow uS_{w'}v$	[18]
$(\mathsf{P})_{\text{gen}}$	$wRw'RuS_wV \Rightarrow (\exists V' \subseteq V)(uS_{w'}V')$	[62]
(W)	$S_w \circ R$ is converse well-founded for each w	[19]
$(\mathsf{W})_{\text{gen}}$	$uS_wV \Rightarrow (\exists U' \subseteq U)(uS_wU' \ \&\ R[U'] \cap S_w^{-1}[U] = \emptyset)$	[28, 49]
(F)	(W)	[65]
$(\mathsf{F})_{\text{gen}}$	See Subsection 2.6.2	[70]
(W^*)	$(\mathsf{M_0})$ and (W)	
$(\mathsf{W}^*)_{\text{gen}}$	$(\mathsf{M_0})_{\text{gen}}$ and $(\mathsf{W})_{\text{gen}}$	
$(\mathsf{P_0})$	$wRuRxS_wvRz \Rightarrow xS_uz$	[27]
$(\mathsf{P_0})_{\text{gen}}$	$wRxRuS_wV \ \&\ (\forall v \in V)R[v] \cap Z \neq \emptyset \ \Rightarrow\ (\exists Z' \subseteq Z)uS_xZ'$	[28]
(R)	$(\mathsf{P_0})$	[28]
$(\mathsf{R})_{\text{gen}}$	See below the table or Section	[28]
$(\mathsf{R_n})$	See [29]	[29]
$(\mathsf{R_n})_{\text{gen}}$	See Chapter 7 for $\mathsf{R_1}$	[44, 45]
$(\mathsf{R^n})$	See [29]	[29]
$(\mathsf{R^n})_{\text{gen}}$	See Chapter 7	[44, 45]
$(\mathsf{W_n})$	See Chapter 8	[46], here
$(\mathsf{W_n})_{\text{gen}}$	See Chapter 8 for discussion	[46], here

A frame condition for W w.r.t. generalised Veltman semantics was first given in [28], and the condition given in the table is from [49].

The condition $(\mathsf{R})_{\text{gen}}$ is the following:

$$wRxRuS_wV \Rightarrow (\forall C \in \mathcal{C}(x,u))(\exists U \subseteq V)(xS_wU \ \&\ R[U] \subseteq C),$$

where $\mathcal{C}(x,u) = \{C \subseteq R[x] : (\forall Z)(uS_xZ \Rightarrow Z \cap C \neq \emptyset)\}$ is the family of 'choice sets'.

As a demonstration, let us look at the frame condition $(\mathsf{M})_{\text{gen}}$.

Proposition 2.11 ([62]) The principle M is valid on a frame if and only if the frame condition $(\mathsf{M})_{\text{gen}}$ holds:

$$uS_wV \Rightarrow (\exists V' \subseteq V)(uS_wV' \ \&\ R[V'] \subseteq R[u]).$$

Proof. Let $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ be a generalised Veltman model that satisfies the given condition. Let $w, u \in W$ and suppose $w \Vdash A \rhd B$, and $wRu \Vdash A \wedge \Box C$. Then there is V with $uS_wV \Vdash B$. The condition implies there is $V' \subseteq V$ with uS_wV' and $R[V'] \subseteq R[u]$. Since $V' \subseteq V \Vdash B$, clearly $V' \Vdash B$. Since $R[u] \Vdash C$, we have $R[V'] \Vdash C$, and thus $V' \Vdash \Box C$.

In the other direction, suppose the principle M is valid on some frame $(W, R, \{S_w : w \in W\})$, and uS_wV. Define the forcing relation so that $[p]_w = \{u\}$, $[q]_w = V$, and $[r]_w = R[u]$. Since $w \Vdash p \rhd q \to p \wedge \Box r \rhd q \wedge \Box r$, and clearly $w \Vdash p \rhd q$, we get $w \Vdash p \wedge \Box r \rhd q \wedge \Box r$. Since $u \Vdash p \wedge \Box r$, there must be a set V' with $uS_wV' \Vdash q \wedge \Box r$. As $[q]_w = V$, we have $V' \subseteq V$. As $[r]_w = R[u]$, we must have $R[V'] \subseteq R[u]$. $\dashv$

2.6.1 Characteristic classes for ILW and ILW*

The first formulation of $(\mathsf{W})_{\text{gen}}$ was published in [28]. Unfortunately this formulation was overlooked and a different (but, luckily, quite a bit shorter) formulation was published in [49]. We discuss the second formulation here.

Definition 2.12 ([49]) A generalised Veltman frame $\mathfrak{F} = (W, R, \{S_w : w \in W\})$ has the property $(\mathsf{W})_{\text{gen}}$ if the following holds:

$$uS_wV \Rightarrow (\exists V' \subseteq V)\Big(uS_wV' \ \&\ R\,[V'] \cap S_w^{-1}\,[V] = \emptyset\Big).$$

As is customary, we omit the implicit universal quantifiers: $(\forall w, u \in W)(\forall V \subseteq W)$. A generalised Veltman model $(\mathfrak{F}, \Vdash)$ has the property $(\mathsf{W})_{\text{gen}}$ if the frame $\mathfrak{F}$ has the property $(\mathsf{W})_{\text{gen}}$.

Since we will often use the negation of the property $(\mathsf{W})_{\text{gen}}$, we will denote this negation as $\overline{(\mathsf{W})_{\text{gen}}}$. Obviously:

$$\overline{(\mathsf{W})_{\text{gen}}} \Leftrightarrow \begin{cases} (\exists w, u \in W)(\exists V \subseteq W) \\ \Big(uS_wV \ \&\ (\forall V' \subseteq V)\Big(uS_wV' \ \Rightarrow \ (R\,[V'] \cap S_w^{-1}\,[V] \neq \emptyset)\Big)\Big) \end{cases}$$

Remark 2.13 ([49]) It is known (see table in the previous section) that the principle W is valid on an ordinary Veltman frame $\mathfrak{F}$ if and only if $\mathfrak{F}$ has the property that for each world w, the relation $S_w \circ R$ is converse well-founded.

It is often the case that for a given principle of interpretability X, the corresponding properties (X) and $(\mathsf{X})_{\text{gen}}$ can be formulated in such a way that they syntactically resemble each other (some examples of this can be seen in the table in the previous section).

It is easy to check that the negation of the property (W) is equivalent to the following property:

$$(\exists w, u \in W)(\exists V \subseteq R[w], V \neq \emptyset)(\forall v \in V)(uS_wv \ \&\ R[v] \cap S_w^{-1}[V] \neq \emptyset).$$

Here, $S_w^{-1}[V] = \{y \mid yS_wz \text{ for some } z \in V\}$. Indeed, if this property holds, there is an infinite $R \circ S_w$ chain starting with u and with worlds alternating between the sets V and $S_w^{-1}[V]$. On the other hand, if there is an infinite $R \circ S_w$ chain $u_1S_wv_1Ru_2S_wv_2R\ldots$, take $V = \{v_1, \ldots\}$ and $u = u_1$.

Lemma 2.14 ([49]) Let $\mathfrak{F} = (W, R, \{S_w : w \in W\})$ be a generalised Veltman frame. Then the principle W is valid on a frame $\mathfrak{F}$ if and only if $\mathfrak{F}$ has the property $(\mathsf{W})_{\text{gen}}$.

Proof. Assume that the principle W is not valid on the frame $\mathfrak{F}$. Then there exists a forcing relation $\Vdash$ on the frame $\mathfrak{F}$, a world $w \in W$, and some formulas A and B such that $w \nVdash A \rhd B \to A \rhd B \wedge \Box\neg A$. Thus:

$$(1) \quad w \Vdash A \rhd B;$$

$$(2) \quad w \nVdash A \rhd B \wedge \Box\neg A.$$

From (2) it follows that there exists a world $u \in R[w]$ such that $u \Vdash A$, and

$$(3) \qquad (\forall V' \subseteq R[w])(uS_wV' \Rightarrow V' \nVdash B \wedge \Box\neg A).$$

Let $V = [B]_w$.

Since $u \Vdash A$, the fact labelled by (1) implies, together with the monotonicity, that uS_wV.

We now prove the following:

$$(\forall V' \subseteq V)\Big(uS_wV' \Rightarrow R[V'] \cap S_w^{-1}[V] \neq \emptyset\Big).$$

Consider any $V' \subseteq V$ such that uS_wV'. Now $V \Vdash B$ and the fact labelled by (3) imply that there exists a world $v \in V'$ such that $v \nVdash \Box\neg A$, i.e. $v \Vdash \Diamond A$. Thus, there is some z such that vRz and $z \Vdash A$. But now the fact (1) and monotonicity imply $zS_w[B]_w$, i.e. zS_wV. So, $z \in R[V'] \cap S_w^{-1}[V]$, i.e. $R[V'] \cap S_w^{-1}[V] \neq \emptyset$. We have proved that the frame $\mathfrak{F}$ does not have the property $(\mathsf{W})_{\text{gen}}$.

We now prove the other implication from the statement of the lemma. Suppose that the frame $\mathfrak{F}$ does not have the property $(\mathsf{W})_{\text{gen}}$, i.e. $\mathfrak{F}$ has the property $\overline{(\mathsf{W})_{\text{gen}}}$. Then there exist worlds $w, u \in W$ and a set $V \subseteq W$ such that uS_wV and the following holds:

$$(4) \qquad (\forall V' \subseteq V)\Big(uS_wV' \Rightarrow R\,[V'] \cap S_w^{-1}\,[V] \neq \emptyset\Big).$$

We define forcing relation $\Vdash$ on $\mathfrak{F}$ so that the following holds:

$$\begin{aligned} x \Vdash p &\quad\Leftrightarrow\quad xS_wV, \\ x \Vdash q &\quad\Leftrightarrow\quad x \in V. \end{aligned}$$

We claim that $w \nVdash p \rhd q \to p \rhd q \wedge \Box\neg p$. It is easy to check that $w \Vdash p \rhd q$. It remains to show that $w \nVdash p \rhd q \wedge \Box\neg p$. Assume the contrary, i.e. that $w \Vdash p \rhd q \wedge \Box\neg p$. Now $z \Vdash p$ and the assumption $w \Vdash p \rhd q \wedge \Box\neg p$ imply that there exists a set $V' \subseteq R\,[w]$ such that uS_wV' and $V' \Vdash q \wedge \Box\neg p$. In particular, we have $V \Vdash q$, thus the definition of $\Vdash$ implies that $V' \subseteq V$. Now the fact labelled with (4) implies there is $v \in V'$ and some $z \in W$ such that vRz and zS_wV. But this implies $z \Vdash p$, and so $v \nVdash \Box\neg p$. This contradicts $v \in V$ and $V \Vdash \Box\neg p$. Hence, the assumption $w \Vdash p \rhd q \wedge \Box\neg p$ leads to a contradiction. Thus, the principle W is not valid on the frame $\mathfrak{F}$. $\dashv$

We will use (this formulation of) $(\mathsf{W})_{\text{gen}}$ in what follows. We note here that the $(\mathsf{W})_{\text{gen}}$ condition can be formulated in a more informative way.

Given a generalised frame $\mathfrak{F} = (W, R, \{S_w : w \in W\})$, we say that (w, u, V) is a counterexample in $\mathfrak{F}$ to $(\mathsf{W})_{\text{gen}}$ if

$$uS_wV \text{ and } (\forall V' \subseteq V)\Big(uS_wV' \;\Rightarrow\; (R\,[V'] \cap S_w^{-1}\,[V] \neq \emptyset)\Big).$$

Proposition 2.15 Let $\mathfrak{F} = (W, R, \{S_w : w \in W\})$ be a generalised Veltman frame. Whenever there are w, u and V such that (w, u, V) is a counterexample in $\mathfrak{F}$ to $(\mathsf{W})_{\text{gen}}$, there is $U \subseteq V$ such that:

1. (w, u, U) is a counterexample in $\mathfrak{F}$ to $(\mathsf{W})_{\text{gen}}$;
2. $R[U] \cap U = \emptyset$;
3. given the set $\overline{U} = \{v \in U : R[v] \cap S_w^{-1}[U] = \emptyset\}$ we have the following:
   (a) $U \setminus \overline{U} \neq \emptyset$;
   (b) for all $v \in U$ we have $(\forall U' \subseteq U)(\forall z)\Big(vRzS_wU' \;\Rightarrow\; U' \cap U \setminus \overline{U} \neq \emptyset\Big)$.

Proof. Let us first show that we can find U with Properties (1) and (2).

For all $v \in V$ we define a world u_v. Fix $v \in V$. If there exists $z \in R[v] \cap V$, let u_v be any such R-maximal z. Otherwise, i.e. if $R[v] \cap V = \emptyset$, let $u_v = v$. Put $U = \{u_v : v \in V\}$.

Quasi-transitivity implies uS_wU, and U has the required property that $U \cap R[U] = \emptyset$. It remains to prove Property (1), i.e. that for an arbitrary $U' \subseteq U$ with uS_wU', the set $R[U'] \cap S_w^{-1}[U]$ is non-empty. Since $U' \subseteq U \subseteq V$ and (w, u, V) is a counterexample to $(\mathsf{W})_{\mathrm{gen}}$, we have that the set $R[U'] \cap S_w^{-1}[V]$ is non-empty. Thus, there is some $x \in R[U']$ such that xS_wV. Since for each $v \in V$ we have $vS_w\{u_v\}$, by quasi-transitivity we have xS_wU. The set $R[U'] \cap S_w^{-1}[U]$ contains the world x and hence is non-empty. Thus, we have verified Properties (1) and (2).

We can assume that V already has Properties (1) and (2) and we will find $U \subseteq V$ that satisfies all three properties (1), (2), and (3).

Let

$$\begin{aligned} U_1 &= \{v \in V : R[v] \cap S_w^{-1}[V] \neq \emptyset\}; \\ U_2 &= \{v \in U_1 : (\forall V' \subseteq V)(\forall z)(vRzS_wV' \Rightarrow\ V' \cap U_1 \neq \emptyset)\}; \\ U &= (V \setminus U_1) \cup U_2; \\ U_3 &= \{v \in U_1 : (\forall U' \subseteq U)(\forall z)(vRzS_wU'\ \Rightarrow\ U' \cap U_2 \neq \emptyset)\}. \end{aligned}$$

Clearly U_1 is non-empty.

Before proving that we have the required properties (1), (2), and (3), let us show that $U_2 = U_3$, prove an auxiliary claim, and show that $U_2 \subseteq U \setminus \overline{U}$. Fix $v \in U_2$, $U' \subseteq U$ and z such that $vRzS_wU'$. Since $U \subseteq V$ and $v \in U_2$, we have $U' \cap U_1 \neq \emptyset$. Since $U' \subseteq U$ and $U = (V \setminus U_1) \cup U_2$, the only part of U_1 that U' could be intersecting must also be a part of U_2. Thus, $U_2 \subseteq U_3$. To show the other direction, fix $v \in U_3$, $V' \subseteq V$ and z such that $vRzS_wV'$. There are two possibilities. One is that $V' \subseteq U$. In this case the definition of U_3 implies $V' \cap U_2 \neq \emptyset$, so, since $U_2 \subseteq U_1$, we have $V' \cap U_1 \neq \emptyset$, as required. The other possibility is that $V' \cap (V \setminus U) \neq \emptyset$, i.e. $V' \cap (U_1 \setminus U_2) \neq \emptyset$. Again, $V' \cap U_1 \neq \emptyset$. Since in both cases we have $V' \cap U_1 \neq \emptyset$, we conclude $v \in U_2$. Thus, we have proven that $U_2 = U_3$.

Next we prove the following auxiliary claim:

$$\text{for all } x \in W,\ xS_wV \text{ implies } xS_wU. \tag{2.1}$$

To prove this claim we first define a set V_v for every $v \in V$. If $v \in (V \setminus U_1) \cup U_2$, let $V_v = \{v\}$. Otherwise, i.e. if $v \in U_1 \setminus U_2$, there are V' and z such that $vRzS_wV' \subseteq V \setminus U_1$. Let $V_v = V'$. Note that for all $v \in V$ we have $V_v \subseteq U$, so for $V_0 := \bigcup_{v \in V} V_v$ we have $V_0 \subseteq U$. If x is such that xS_wV then xS_wV_0 by quasi-transitivity, and xS_wU by monotonicity. This concludes the proof of the auxiliary claim (2.1).

Finally, let us show that $U_2 \subseteq U \setminus \overline{U}$. Let $v \in U_2$, clearly $v \in U$. Suppose $v \in \overline{U}$. Since $v \in U_2 \subseteq U_1$, there is z such that $vRzS_wV$. Then $vRzS_wU$ by (2.1). So, $R[v] \cap S_w^{-1}[U] \neq \emptyset$, and $v \notin \overline{U}$.

Now we are ready to prove (1), (2) and (3) for the triple (w, u, U).

By (2.1) we have uS_wU, so to see that (1) holds let $U' \subseteq U$ be such that uS_wU'. Since Property (1) holds for the triple (w, u, V) and $uS_wU' \subseteq V$, there must be $v \in U'$ such that for some z we have $vRzS_wV$. Therefore, $vRzS_wU$ by (2.1). Thus, $R[U'] \cap S_w^{-1}[U] \neq \emptyset$. This concludes the proof of Property (1) for the set U.

Since the world v from the preceding argument is contained in $U \setminus \overline{U}$, we also proved Property (3a).

Property (2) for (w, u, U) holds because it holds for (w, u, V) and $U \subseteq V$.

It remains to verify Property (3b).

Suppose $U \ni vRzS_wU' \subseteq U$ and $U' \cap U \setminus \overline{U} = \emptyset$ for a contradiction. Since $v \in U$, either $v \in V \setminus U_1$ or $v \in U_2$. In the first case we would have $R[v] \cap S_w^{-1}[V] = \emptyset$, which is not the case ($vRzS_wU' \subseteq U \subseteq V$ and apply monotonicity). So, $v \in U_2 = U_3$, and by the definition of U_3, $U' \cap U_2 \neq \emptyset$. Since $U_2 \subseteq U \setminus \overline{U}$ as we showed earlier, $U' \cap U \setminus \overline{U} \neq \emptyset$, a contradiction. ⊣

This new formulation tells us that we can pick a set U and a quasi-partition $\{\overline{U}, U \setminus \overline{U}\}$ of U such that points in $\overline{U}$ cannot 'return' (via $S_w \circ R$) to U, while the points in $U \setminus \overline{U}$ can 'return' to U, and have an additional property that every set these points 'return' to intersects (not only U but also) $U \setminus \overline{U}$.

Goris and Joosten emphasised in [27] that the properties (W) and ($\mathsf{M_0}$) determine the characteristic class of Veltman frames for the logic **IL**W* because we have **IL**W* =**IL**W$\mathsf{M_0}$ (see [68]). In the following corollary we conclude the completely analogous fact w.r.t. generalised semantics.

Corollary 2.16 ([49]) For any generalised Veltman frame $\mathfrak{F}$ we have that schema W* is valid on $\mathfrak{F}$ if and only if both conditions $(\mathsf{M_0})_{\text{gen}}$ and $(\mathsf{W})_{\text{gen}}$ hold.

2.6.2 On ILF

The logic **IL**F is an example of a logic incomplete with respect to ordinary Veltman semantics. Even though the characteristic classes (w.r.t. ordinary semantics) of **IL**W and **IL**F are the same (see [65]), the logics themselves are not. A straightforward way to see this is to look at the frame conditions w.r.t. generalised Veltman semantics. In [62] a generalised Veltman model $\mathfrak{M}$ is defined such that $\mathfrak{M} \Vdash \mathsf{F}$ but $\mathfrak{M} \nVdash \mathsf{W}$. Hence **IL**F $\nvdash$ W.

The condition $(\mathsf{F})_{\text{gen}}$ (see [70]) is that $\overline{S_w} \circ \overline{R_w}$ is converse well-founded for each w, where the relations $\overline{S_w}$ and $\overline{R_w}$ are defined as follows:

- for any $A \in 2^{R[w]} \setminus \{\emptyset\}$ and $\mathcal{B} \subseteq 2^{R[w]} \setminus \{\emptyset\}$ we define $A\overline{S_w}\mathcal{B}$ if and only if $(\forall a \in A)(\exists B \in \mathcal{B})(aS_wB)$,
- for any $\mathcal{C} \subseteq 2^{R[w]} \setminus \{\emptyset\}$ and $D \in 2^{R[w]} \setminus \{\emptyset\}$ we define $\mathcal{C}\overline{R}D$ if and only if $(\forall C \in \mathcal{C})(\forall c \in C)(\exists d \in D)(cRd)$.

Proposition 2.17 ([49]) If a generalised Veltman frame $\mathfrak{F} = (W, R, \{S_w : w \in W\})$ possesses the property $(\mathsf{W})_{\text{gen}}$ then the frame $\mathfrak{F}$ also possesses the property $(\mathsf{F})_{\text{gen}}$.

Proof. Suppose for a contradiction that a generalised Veltman frame $\mathfrak{F}$ that possesses the property $(\mathsf{W})_{\text{gen}}$ does not have the property $(\mathsf{F})_{\text{gen}}$.

Then there exists a world $w \in W$ and sequences of sets (A_n) and $(\mathcal{B}_n)$ such that

$$A_0\overline{S_w}\mathcal{B}_0\overline{R_w}A_1\overline{S_w}\mathcal{B}_1\dots$$

Let $V = \bigcup_n(\bigcup\mathcal{B}_n)$. Let $u \in \bigcup_n A_n$ be an arbitrary world. Then there exists $n \in \omega$ such that $u \in A_n$. The fact $A_n\overline{S_w}\mathcal{B}_n$ implies that there exists $U \in \mathcal{B}_n$ such that uS_wU. Since $U \subseteq V \subseteq R[w]$, by monotonicity we have uS_wV. We claim the following:

$$(\forall V' \subseteq V)\Big(uS_wV' \Rightarrow R[V'] \cap S_w^{-1}[V] \neq \emptyset\Big).$$

Let $V' \subseteq V$ be a set such that uS_wV' and fix an arbitrary $v \in V'$. There is a number $m \in \omega$ such that $v \in \bigcup\mathcal{B}_m$. Since $\mathcal{B}_m\overline{R_w}A_{m+1}$, there is a world $z \in A_{m+1}$ such that vRz. Obviously we have zS_wZ for some $Z \in B_{m+1}$. By monotonicity, zS_wV.

So, we have proved that the frame $\mathfrak{F}$ possesses the property $(\mathsf{W})_{\text{gen}}$. $\dashv$

We now give an alternative proof that $\mathbf{IL}\mathsf{F} \nvdash \mathsf{W}$, by using only the conditions $(\mathsf{F})_{\text{gen}}$ and $(\mathsf{W})_{\text{gen}}$.

Corollary 2.18 ([49]) The conditions $(\mathsf{W})_{\text{gen}}$ and $(\mathsf{F})_{\text{gen}}$ are not equivalent. So, the principle W is not provable in $\mathbf{IL}\mathsf{F}$.

Proof. Let $\mathfrak{F}$ be the smallest generalised Veltman frame (with respect to the definition of a generalised Veltman frame) such that we have $W = \{w, x_1, x_2, z\}$, wRx_1, wRx_2, wRz, and x_1Rz, and $zS_w\{x_1, x_2\}$.

Let us suppose that the frame $\mathfrak{F}$ does not satisfy the condition $(\mathsf{F})_{\text{gen}}$. Then there exists a world $u \in W$ and sequences of sets (A_n) and $(\mathcal{B}_n)$ such that

$$A_0\overline{S_u}\mathcal{B}_0\overline{R_u}A_1\overline{S_u}\mathcal{B}_1\dots$$

If $u = x_2$ or $u = z$ then we have $R[u] = \emptyset$. So, in this case the relation S_u is empty, which contradicts the fact that $A_0\overline{S_u}\mathcal{B}_0$.

Let us now consider the case $u = x_1$. Since $R[u] = \{z\}$, we have $S_u = \{(z, \{z\})\}$. Since $A_0\overline{S_u}\mathcal{B}_0$, it is necessary that $A_0 = \{z\}$ and $z \in \bigcup\mathcal{B}_0$. But $\mathcal{B}_0\overline{R_u}A_1$ then implies that there exists a world $v \in A_1$ such that zRv. This is impossible, because $R[z] = \emptyset$.

It remains to check the case $u = w$. Denote $B_n = \bigcup\mathcal{B}_n$ for each $n \in \omega$. First note that the set B_n cannot contain the world w for any $n \in \omega$. Suppose the contrary. Then $A_n\overline{S_w}\mathcal{B}_n$ and the definition of $\overline{S_w}$ implies $\mathcal{B}_n \subseteq 2^{R[w]} \setminus \{\emptyset\}$. This implies $B_n \subseteq R[w]$. But then we have $w \in R[w]$, contrary to the converse well-foundedness of R.

The set B_n also cannot contain the worlds x_2 and z. This is because we have $\mathcal{B}_n\overline{R_u}A_{n+1}$, but $R[x_2] = \emptyset$ and $R[z] = \emptyset$.

Let us now prove that the set B_n must be non-empty. We have $A_n\overline{S_w}\mathcal{B}_n$. The definition of $\overline{S_w}$ requires the set A_n to be non-empty. Thus, there exists a world $v \in A_n$ and a set $V \in \mathcal{B}_n$ such that vS_wV. But the definition of S_w requires V to be non-empty. Since $V \subseteq B_n$, the set B_n is also non-empty.

So, the set B_n does not contain the worlds w, x_2 or z, but is non-empty. Therefore, $B_n = \{x_1\}$, and thus $\mathcal{B}_n = \{\{x_1\}\}$. Since $\mathcal{B}_n\overline{R_w}A_{n+1}$, and $R[x_1] = \{z\}$, we have $z \in A_{n+1}$. Hence, for all $n \in \omega$ we have $B_n = \{x_1\}$ and $z \in A_{n+1}$. Choose any $n \in \omega$. Now $z \in A_{n+1}$, $A_{n+1}\overline{S_w}\mathcal{B}_{n+1}$ and the fact that $\mathcal{B}_{n+1} = \mathcal{B}_n = \{x_1\}$ imply $zS_w\{x_1\}$. This is impossible because we only have $zS_w\{z\}$, $zS_w\{z, x_1\}$, $zS_w\{z, x_2\}$, $zS_w\{z, x_1, x_2\}$ and $zS_w\{x_1, x_2\}$. So, we have proved that the frame $\mathfrak{F}$ satisfies the condition $(\mathsf{F})_{\text{gen}}$.

Now let us show that the frame $\mathfrak{F}$ satisfies $\overline{(\mathsf{W})_{\text{gen}}}$, i.e. does not satisfy the property $(\mathsf{W})_{\text{gen}}$. Let $V = \{x_1, x_2\}$. Obviously we have $V \subseteq R[w]$ and zS_wV. It remains to verify that the following holds:

$$(\forall V' \subseteq v)\Big(zS_wV \Rightarrow R[V'] \cap S_w^{-1}[V] \neq \emptyset\Big).$$

It is easy to see that the only subset V' of V such that zS_wV' holds is the set V itself (as we only have $zS_w\{z\}$, $zS_w\{z, x_1\}$, $zS_w\{z, x_2\}$, $zS_w\{z, x_1, x_2\}$ and $zS_w\{x_1, x_2\}$). Therefore, we only need to prove that $(\exists v \in \{x_1, x_2\})(R[v] \cap S_w^{-1}[V] \neq \emptyset)$. We indeed have x_1Rz and zS_wX. So, the frame $\mathfrak{F}$ does not satisfy the property $(\mathsf{W})_{\text{gen}}$. $\dashv$

2.7 Thesis structure

The first topic of this thesis is completeness with respect to generalised Veltman semantics. We explore this topic in the next two chapters. First we introduce the key tool: *assuring labels*, and then we employ this tool to obtain various completeness results. We touch again on completeness in the final chapter of the thesis.

Next, in Chapter 5, we apply completeness results and obtain decidability results. This is another, and perhaps the most useful, application of generalised semantics: the ability to define well-behaved filtrations.

Chapter 6 is on complexity; we prove that **IL**, **IL**W and **IL**P are all PSPACE-complete. We also provide commentary regarding the complexity of other decidable logics.

In Chapter 7 we work with the arithmetical aspect of interpretability logics. Namely, we give a new soundness proof for the two recently introduced series of principles $\mathsf{R_n}$ and $\mathsf{R^n}$ using an extended version of the system AtL presented in [40].

In the final chapter we introduce another series of principles, prove it arithmetically sound and provide ordinary Veltman semantics for it. We also provide conditions under which the resulting logic is modally complete w.r.t. generalised Veltman semantics; the validity of the conditions themselves remains an open problem.

2.8 Notation

In this final section we present some notational choices we fix for the remainder of this thesis.

Notation	Meaning	Context
$\mathcal{R}[x]$	$\{y : x\mathcal{R}y\}$	$\mathcal{R} \subseteq \mathcal{A} \times \mathcal{B}$ is a binary relation and $x \in \mathcal{A}$
$\mathcal{R}[X]$	$\{y : \exists x \in X, x\mathcal{R}y\}$	$\mathcal{R} \subseteq \mathcal{A} \times \mathcal{B}$ is a binary relation and $X \subseteq \mathcal{A}$
$\mathcal{R}^{-1}[x]$	$(\mathcal{R}^{-1})[x]$	$\mathcal{R} \subseteq \mathcal{A} \times \mathcal{B}$ is a binary relation and $x \in \mathcal{B}$ or $x \subseteq \mathcal{B}$
$\mathbf{L} \vdash A$	$\mathbf{L}$ proves A	$\mathbf{L}$ is a logic (possibly implicit) and A is a formula in the language of $\mathbf{L}$
$\mathcal{X} \Vdash A$	$\mathcal{X}$ satisfies A	$\mathcal{X}$ is a world, a model, or a frame. Depending on the type of $\mathcal{X}$, *satisfaction* is satisfaction in a world, global satisfiability in a model, or validity on a frame
$[A]_w$	$\{x \in R[w] : x \Vdash A\}$	w is a world and A is a modal formula

For example, $S_w^{-1}[V] = \{u : uS_wV\}$ if we are working with generalised semantics (where $S_w \subseteq W \times 2^W$).

Chapter 3

Labelling

In this chapter we introduce *assuringness*, a concept we will use heavily in the remainder of the thesis. In particular, it will be the key ingredient of our completeness proofs.

The first version of this text, containing some of the content of this chapter (definition and basic properties of assuring labels and labelling lemmas for W, P, M, $\mathsf{M_0}$ and R), has been published in [37] and [11]. The edited version, where we explore the so-called Γ-full assuring labels and iterated labelling systems exists as a preprint [26]. The content of this chapter is largely taken from this preprint.

Introduction

A large part of this thesis is concerned with connections between syntax and semantics. In order to prove semantic completeness of modal logics with respect to various forms of relational semantics, the usual approach is to let the set of worlds equal the set of maximal consistent sets w.r.t. the logic in question. With interpretability logics we sometimes use more elaborate definitions, for example worlds are sometimes identified with sequences of sets of formulas. Nevertheless, the key idea is the same: models are built out of maximal consistent sets.

The next issue, once we defined the set of worlds, is what will relations look like. In the case of Veltman semantics, the relations that we care about are the binary relation R and the ternary relation S. Bearing in mind that the set of worlds somehow corresponds to a set of maximal consistent sets, clearly R and S have to be reflected in some syntactic relations between maximal consistent sets.

For example, most modal logics with the unary modality $\Box$ admit the usual Kripke-style semantics with: xRy if and only if $\Box A \in x$ implies $A \in y$. Sometimes this definition is tweaked, depending on, for example, are the maximal consistent sets truncated (*maximal* with respect to some finite set) or we want to eliminate infinite chains. However, again, the key idea is the same: we want the relations of our model to be as large as possible, while staying compatible with the definition of truth. Usually it is the case that the less restricted the relations, the easier it is find a model for a consistent formula.

Thus, the relations S_w for $w \in W$ remain to be defined. This problem is central to Veltman semantics, and all known completeness proofs have the same general solutions: labels.

For example, if we are working in **IL** and wish to prove completeness w.r.t. ordinary Veltman semantics, we might want to let xS_wy hold if and only if x and y have the same label with respect to w (the approach used in [18]). We distinguish two main kinds of labels: critical labels and assuring labels. As mathematical objects, critical labels are single modal formulas and assuring labels are sets of formulas. Both types of labels have the associated relations of precedence between maximal consistent sets. Namely, given two maximal consistent sets w and x we can define $w \prec_L x$ where L is a label. We usually don't think of labels as existing in isolation, but rather in the context of some such maximal consistent sets w and x.

3.1 Preliminaries

Uppercase Greek, like Γ and Δ, will denote maximal consistent sets (MCS's). It will be clear from the context with respect to what logic the consistency will refer. Uppercase Roman denotes modal interpretability formulas $A, B, C, \ldots$ or sets of such formulas $S, T, U, \ldots$. An exception to this rule is that we might write formulas from a set S as S_i, S_j etc. In particular if S is a set of formulas, then $\bigvee S_i$ denotes a finite disjunction over some formulas in S. If we talk of logics we mean extensions of **IL**. As usual, we use $\boxdot A$ as an abbreviation for $A \wedge \Box A$. If S is a set of formulas then we write $\Box S$ for $\{\Box A \mid A \in S\}$.

Definition 3.1 For MCS's Γ and Δ we define $\Gamma \prec \Delta$ if

$$\forall A \left(\Box A \in \Gamma \Rightarrow A, \Box A \in \Delta\right).$$

When building models for consistent formulas we are to ensure a truth lemma:

$$\forall \Delta \forall B \ \left(B \in \Delta \ \Leftrightarrow \ \Delta \Vdash B\right). \tag{3.1}$$

We will now investigate what (3.1) imposes on relations S_w for $w \in W$. In particular, let us consider the condition for a formula $\neg(A \rhd B)$ to be true in some world x in some particular model. Recall that $x \Vdash \neg(A \rhd B)$ if and only if there is some world y so that xRy, so that $y \Vdash A$ but for no z for which yS_xz will we have $z \Vdash B$. In particular, since yS_xy we see that $y \Vdash \neg B$. Moreover, since yRu implies yS_xu we also see that $y \Vdash \Box\neg A$.

Thus, certain transitions $\Gamma R \Delta$ should come with a promise that for any Δ' with $\Delta S_\Gamma \Delta'$ we will have $\neg B, \Box\neg B \in \Delta'$. Of course, we should also have $\neg C, \Box\neg C \in \Delta'$ for any C so that $C \rhd B \in \Gamma$. Let us introduce the notion of criticality from [18].

Definition 3.2 For MCS's Γ and Δ, and for C a formula, we say that Δ is a C-*critical* successor of Γ whenever

$$\forall B \left(B \rhd C \in \Gamma \ \Rightarrow \ \neg B, \Box\neg B \in \Delta \right).$$

We will write $\Gamma \prec^C \Delta$ in this case.[1]

It is easy to see that C-criticality naturally extends the $\prec$ relation as reflected by the following easy lemma.

Lemma 3.3 For MCS's Γ and Δ we have $\Gamma \prec \Delta$ if and only if $\Gamma \prec^{\bot} \Delta$.

Proof. Immediate since $\Box A \in \Gamma$ holds if and only if $\neg A \rhd \bot \in \Gamma$. ⊣

We can see C-criticality as a promise that the formula C will be avoided in a strong sense. All completeness proofs before [11] made essential use of critical successors. Whenever in a structure of MCS's a $\Gamma \prec^C \Delta$ was there, the definition of the S_Γ relation should reflect the promise that C should be avoided. This strategy, although successful, resulted in a need for complicated book-keeping to keep all promises.

An improvement can be made if we can deal with various promises at the same time. Suppose we wished to define $\Gamma \prec^{B,C} \Delta$ in such a way that it promises that both B and C are avoided in Δ in a strong sense. Requiring that simultaneously both $\Gamma \prec^B \Delta$ and $\Gamma \prec^C \Delta$ is not sufficient since the promises may interact. In particular

$$\text{if } A \rhd B \vee C \in \Gamma \text{ we should also require that } \neg A, \Box\neg A \in \Delta.$$

It is this simple idea that adds a lot of power to the notion of criticality. However, there is one more subtlety to it. It turns out to be fruitful to apply a change of perspective. Instead of speaking of a promise to avoid certain formulas it turns out to be a very fruitful perspective to rather speak of *assuring* certain formulas. If we do so, the set of promises has certain nice properties. In particular, it can be closed under derivability as proven in Lemma 3.12. These considerations give rise to the following definition.

Definition 3.4 (Assuring successor) Let S be a set of formulas. We define $\Gamma \prec_S \Delta$, and say that Δ is an S-*assuring successor* of Γ, if for any finite $S' \subseteq S$ we have that $A \rhd \bigvee_{S_j \in S'} \neg S_j \in \Gamma$ implies $\neg A, \Box\neg A \in \Delta$ and for some $\Box C \in \Delta$ we have $\Box C \notin \Gamma$. We will call S a *label* for Γ and Δ or simply a *label*.[2]

[1] The usual notation for criticality is $\Gamma \prec_C \Delta$. We write $\Gamma \prec^C \Delta$ for criticality in this chapter in order to more clearly distinguish it from assuringness, which we denote with $\Gamma \prec_S \Delta$.

[2] While this definition is more in line with the old notion of criticality, occasionally we will not require that for some $\Box C \in \Delta$ we have $\Box C \notin \Gamma$. For example, we work without this condition throughout the whole chapter concerning modal completeness. The property will still hold one way or the other, it's just that sometimes it is convenient not to have this condition be required by the definition.

In the following lemma we shall see that the notion of assuring successor on sets of formulas naturally extends the regular successor relation as well as the critical successor relation.

Lemma 3.5 1. $\Gamma \prec_{\emptyset} \Delta$ if and only if $\Gamma \prec \Delta$;

2. Δ is a B-critical successor of Γ if and only if $\Gamma \prec_{\{\neg B\}} \Delta$, if and only if $\Gamma \prec^{B} \Delta$.

Proof. For the first item, we observe that the empty disjunction is per definition equivalent to $\bot$. We have $A \rhd \bot \in \Gamma$ if and only if $\Box\neg A \in \Gamma$. Consequently,

$$\forall A\Big(A \rhd \bot \in \Gamma \ \Rightarrow \ \neg A, \Box\neg A \in \Delta\Big) \text{ if and only if } \forall A\Big(\Box\neg A \in \Gamma \ \Rightarrow \ \neg A, \Box\neg A \in \Delta\Big).$$

Since we work in classical logic, the right-hand side is easily seen to be equivalent to $\forall A\Big(\Box A \in \Gamma \ \Rightarrow \ A, \Box A \in \Delta\Big)$.

The $\Leftarrow$ direction of the second item is easy and the other direction follows from the first item of this lemma: if we take a finite subset of $\{\neg B\}$ this is either the empty set, or $\{\neg B\}$ itself. Now, the fact that $A \rhd \neg\neg B \in \Gamma$ implies $\neg A, \Box\neg A \in \Delta$ follows from the assumption that Δ is a B-critical successor of Γ and that the fact $A \rhd \bot \in \Gamma$ implies $\neg A, \Box\neg A \in \Delta$ follows from the first item since critical successors are in particular successors. $\dashv$

3.2 Assuring and full labels

In this section we will expose a general theory of assuring successors. In the next section we will show how assuring successors can be used to solve, in a uniform way, certain problematic aspects of modal completeness proofs.

As the name suggests, assuring labels assure certain formulas to be present. The relation $\prec_S$ assures elements in Γ and Δ, and in a sense it is not allowed to "speak" of consistency formulas, i.e. $\Diamond$-formulas cannot be contained in a label. This is made explicit in the following lemma.

Lemma 3.6 We have the following:

1. if $\Gamma \prec_S \Delta$ then $S, \Box S \subseteq \Delta$;

2. if $\Gamma \prec_S \Delta$ then $\Diamond S \subseteq \Gamma$;

3. if $\Gamma \prec_S \Delta$ then the label S does not contain any formula of the form $\Diamond A$.

Proof. The first item is clear since for any $A \in S$ we have that $\neg A \rhd \neg A$ is a theorem and whence in Γ. By the definition of $\Gamma \prec_S \Delta$ we get that $A, \Box A \in \Delta$.

The second item follows from the first: since Γ is maximal, for any $A \in S$, either $\Diamond A \in \Gamma$ or $\Box\neg A \in \Gamma$. However, the latter would imply $\neg A \in \Delta$ contradicting our first item.

For the last item we reason as follows. Suppose for a contradiction that there is some $\Diamond A$ in S. Then, by the first item we have both $\Diamond A \in \Delta$ and $\Box\Diamond A \in \Delta$. However, over **GL** we have that $\Box\Diamond A$ is equivalent to $\Box\bot$. But $\Box\bot \in \Delta$ clearly contradicts $\Diamond A \in \Delta$. $\dashv$

A label S between $\Gamma \prec_S \Delta$ keeps track of the formulas that are promised to be in Δ in virtue of certain interpretability formulas in Γ. The larger the label, the more promises it stores.

Often we can enlarge the label for free. To see how much we can add we need the following definition.

Definition 3.7 For any set of formulas T and maximal consistent set Δ we define

$$\Delta_T^{\Box} = \{\Box\neg A \mid A \rhd \bigvee_{T_i \in T'} \neg T_i \in \Delta \text{ for some finite } T' \subseteq T\},$$
$$\Delta_T^{\boxdot} = \{\Box\neg A, \neg A \mid A \rhd \bigvee_{T_i \in T'} \neg T_i \in \Delta \text{ for some finite } T' \subseteq T\}.$$

Note that $\Delta_\emptyset^{\Box} = \{\Box\neg A \mid A \rhd \bot \in \Delta\}$. However, we want to think about this set as $\{\Box C \mid \Box C \in \Delta\}$. Clearly the two sets, although not literally equal, behave the same when used within labels. The next lemma tells us how promises propagate over composition of successors.

Lemma 3.8 For the relation $\prec_S$ we claim the following:

1. if $S \subseteq T$ & $\Gamma \prec_T \Delta$ then $\Gamma \prec_S \Delta$;
2. if $\Gamma \prec_S \Delta \prec \Delta'$ then $\Gamma \prec_S \Delta'$.

Proof. The first item is obvious since any finite subset of S is also a finite subset of T whenever $S \subseteq T$. For the second item we observe that $\Gamma \prec_S \Delta$ implies $\Gamma_S^{\boxdot} \subseteq \Delta$ whence by $\Delta \prec \Delta'$ and $\Box\Gamma_S^{\boxdot} \subseteq \Gamma_S^{\boxdot}$ we see that $\Gamma_S^{\boxdot} \subseteq \Delta'$. $\dashv$

Notation 3.9 Often we shall simply write $\bigvee \neg S_i$ to indicate some particular finite disjunction without really specifying it. If in the same context we will need another particular but otherwise unspecified big disjunction we will flag this by using a different index. Thus, $\bigvee \neg S_i \vee \bigvee \neg S_j$ stands for the disjunction of two particular but unspecified finite disjunctions of negated formulas from some label set S.

Often we will consider a finite collection of formulas C_j such that each C_j will interpret some finite disjunction of negated formulas from the label S. For each particular formula C_j we will denote the corresponding disjunction by $\bigvee \neg S_k^j$ and thus write $C_j \rhd \bigvee \neg S_k^j$. Subsequently, we will denote the big disjunction over all k and all corresponding $\neg S_j^k$ by $\bigvee \neg S_j^k$ so that (with the help of a few applications of axioms of **IL**) $\bigvee C_k \rhd \bigvee \neg S_j^k$.

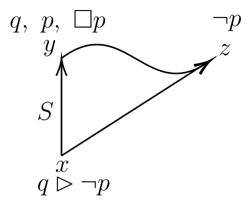


Figure 3.1: Situation described in Lemma 3.11.

The following lemma gives us a way to extend labels.

Lemma 3.10 For any logic (i.e. extension of **IL**) we have that $\Gamma{\prec_S}\Delta$ implies $\Gamma{\prec_{S\cup\Gamma^{\boxdot}_S}}\Delta$.

Proof. Suppose $\Gamma{\prec_S}\Delta$ and $C \rhd \bigvee\neg S_i \vee \bigvee A_j \vee \Diamond A_j \in \Gamma$ for some finite collection of formulas $\neg A_j, \Box\neg A_j \in \Gamma^{\boxdot}_S$. In particular, for each j we have $A_j \rhd \bigvee \neg S^j_k \in \Gamma$ for some finite collection (depending on j) of formulas S^j_k from the label S. Then $C \rhd \bigvee\neg S_i \vee \bigvee A_j \in \Gamma$ and thus $C \rhd \bigvee\neg S_i \vee \bigvee\neg S^j_k \in \Gamma$ which implies $\neg C, \Box\neg C \in \Delta$ since we assumed $\Gamma{\prec_S}\Delta$. ⊣

This lemma tells us in a sense that when we have $\Gamma{\prec_S}\Delta$, then certain sentences in Γ justify that we may extend the label S. Will likewise the occurrence of sentences in Δ allow us to extend the label S? The next lemma tells us that this is not the case. In particular, if $A \rhd \bigvee\neg S_i$ for some $S_i \in S' \subseteq_{\mathsf{fin}} S$, then by definition $\neg A, \Box\neg A \in \Delta$. However, when for some arbitrary A we have $\neg A, \Box\neg A \in \Delta$, this does not allow us to extend our label S.

Lemma 3.11 There are MCS's Γ and Δ, and a set S such that $\Gamma{\prec_S}\Delta$, and for some propositional variable p we have $p, \Box p \in \Delta$ but $\Gamma \not\prec_{S\cup\{p\}} \Delta$.

Proof. Consider the model consisting of three points x, y and z given in Figure 3.1. Let $\Gamma = \{A : x \Vdash A\}$ and $\Delta = \{A : y \Vdash A\}$. Since $q \in \Delta$ and $(q \rhd \neg p) \in \Gamma$, whatever we take for S with $\Gamma{\prec_S}\Delta$, we will never have $\Gamma{\prec_{S\cup\{p\}}}\Delta$. ⊣

Thus, via the previous two lemmas we see that the S-assuringness between two sets $\Gamma{\prec_S}\Delta$ can only be automatically extended via Γ. The next lemma tells us that there are other ways to 'freely extend' a label.

Lemma 3.12 For any logic we have

1. $\Gamma{\prec_S}\Delta$ and $S \vdash A$ implies $\Gamma{\prec_{S\cup\{A\}}}\Delta$;

2. $\Gamma{\prec_S}\Delta$ implies $\Gamma{\prec_{S\cup\Box S}}\Delta$.

Proof. For the first item—that labels can be closed under derivability—we assume that $S \vdash A$ where the notion of derivability depends on the logic in question. Thus, for some $S_1, \ldots, S_n \in S$ we have $S_1 \wedge \ldots \wedge S_n \vdash A$. Consequently, $\vdash \neg A \to \bigvee\neg S_j$ and also $\vdash \Box(\neg A \to \bigvee\neg S_j)$. Thus, if $\Gamma{\prec_S}\Delta$ and $(B \rhd \bigvee\neg S_i \vee \neg A) \in \Gamma$, also $(B \rhd \bigvee\neg S_i) \in \Gamma$ so that $\neg B, \Box\neg B \in \Delta$ and we conclude $\Gamma{\prec_{S\cup\{A\}}}\Delta$.

For the second item, we consider $(A \rhd \bigvee \neg S_i \vee \bigvee \neg \Box S_j) \in \Gamma$. But since $\neg \Box S_j \equiv \Diamond \neg S_j$ and $\Diamond \neg S_j \rhd \neg S_j$ we conclude $(A \rhd \bigvee \neg S_i \vee \bigvee \neg S_j)$ so that $\neg A, \Box \neg A \in \Delta$. ⊣

This Lemma 3.12 tells us that given an extension **IL**X of the logic **IL**, we can freely extend labels to be closed under **IL**X-derivability (where **IL**X is an arbitrary extension of **IL**) and to be closed under necessitation. Thus, we can identify labels with **IL**X theories.

Moreover, Lemma 3.10 tells us that we can freely close off a label S for $\Gamma \prec_S \Delta$ under $\Gamma^{\boxdot}_S$. These observations lead us to the definition of Γ-*full labels*. When the context makes clear which Γ is meant, we shall simply speak of *full labels*.

Definition 3.13 For Γ a maximal consistent set we call S a Γ-*full label* whenever S is a logic extending $\Gamma^{\boxdot}_S$.

In concrete, S is a Γ-full label whenever we have the following:

1. $A \rhd \bigvee \neg S_i \in \Gamma$ implies $\neg A, \Box \neg A \in S$;
2. in particular $\Box A \in \Gamma$ implies $A \in S$;
3. the label S is closed under derivability, that is, if $S \vdash A$, then $A \in S$;
4. the label S is closed under necessitation, that is, if $B \in S$, then $\Box B \in S$.

If we stick to full labels, there is a close correspondence between theories and labels. We find this observation so essential that we formulate it explicitly as a lemma:

Lemma 3.14 If $\Gamma \prec_S \Delta$ and S is a full label, then S is an **IL**X-theory.

We pose as open question whether for any consistent **IL**X-theory S we can find MCS's Γ and Δ so that $\Gamma \prec_S \Delta$. In case this could be answered in the affirmative it would be interesting to know whether the result can be extended to arbitrary chains of increasing theories.

Full labels contain as many free promises as possible and posses certain nice closure properties. In particular, we have the following lemma that justify the name 'full'.

Lemma 3.15 Given a MCS Γ and a label S, then S is Γ-*full* if and only if the following holds:

$$\forall T \Big(S \subseteq T \wedge \forall \Delta \big(\Gamma \prec_S \Delta \Rightarrow \Gamma \prec_T \Delta \big) \implies S = T \Big).$$

The sets of formulas S and T range here over **IL**X-theories, and Δ over MCS's.

Proof. First assume that S is a Γ-full label and $S \subsetneq T$. We want to show there is a MCS Δ with $\Gamma \prec_S \Delta$ but $\Gamma \not\prec_T \Delta$. As $S \subsetneq T$, there is some $A \in T$ for which we have $A \notin S$, and therefore, by S being Γ-full, $A \notin \Gamma^{\boxdot}_S \subseteq S$. Since S is a theory also $S \nvdash A$ and $\Gamma^{\boxdot}_S \nvdash A$. Then there exists a MCS Δ containing $\Gamma^{\boxdot}_S$ with $A \notin \Delta$. Clearly $\Gamma \prec_S \Delta$, and as $\neg A \rhd \neg A \in \Gamma, A \in T$ and $A \notin \Delta$, we see that $\Gamma \not\prec_T \Delta$.

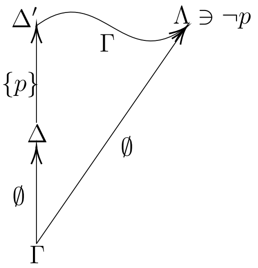


Figure 3.2: Downward influence

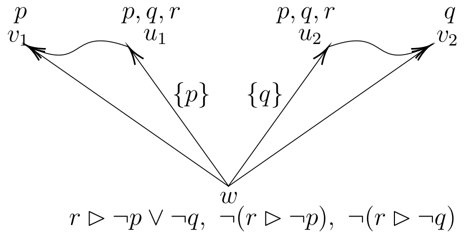


Figure 3.3: Incomparable labels

For the other direction assume $\Gamma^{\boxdot}_{S} \not\subseteq S$. We want to find a theory $T \supset S$ with $\forall\Delta\big(\Gamma{\prec_S}\Delta \Rightarrow \Gamma{\prec_T}\Delta\big)$. Take T to be the theory generated by $S \cup \Gamma^{\boxdot}_{S}$: it certainly is bigger than S. Now assume $\Gamma{\prec_S}\Delta$, but then $\Gamma{\prec_{S\cup\Gamma^{\boxdot}_{S}}}\Delta$ by Lemma 3.10. ⊣

Full labels can at times simplify matters. In particular, they clearly propagate along successors as expressed by the following lemma.

Lemma 3.16 If $\Gamma{\prec_S}\Delta{\prec_T}\Lambda$, for some full labels S and T, then $S \subseteq T$.

Proof. For any $S_i \in S$ we have $\Box S_i \in \Delta$ so, by fullness, $S_i \in T$. ⊣

Thus, this Lemma states that full labels accrue information along the top successor relation. Does information between related full labels also 'reflect down'? To put it otherwise, it may be natural to ask if Lemma 3.8, Item 2 (that $\Gamma{\prec_S}\Delta \prec \Delta'$ implies $\Gamma{\prec_S}\Delta'$), can be strengthened. That is to say, suppose we have $\Gamma{\prec_S}\Delta{\prec_T}\Delta'$, can we say something more than just $\Gamma{\prec_S}\Delta'$? As we shall see in the next section, it turns out that for extensions of **IL** we often can. In general this does not seem to hold, at least if we do not require our labels to be full. Suppose $\Gamma{\prec_\emptyset}\Delta{\prec_{\{p\}}}\Delta'$ (see Figure 3.2). If $p \rhd \neg p \in \Gamma$ and $p \in \Delta'$, there is an MCS Λ with $\neg p \in \Lambda$. Clearly, the fact that we have $\{p\}$ between Δ and Δ' did not stop $\neg p \in \Lambda$. Let us mention a question that we do not have a definite answer for. Suppose $\Gamma{\prec_S}\Delta{\prec_T}\Delta'$ and S and T are full labels. Is there a (non-trivial) notion of a "T-influenced formula" such that we may put the T-influenced formulas between Γ and Δ'?

Although a label can be full, this does not mean we can always find a maximum among the possible labels. We shall now exhibit a model that generates maximal consistent sets Γ and Δ with two incomparable labels between them.

Lemma 3.17 There are maximal consistent sets Γ and Δ, and labels S and T with $\Gamma{\prec_S}\Delta$ and $\Gamma{\prec_T}\Delta$ so that $\Gamma \not\prec_{S\cup T} \Delta$.

Proof. We let $S := \{p\}$, $T := \{q\}$ and consider the model in Figure 3.3. Let $\Gamma = \{A : w \Vdash A\}$, $\Delta_1 = \{A : u_1 \Vdash A\}$ and $\Delta_2 = \{A : u_2 \Vdash A\}$. Clearly, $\Delta_1 = \Delta_2$. However, we do not have $\Gamma{\prec_{\{p,q\}}}\Delta_1$, since this would imply $v_1 \Vdash q$ which is not the case. Similarly for u_2, Δ_2 and p. ⊣

When $\Gamma \prec_S \Delta$, this enforces many formulas of the form $\neg(A \rhd B)$ to be in Γ as we can see in the next lemma.

Lemma 3.18 Let $\Gamma \prec_S \Delta$ with $A \in \Delta$ and $S' \subseteq_{\mathsf{fin}} S$. We then have

$$\neg(A \rhd \bigvee_{S_i \in S'} \neg S_i) \in \Gamma.$$

Proof. Suppose $A \rhd \bigvee_{S_i \in S'} \neg S_i \in \Gamma$. Then by $\Gamma \prec_S \Delta$ we would have $\neg A \in \Delta$ which is a contradiction. Thus, $A \rhd \bigvee_{S_i \in S'} \neg S_i \notin \Gamma$ and by maximality $\neg(A \rhd \bigvee_{S_i \in S'} \neg S_i) \in \Gamma$. ⊣

Conversely, the next lemma will show that given a label S and maximal consistent set Γ we have: if there are sufficiently many negated interpretability formulas related to S in Γ, then we can conclude that there exists some MCS Δ with $\Gamma \prec_S \Delta$.

Theorem 3.19 Let Γ be an MCS, B a formula and S a set of formulas. If for any choice of $S' \subseteq S$ we have that $\neg(B \rhd \bigvee_{S_i \in S'} \neg S_i) \in \Gamma$, then[3] there exists an MCS Δ such that $\Gamma \prec_S \Delta \ni B, \Box\neg B$.

Proof. Suppose for a contradiction there is no such Δ. Then there is a formula[4] A such that for some $S_i \in S$ we have $(A \rhd \bigvee \neg S_i) \in \Gamma$ and $B, \Box\neg B, \neg A, \Box\neg A \vdash \bot$. Then $\vdash \Box\neg B \wedge B \rhd A \vee \Diamond A$, and we get $\vdash B \rhd A$. As $(A \rhd \bigvee \neg S_i) \in \Gamma$, also $(B \rhd \bigvee \neg S_i) \in \Gamma$. A contradiction. ⊣

The following lemma is sometimes called a "problem-solving lemma" (see [27]).

Lemma 3.20 Let Γ be an MCS such that $\neg(B \rhd C) \in \Gamma$. Then there is an MCS Δ such that $\Gamma \prec_{\{\neg C\}} \Delta$ and $B, \Box\neg B \in \Delta$.

Proof. Taking $S = \{\neg C\}$ in Theorem 3.19. ⊣

The following lemma is sometimes called a "deficiency solving lemma" (see [27]).

Lemma 3.21 Let Γ and Δ be MCS's such that $A \rhd B \in \Gamma \prec_S \Delta \ni A$. Then there is an MCS Δ' such that $\Gamma \prec_S \Delta' \ni B, \Box\neg B$.

Proof. First we see that for any choice of S_i, $\neg(B \rhd \bigvee \neg S_i) \in \Gamma$. Suppose not. Then for some S_i, $(B \rhd \bigvee \neg S_i) \in \Gamma$ because Γ is an MCS. But then $(A \rhd \bigvee \neg S_i) \in \Gamma$ and by $\Gamma \prec_S \Delta$ we have $\neg A \in \Delta$. A contradiction. So $\neg(B \rhd \bigvee \neg S_i) \in \Gamma$ for any choice of S_i and we can apply Theorem 3.19. ⊣

[3] Lemma 3.18 tells us that we actually have *if and only if*.

[4] There are finitely many A_j such that there exist some formulas $S_i^j \in S$ with $(A_j \rhd \bigvee \neg S_i^j) \in \Gamma$ and $\Box\neg B, B, \neg A_j, \Box\neg A_j \vdash \bot$. We can take A to be $\bigvee_j A_j$.

3.3 Frame conditions and labelling lemmas

In this section we recall what steps there are along the way when constructing a counter-model to an unprovable formula. We can think of the step-by-step method of constructing a counter-model used e.g. in [27] for now. In the next chapter we use these results in order to prove completeness of various systems w.r.t. generalised Veltman semantics. In addition to that, these results have been used in order to prove completeness and the finite model property for **IL**W w.r.t. ordinary Veltman semantics. We do not include the latter here since the author of the thesis did not contribute to it (this proof first appeared in [11]).

The idea is to build a model from MCS's and define the R and S_Γ accessibility relations on them, where in particular the R relation is to be defined using $\prec$. We wish to use the labels along $\prec$ to keep track of the promises posed on later added worlds by already contained interpretability formulas, and, as we shall see, also to be able to ensure we can "locally satisfy" the frame conditions corresponding to the additional axiom schemas, i.e. we can close the model under the characteristic property of the schema.

Let W be a multiset of MCS's used in the model we wish to define. The main points one has to address are the following three:

1. For each $\Gamma \in W$ with $\neg(A \rhd B) \in \Gamma$ we need to include a $\{\neg B\}$-assuring successor Δ in W for which $A \in \Delta$.
2. For each $\Gamma, \Delta \in W$ with $C \rhd D \in \Gamma \prec \Delta \ni C$ we need to include a Δ' in W for which $\Gamma \prec \Delta' \ni D$. Moreover if Δ is a T-assuring successor of Γ then we should be able to choose Δ' a T-assuring successor of Γ as well (to carry promises along the S_Γ relation).
3. We need to make sure all the appropriate frame conditions are satisfied.

The existence Lemmas 3.20, 3.21 of the previous section ensure existence of MCS's required to witness modal formulas as specified in Item 1 and Item 2 above. When working in **IL** alone, making sure that the frame conditions are satisfied does not pose any problems [35], as they are just the basic properties of R and S_Γ, but with various extensions of **IL** the situation regarding the frame conditions for the additional modal principles becomes more complicated (cf. [18, 27]). Note that many of these issues disappear when one uses generalised Veltman semantics, as we shall see in the next chapter. However, ordinary Veltman semantics is easier to visualise and demonstrate key ideas.

When MCS $\Sigma \ni D$ is chosen witnessing a formula $C \rhd D \in \Gamma \prec \Delta \ni C$ by Item 2 (Page 39), we want to be able to do so in a way where not only $\Gamma \prec \Sigma$ (and the same formulas are assured), but also $\Lambda \prec \Sigma$. Moreover, if $\Lambda \prec_T \Delta$, it should be possible to

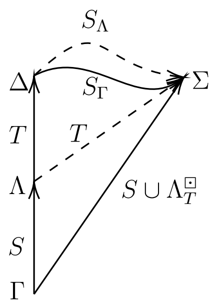


Figure 3.4: Ensuring the frame condition for P

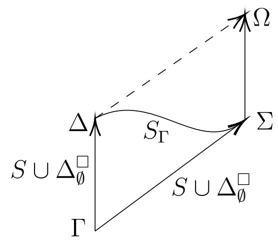


Figure 3.5: Ensuring the frame condition for M

choose Σ so that $\Lambda \prec_T \Sigma$. Only then it is consistent to draw the $\Delta S_\Lambda \Sigma$ arrow required by the frame condition, as depicted in Figure 3.4.

To see such requirements are indeed possible to meet, we will prove, for each principle, a *labeling lemma*. Labeling lemmas tell us how to label the $\prec$ relation in a sufficient way to ensure we can meet the requirements imposed by frame conditions locally. Note that we do not require labels to be full in the remainder of this chapter; unless stated otherwise, labels can be any sets of modal formulas, that is, they *do not* have to be Γ-full for any Γ.

Principle P. Let us see how frame conditions locally impose requirements on MCS's, taking **IL**P as the first example. The frame condition for P is the following [18]:

$$wRw'RuS_wv \Rightarrow uS_{w'}v.$$

The frame condition for P imposes on MCS's the following:

$$\Gamma \prec \Lambda \prec \Delta S_\Gamma \Sigma \Rightarrow \Delta S_\Lambda \Sigma.$$

Lemma 3.22 For logics containing P we have

$$\Gamma \prec_S \Lambda \prec_T \Delta \Rightarrow \Gamma \prec_{S \cup \Lambda_T^{\boxdot}} \Delta.$$

Proof. Suppose $C \rhd \bigvee \neg S_i \vee \bigvee A_j \vee \Diamond A_j \in \Gamma$, where $\Box\neg A_j, \neg A_j \in \Lambda_T^{\boxdot}$. Then $C \rhd \bigvee \neg S_i \vee \bigvee A_j \in \Gamma$ and thus by P we obtain $C \rhd \bigvee \neg S_i \vee \bigvee A_j \in \Lambda$. Since $\Gamma \prec_S \Lambda$ we have $\Box \bigwedge S_i \in \Lambda$ so we obtain $C \rhd \bigvee A_j \in \Lambda$. But for each A_j we have $A_j \rhd \bigvee \neg T_{jk} \in \Lambda$ and thus $C \rhd \bigvee \neg T_{jk} \in \Lambda$. Since $\Lambda \prec_T \Delta$ we conclude $\neg C, \Box\neg C \in \Delta$. ⊣

In the case of P, a simpler labelling lemma can be used to ensure the frame condition locally, provided we consider the labels that are *full* (S a Γ-full label, and T a Λ-full label).

Lemma 3.23 For logics containing P we have

$$\Gamma \prec_S \Lambda \prec_T \Delta \Rightarrow \Gamma \prec_T \Delta$$

Proof. Assume $\Gamma \prec_S \Lambda \prec_T \Delta$, and $C \rhd \bigvee \neg T_i \in \Gamma$. Then by P we know $C \rhd \bigvee \neg T_i \in \Lambda$. Since $\Lambda \prec_T \Delta$ we conclude $\neg C, \Box\neg C \in \Delta$. ⊣

Note that the lemma is true in the case of ordinary labels, but in that case, the previous lemma gives us more precise labelling information to ensure the frame condition locally. This is because only for full labels we in fact have $S \cup \Lambda_T^{\boxdot} \subseteq T$.

Principle M. The frame condition for M is the following [18]:

$$wRuS_wvRz \Rightarrow uRz.$$

The frame condition for M imposes on MCS's the following:

$$\Delta S_\Gamma \Sigma \prec \Omega \Rightarrow \Delta \prec \Omega.$$

When MCS $\Sigma \ni D$ is chosen witnessing a formula $C \rhd D \in \Gamma \prec \Delta \ni C$ by Item 2 (Page 39), we want to do so in such a way that whenever we later need to add a MCS Ω with $\Sigma \prec \Omega$, we can also draw the $\Delta \prec \Omega$ arrow. Therefore, we need to ensure $\Delta_\emptyset^\Box$ along the $\Gamma \prec \Sigma$ arrow (as we remarked previously, one can think of the set $\Delta_\emptyset^\Box$ as simply $\{\Box C \mid \Box C \in \Delta\}$), we achieve this by ensuring $\Delta_\emptyset^\Box$ along the $\Gamma \prec \Delta$ arrow. The situation is depicted in Figure 3.5. The corresponding labelling lemma is the following:

Lemma 3.24 For logics containing M we have $\Gamma \prec_S \Delta \Rightarrow \Gamma \prec_{S \cup \Delta_\emptyset^\Box} \Delta$.

Proof. Assume that for some $\Box C_j \in \Delta_\emptyset^\Box$ we have $(A \rhd \bigvee \neg S_i \vee \bigvee \neg \Box C_j) \in \Gamma$. By M, $(A \wedge \bigwedge \Box C_j \rhd \bigvee \neg S_i) \in \Gamma$, whence $\boxdot\neg(A \wedge \bigwedge \Box C_j) \in \Delta$. As $\bigwedge \Box C_j \in \Delta$, we conclude $\neg A, \Box\neg A \in \Delta$. ⊣

In the case of M, we have no simpler labelling lemma in case S is a Γ-full label.

Principle $\mathsf{M_0}$. The frame condition for $\mathsf{M_0}$ is the following [27]:

$$wRuRxS_wvRz \Rightarrow uRz.$$

The frame condition for $\mathsf{M_0}$ imposes on MCS the following:

$$\Gamma \prec \Delta \prec \Delta' S_\Gamma \Sigma \prec \Omega \Rightarrow \Delta \prec \Omega.$$

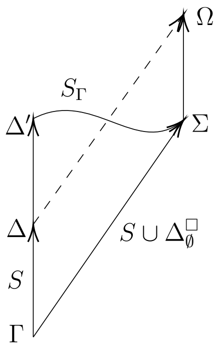


Figure 3.6: Ensuring the frame condition for M_0

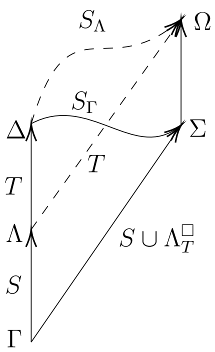


Figure 3.7: Ensuring the frame condition for R

When MCS $\Sigma \ni D$ is chosen to witness a formula $C \rhd D \in \Gamma \prec \Delta \prec \Delta' \ni C$ by Item 2 (Page 39), we want to do so in such a way that whenever we later need to add a MCS Ω with $\Sigma \prec \Omega$, we can also draw the $\Delta \prec \Omega$ arrow. Therefore, we again need to ensure $\Delta_\emptyset^\Box$ along the $\Gamma \prec \Sigma$ arrow. The situation is depicted in Figure 3.6, and the corresponding labelling lemma is the following (as before, we do not have a special lemma in case the labels are full):

Lemma 3.25 For logics containing M_0 we have $\Gamma \prec_S \Delta \prec \Delta' \Rightarrow \Gamma \prec_{S \cup \Delta_\emptyset^\Box} \Delta'$.

Proof. Suppose $C \rhd \bigvee \neg S_i \vee \bigvee \Diamond A_j \in \Gamma$, where $\Box \neg A_j \in \Delta_\emptyset^\Box$. By M_0 we obtain $\Diamond C \wedge \bigwedge \Box \neg A_j \rhd \bigvee \neg S_i \in \Gamma$. So, since $\Gamma \prec_S \Delta$ and $\bigwedge \Box \neg A_j \in \Delta$ we obtain $\Box \neg C \in \Delta$ and thus $\Box \neg C, \neg C \in \Delta'$. ⊣

Principle R. Last we will look at a more complicated case of **IL**R.

The frame condition for the principle R is the following [28]:

$$wRxRyS_w y' Rz \Rightarrow yS_x z.$$

On MCS's the condition imposes the following:

$$\Gamma \prec \Lambda \prec \Delta S_\Gamma \Sigma \prec \Omega \Rightarrow \Delta S_\Lambda \Omega.$$

The frame condition is depicted in Figure 3.7. Assume $\Sigma \ni D$ was chosen as a witness for $C \rhd D \in \Gamma R \Delta \ni C$. Since Δ lies T-assuring above Λ, we should not only make sure that Σ lies S-assuring above Γ, but also that any successor Ω of Σ lies T-assuring above Λ. Only then we would be justified to draw the required $\Delta S_\Lambda \Omega$ arrow. One way to guarantee $\Lambda \prec_T \Omega$ is to ensure $\Lambda_T^\Box$ along the $\Gamma \prec \Sigma$ arrow: whenever $B \rhd \bigvee \neg T_i \in \Lambda$, we have $\Box \neg B \in \Lambda_T^\Box$ and this puts $\Box \neg B \in \Delta$ and $\Box \neg B, \neg B \in \Omega$ as required.

The corresponding labelling lemma is the following:

Lemma 3.26 For logics containing R we have $\Gamma \prec_S \Lambda \prec_T \Delta \Rightarrow \Gamma \prec_{S \cup \Lambda_T^\Box} \Delta$.

Proof. We consider A such that for some $S_i \in S$ and some $\Box\neg A_j \in \Lambda_T^{\Box}$, we have $(A \rhd \bigvee \neg S_i \vee \bigvee \Diamond A_j) \in \Gamma$. By R we obtain $(\neg(A \rhd \bigvee A_j) \rhd \bigvee \neg S_i) \in \Gamma$, thus by $\Gamma \prec_S \Lambda$ we get $(A \rhd \bigvee A_j) \in \Lambda$. As $(A_j \rhd \bigvee \neg T_{kj}) \in \Lambda$, also $(A \rhd \bigvee \neg T_{kj}) \in \Lambda$. By $\Lambda \prec_T \Delta$ we conclude $\boxdot\neg A \in \Delta$. ⊣

In the case of R, a simpler labelling lemma can be used to ensure the frame condition locally if T is Λ-full:

Lemma 3.27 For logics containing R we have $\Gamma \prec_S \Lambda \prec_T \Delta \Rightarrow \Gamma \prec_{S \cup \Box T} \Delta$.

Proof. Assume $A \rhd \bigvee \neg S_i \vee \bigvee \neg\Box T_j \in \Gamma$. Then, by R, we obtain $\neg(A \rhd \bigvee \neg T_j) \rhd \bigvee \neg S_i \in \Gamma$ and by $\Gamma \prec_S \Lambda$ we know $\boxdot(A \rhd \bigvee \neg T_j) \in \Lambda$, and $\boxdot\neg A \in \Sigma$ as required. ⊣

As before in the case of logics containing P and Lemma 3.23, this lemma ensures the frame condition locally provided the labels are full: for in this case $\Lambda_T^{\Box} \subseteq T$ and therefore, because T is a theory, $\Lambda_T^{\Box} \subseteq \Box T$, and consequently $S \cup \Lambda_T^{\Box} \subseteq S \cup \Box T$. Thus, sufficient information is carried by the composed label.

Principle W. Let us state two existence lemmas for **IL**W, a logic for which only second-order frame properties are known ([28], [49]).

We do not know if there is a labelling lemma for this logic in the same sense as before. However, we know that the following lemmas are sufficiently strong for completeness proofs w.r.t. both ordinary and generalised semantics (see [11] and the following chapter).

Lemma 3.28 Suppose Γ is an **IL**W-MCS. Suppose $\neg(A \rhd B) \in \Gamma$. There exists some **IL**W-MCS Δ with $\Gamma \prec_{\{\Box\neg A, \neg B\}} \Delta$ and $A \in \Delta$.

Proof. Suppose for a contradiction that there is no such Δ. Then there are finitely many formulas E_i such that $(E_i \rhd \Diamond A \vee B) \in \Gamma$ and $A, \{\neg E_i, \Box\neg E_i\}_i \vdash \bot$. Let $E = \bigvee_i E_i$. By **IL** and maximal consistency we have $(E \rhd \Diamond A \vee B) \in \Gamma$ and $A, \neg E, \Box\neg E \vdash \bot$. Thus, $\vdash A \rhd E$. Then $(A \rhd \Diamond A \vee B) \in \Gamma$ and by the principle W we have $A \rhd B \in \Gamma$. The contradiction. ⊣

Lemma 3.29 For logics containing W we have that if $B \rhd C \in \Gamma \prec_S \Lambda \ni B$ then there exists Δ with $\Gamma \prec_{S \cup \{\Box\neg B\}} \Delta \ni C, \Box\neg C$.

Proof. Suppose for a contradiction that no such Δ exists. Then for some formula A with $(A \rhd \bigvee \neg S_i \vee \Diamond B) \in \Gamma$, we get $C, \Box\neg C, \neg A, \Box\neg A \vdash \bot$, whence $\vdash C \rhd A$. Thus, $B \rhd C \rhd A \rhd \bigvee \neg S_i \vee \Diamond B \in \Gamma$. By W, $B \rhd \bigvee \neg S_i \in \Gamma$ which contradicts $\Gamma \prec_S \Lambda \ni B$. ⊣

In this chapter we introduced the main tool we use in our completeness proof with respect to generalised Veltman semantics, the assuring labels. There is a bit more to labels than what has been said here: in particular, at one point we will be needing *iterated label systems*. For the most part of the following chapter (on completeness) what we've seen so far will suffice. Thus, we will return to discussing labels near the end of the following chapter, when a new approach is required.

Chapter 4

Modal completeness

The aim of this chapter is to explore modal completeness with respect to generalised Veltman semantics. We will first say a few words on the history of modal completeness proofs concerning interpretability logics.

The content of this chapter, except for the final section, is largely taken from the recent paper [50]. If not stated otherwise, it is safe to assume that those results and proofs are taken from [50]. The content of [50] is copyrighted by Cambridge University Press.

The content of the final section is largely taken from the preprint [26].

Introduction

In this chapter, and in fact in the whole thesis, the only notion of semantic completeness we discuss is that of weak completeness. More precisely, let X be any (possibly empty) set of modal formulas and $\mathbf{IL}\mathsf{X}$ the result of extending the base logic $\mathbf{IL}$ with X. Let $\mathcal{C}(\mathsf{X})$ be the characteristic class of $\mathbf{IL}\mathsf{X}$ w.r.t. ordinary Veltman semantics, i.e. $\mathcal{C}(\mathsf{X}) = \{\mathfrak{F} : \mathfrak{F}$ is an ordinary Veltman frame and $\mathfrak{F} \Vdash \mathsf{X}\}$. We say that $\mathbf{IL}\mathsf{X}$ is complete w.r.t. ordinary Veltman semantics if for all formulas A we have that: $\mathcal{C}(\mathsf{X}) \Vdash A$ implies $\mathbf{IL}\mathsf{X} \vdash A$.

We define completeness w.r.t. generalised semantics similarly, by replacing "ordinary" with "generalised" in the preceding definition.

Note that we do not have any sort of strong completeness result for any conservative extension of $\mathbf{GL}$ (which includes all logics studied in this thesis), since the counterexamples for $\mathbf{GL}$ (see e.g. [5]) are also counterexamples for such extensions. Since for $\mathbf{GL}$ this issue can be overcome by switching to topological semantics, it is possible that with topological semantics we have strong completeness for interpretability logics. However, at the time of writing, topological semantics for interpretability logics has not been studied, and the formulation of a strongly complete semantics for interpretability logics is an open problem.

De Jongh and Veltman proved the completeness of $\mathbf{IL}$, $\mathbf{IL}\mathsf{M}$ and $\mathbf{IL}\mathsf{P}$ w.r.t. the corresponding characteristic classes of ordinary (and finite) Veltman frames in [18]. As is usual for extensions of the provability logic **GL**, all completeness proofs suffer from compactness-related issues. One way to go about this is to define a (large enough) adequate set of formulas and let worlds be maximal consistent subsets of such sets (used e.g. in [18]). With interpretability logics and ordinary Veltman semantics, worlds have not been identified with (only) sets of formulas. It seems that with ordinary Veltman semantics it is sometimes necessary to duplicate worlds (that is, have more than one world correspond to a single maximal consistent set) in order to build models for certain consistent sets (see e.g. [18]). In [19], de Jongh and Veltman proved completeness of the logic **IL**W w.r.t. its characteristic class of ordinary (and finite) Veltman frames.

Goris and Joosten, inspired by Dick de Jongh, introduced a more robust approach to proving completeness of interpretability logics, the *construction method* [27, 28]. In this type of proofs, one builds models step by step[1], and the final model is retrieved as a union. While closer to the intuition and more informative than the standard proofs, these proofs are hard to produce and verify due to their size. (They might have been shorter if assuring labels—see the previous chapter or [11, 26]—have been used from the start.) For the purpose for which this type of proofs was invented (completeness of **IL**M_0 and **IL**W^* w.r.t. the ordinary semantics), this type of proofs is still the only known approach that works.

In [50] a very direct type of proofs of completeness is presented; similar to [18] in the general approach, but this time with respect to generalised Veltman semantics. The so-called *assuring labels* from [11, 26] were used as a key step (in this thesis, assuring labels are studied in the preceding chapter). These completeness proofs are the ones that we aim to explore here. An example that illustrates benefits of using the generalised semantics will be given in the subsection dedicated to **IL**M_0. The most interesting of these results are completeness of **IL**R and **IL**P_0. The principle R is important because it forms the basis of the, at the moment, best explicit candidate for **IL**(All). Results concerning the principle **IL**P_0 are interesting in a different way; they answer an old question: is there an unravelling technique that transforms generalised **IL**X-models to ordinary **IL**X-models, that preserves satisfaction of relevant characteristic properties? The answer is *no*: **IL**P_0 is complete w.r.t. generalised Veltman semantics, but it is known to be incomplete w.r.t. the ordinary semantics ([28]).

4.1 Preliminaries

In what follows, "formula" will always mean "modal formula". A maximal consistent set w.r.t. **IL**X will be called an **IL**X-MCS. We require the notion of *assuring labels* (Definition 3.4) which was originally introduced in [11] (see also [26]). However, here we use

[1] In [8] the term "step-by-step method" is coined for this type of proofs. Even before that, this was known as *Completeness by construction* in a reader from de Jongh and Veltman [20]. See also [7] and [21].

a slightly modified version of this notion. The difference is a new strategy of ensuring converse well-foundedness for the relation R. Instead of asking for the existence of some $\Diamond F \in w \setminus u$ whenever wRu, as is usual in the context of provability (and interpretability) logics, we will go for a stronger condition (see Definition 4.6). Since we will later put $R := \prec$, this choice of ours is reflected already at this point, in the definition of an assuring successor.

Definition 4.1 ([11], a slightly modified Definition 3.1) Let w and u be some **IL**X-MCS's, and let S be an arbitrary set of formulas. We write $w \prec_S u$ if for any finite $S' \subseteq S$ and any formula A we have that $A \rhd \bigvee_{G \in S'} \neg G \in w$ implies $\neg A, \Box\neg A \in u$.

We will also require the notation $w_S^{\Box}$ and $w_S^{\boxdot}$ (see Definition 3.7). Note that with the new definition of $\prec_S$, we have the following: $w \prec_S u$ if and only if $w_S^{\boxdot} \subseteq u$. Recall that since w is maximal consistent, use of $w_{\emptyset}^{\Box}$ usually amounts to the same as the use of the set $\{\Box A : \Box A \in w\}$.

We will usually write $w \prec u$ instead of $w \prec_{\emptyset} u$. This should not be confused with the standard notion of a predecessor/successor that can be found in the literature (in virtually all sources that deal with the relation between syntax and semantics of interpretability logics, including the previous chapter of this thesis), which usually requires an additional property that there is some $\Diamond F \in w \setminus u$ whenever $w \prec u$.

We will often require properties stated in lemmas 3.6 and 3.8. Here we emphasise that these properties hold with the new definition of assuringness. We will use these properties tacitly in the remainder of this chapter.

Lemma 4.2 ([11], Lemma 3.2) Let w, u and v be some **IL**X-MCS's, and let S and T be some sets of formulas. It follows that:

a) if $S \subseteq T$ and $w \prec_T u$, then $w \prec_S u$;

b) if $w \prec_S u \prec v$, then $w \prec_S v$;

c) if $w \prec_S u$, then $S \subseteq u$.

We need two lemmas that can be used to construct (or in our case, find) an MCS with the required properties. We already discussed them (see lemmas 3.20 and 3.21) in the previous chapter, but here we stress that they hold with our definition of the relation $\prec_S$:

Lemma 4.3 ([11], Lemma 3.4) Let w be an **IL**X-MCS, and let $\neg(B \rhd C) \in w$. Then there is an **IL**X-MCS u such that $w \prec_{\{\neg C\}} u$ and $B, \Box\neg B \in u$.

Lemma 4.4 ([11], Lemma 3.5) Let w and u be some **IL**X-MCS's such that $B \rhd C \in w$, $w \prec_S u$ and $B \in u$. Then there is an **IL**X-MCS v such that $w \prec_S v$ and $C, \Box\neg C \in v$.

We need a notion of adequacy which is mainly used to specify how far we want the truth lemma to stretch.

Definition 4.5 We say a set Γ is *adequate* if it is a finite set of formulas that is closed under taking subformulas and single negations, and $\top \in \Gamma$.

Any finite set of formulas can be extended to some adequate set.

In the remainder of this chapter, we will assume that $\mathcal{D}$ always stands for an adequate set. The following definition is central to most of the results of this chapter.

Definition 4.6 Let X be a subset of $\{\mathsf{M}, \mathsf{M_0}, \mathsf{P}, \mathsf{P_0}, \mathsf{R}\}$. We say that $\mathfrak{M} = (W, R, \{S_w : w \in W\}, V)$ is *the* $\mathbf{IL}\mathsf{X}$*-structure for a set of formulas* $\mathcal{D}$ if:

$$\begin{array}{rcl} W & := & \{w : w \text{ is an } \mathbf{IL}\mathsf{X}\text{-MCS and for some } G \in \mathcal{D}, G \wedge \Box\neg G \in w\}; \\ wRu & :\Leftrightarrow & w \prec u; \\ uS_wV & :\Leftrightarrow & wRu \text{ and, } V \subseteq R[w] \text{ and, } (\forall S)(w \prec_S u \Rightarrow (\exists v \in V) w \prec_S v); \\ w \in V(p) & :\Leftrightarrow & p \in w. \end{array}$$

We note that the $\mathbf{IL}\mathsf{X}$-structure for $\mathcal{D}$ is a unique object. In fact, we could work with just one "$\mathbf{IL}\mathsf{X}$-structure" (that would not depend even on $\mathcal{D}$): the disjoint union of $\mathbf{IL}\mathsf{X}$-structures for all choices of $\mathcal{D}$. We also observe that the definition entails that when uS_wV, then $V \neq \emptyset$ since wRu implies $w \prec_\emptyset u$, so there is $v \in V$ with $w \prec_\emptyset v$.

Notice that worlds in the definition above are somewhat more restricted than what is usually found in similar proofs: every world is required to be R-maximal with respect to some formula. That is, for every world $w \in W$ we want to have a formula G_w such that $w \Vdash G_w$ and for any R-successor u of w, $u \nVdash G_w$. This is equivalent to the requirement that for some formula G_w, $w \Vdash G_w \wedge \Box\neg G_w$. Of course, before we prove our truth lemma we can only require that $G_w \wedge \Box\neg G_w \in w$. Because of this we need the following lemma whose proof boils down to an instance of Löb's axiom.

Lemma 4.7 If $\mathbf{IL}\mathsf{X} \nvdash \neg A$ then there is an $\mathbf{IL}\mathsf{X}$-MCS w such that $A \wedge \Box\neg A \in w$.

Proof. We are to show that $\{A \wedge \Box\neg A\}$ is an $\mathbf{IL}\mathsf{X}$-consistent set. Suppose $A, \Box\neg A \vdash \bot$. It follows that $\vdash \Box\neg A \to \neg A$. Applying generalisation (necessitation) gives $\vdash \Box(\Box\neg A \to \neg A)$. The Löb axiom implies $\vdash \Box\neg A$. Now, $\vdash \Box\neg A$ and $A, \Box\neg A \vdash \bot$ imply $A \vdash \bot$, i.e. $\vdash \neg A$, a contradiction. $\dashv$

We are now ready to prove the main lemma of this section, which tells us that the structure defined in Definition 4.6 really is a generalised Veltman model. Notice that we do not claim that it is also an $\mathbf{IL}_{\mathrm{set}}\mathsf{X}$-model; we prove that later. Recall that if B is a formula, and w a world in a generalised Veltman model, we write $[B]_w$ for $\{u : wRu \text{ and } u \Vdash B\}$.

Lemma 4.8 Let X be a subset of $\{\mathsf{M}, \mathsf{M_0}, \mathsf{P}, \mathsf{P_0}, \mathsf{R}\}$. The $\mathbf{IL}\mathsf{X}$-structure $\mathfrak{M}$ for a set of formulas $\mathcal{D}$ is a generalised Veltman model. Furthermore, the following truth lemma holds:

$$\mathfrak{M}, w \Vdash G \text{ if and only if } G \in w,$$

for all $G \in \mathcal{D}$ and $w \in W$.

Proof. Let us verify that the **IL**X-structure $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ for $\mathcal{D}$ is a generalised Veltman model. Since $\mathbf{IL}X \nvdash \bot$ and $\top \in \mathcal{D}$, Lemma 4.7 implies $W \neq \emptyset$. Transitivity of R is immediate. To see converse well-foundedness, assume there are more than $|\mathcal{D}|$ worlds in an R-chain. Then there are x and y with xRy and for some $G \in \mathcal{D}$, $G, \Box\neg G \in x, y$. However, $\Box\neg G \in x$ and $G \in y$ obviously contradict the assumption that xRy ($x \prec y$). Next, let us prove the properties of S_w for $w \in W$. Clearly $S_w \subseteq R[w] \times \mathcal{P}(R[w])$. If xS_wV, then $w \prec_\emptyset x$ implies there is at least one element v in V (with $w \prec_\emptyset v$). Quasi-reflexivity and monotonicity are obvious. Next, assume $wRxRu$ and $w \prec_S x$. Lemma 4.2 and $w \prec_S x \prec u$ imply $w \prec_S u$. Thus, $xS_w\{u\}$. It remains to prove quasi-transitivity. Assume xS_wV and vS_wU_v for all $v \in V$. Put $U = \bigcup_v U_v$. We claim that xS_wU. We have $U \subseteq R[w]$. Assume $w \prec_S x$. This and xS_wV imply there is $v \in V$ such that $w \prec_S v$. This and vS_wU_v imply there is $u \in U_v$ (thus also $u \in U$) such that $w \prec_S u$. Let us prove the truth lemma with respect to the formulas contained in $\mathcal{D}$. The claim is proved by induction on the complexity of $G \in \mathcal{D}$. We will only consider the case $G = B \rhd C$. Assume $B \rhd C \in w$, wRu and $u \Vdash B$. Induction hypothesis implies $B \in u$. We claim that $uS_w[C]_w$. Clearly $[C]_w \subseteq R[w]$. Assume $w \prec_S u$. Lemma 4.4 implies there is an **IL**X-MCS v with $w \prec_S v$ and $C, \Box\neg C \in v$ (thus also wRv and $v \in W$). Induction hypothesis implies $\mathfrak{M}, v \Vdash C$. To prove the converse, assume $B \rhd C \notin w$. Lemma 4.3 implies there is u with $w \prec_{\{\neg C\}} u$ and $B, \Box\neg B \in u$ (thus $u \in W$). It is immediate that wRu and the induction hypothesis implies that $u \Vdash B$. Assume uS_wV. We are to show that $V \nVdash C$. Since $w \prec_{\{\neg C\}} u$ and uS_wV, there is $v \in V$ such that $w \prec_{\{\neg C\}} v$. Lemma 4.2 implies $\neg C \in v$. The induction hypothesis implies $v \nVdash C$; thus $V \nVdash C$. ⊣

Theorem 4.9 Let $X \subseteq \{\mathsf{M}, \mathsf{M_0}, \mathsf{P}, \mathsf{P_0}, \mathsf{R}\}$. Assume that for every set $\mathcal{D}$ the **IL**X-structure for $\mathcal{D}$ possesses the property $(\mathsf{X})_{\mathrm{gen}}$. Then **IL**X is complete w.r.t. $\mathbf{IL}_{\mathrm{set}}$X-models.

Proof. Let A be a formula such that $\mathbf{IL}X \nvdash \neg A$. Lemma 4.7 implies there is an **IL**X-MCS w such that $A \wedge \Box\neg A \in w$. Let $\mathcal{D}$ have the usual properties, and contain A. Let $\mathfrak{M} = (W, R, \{S_w : w \in W\}, V)$ be the **IL**X-structure for $\mathcal{D}$. Since $A \wedge \Box\neg A \in w$ and $A \in \mathcal{D}$, we have $w \in W$. Lemma 4.8 implies $\mathfrak{M}, w \nVdash \neg A$. ⊣

The logic **IL** is the logic of all Veltman frames (both ordinary or generalised). Thus we immediately have the following:

Corollary 4.10 The logic **IL** is complete w.r.t. generalised Veltman semantics.

In the next section we comment on the completeness of the following logics w.r.t. generalised Veltman semantics: **IL**M, **IL**$\mathsf{M_0}$, **IL**P, **IL**$\mathsf{P_0}$, **IL**R, **IL**W and **IL**W*.

In the end we will explore the logic **IL**WR, and another proof for **IL**P. The logic **IL**WR is explored in more detail in the last chapter of this thesis.

4.1.1 A note on generalised Veltman semantics and labelling

Labels are primarily a syntactic notion. However, there is an aspect of their interaction with generalised semantics that is not present when working with ordinary semantics. In this subsection we comment on this—there are no directly usable implications obtained here; still, we think it's an interesting property both of labelling and generalised semantics.

In all studied extensions of **IL** we have to duplicate maximal consistent sets when building ordinary Veltman models for consistent sets of formulas. More accurately, no one seems to have come up with a natural way of assigning just one purpose to every maximal consistent set of formulas. For example, when building a model where $\{\neg(p \rhd q), \neg(p \rhd r), p \rhd (q \vee r)\}$ is true in some world w, we could try to use the same set/world u visible from w as a witness for the formulas $\neg(p \rhd q)$ and $\neg(p \rhd r)$ in w. For example, this may be the set where the only propositional formula is p, and no formula of form $\neg(A \rhd B)$ is contained. But, due to $p \rhd (q \vee r)$, in any model where w is present, we do require two worlds like u within that model. One of these worlds will have an S_w-successor satisfying q but not r, and the other one an S_w-successor satisfying r but not q.

Generalised Veltman semantics doesn't share this problem of duplication, at least not in any known case of a complete extension of **IL**. A generalised model for the problem above is simple. Let $w = \{\neg(p \rhd q), \neg(p \rhd r), p \rhd (q \vee r)\}$, $u = \{p\}$, $x = \{q\}$, $y = \{r\}$, and let $wRuS_w\{x, y\}$. Unspecified propositional formulas are assumed to be false, and unspecified $\rhd$-formulas are assumed to be true.

Now, having in mind this generalised model, what can be said about the wRu transition in terms of labels? This might be important if we are building a generalised model step-by-step. Since u has two roles, it would be natural to allow (even with assuringness) two labels: $\{\neg q\}$ and $\{\neg r\}$. Note that the S_w from Definition 4.6 indeed takes multiple labels into account. And these labels are justified, since indeed $\{x, y\} \nVdash q, r$. Both these labels are expressible without sets (in terms of criticality, for example, the labels would be formulas q, and r, respectively).

However, there is another bit of label-related information that these facts do not express: which labels *do not* hold. Although $\{\neg q\}$ and $\{\neg r\}$ are justified choices, the label $\{\neg q, \neg r\}$ is not a good choice. This label would require $\neg p \notin u$, which is clearly not the case. This is the information the assuringness allows us to express, and criticality does not. Granted, one might say that the inadequacy of the assuring label $\{\neg q, \neg r\}$ is equivalent to the inadequacy of the critical label $q \vee r$. However, expressing this fact in terms of criticality does not retain structural information of our situation; we see a disjunction where really we are only interested in disjuncts. Note that such a situation cannot happen in ordinary semantics: if the label $\{\neg q, \neg r\}$ is inappropriate for some wRu transition, that means there is $A \in u$ with $A \rhd q \vee r \in w$. This, since we are now working in ordinary semantics, means there should be an S_w-successor of u satisfying $q \vee r$. So, either this new world satisfies q or r. So, $\{\neg q\}$ or $\{\neg r\}$ had to be inappropriate labels (for wRu) too.

4.2 Completeness results

We have seen an easy proof that **IL** is complete w.r.t. generalised Veltman semantics. In this section we explore completeness proofs for various extensions of **IL**.

Note that we can always reuse completeness results for ordinary semantics (when such result exists), as the following proposition shows. In this proposition we reuse the construction $gen(\mathfrak{M})$ from [73].

Definition 4.11 ([73]) Let $\mathfrak{M} = (W, R, S, V)$ be an arbitrary ordinary Veltman model. We define a generalised model $gen(\mathfrak{M}) = (W, R, S', V)$ where uS'_wV if and only if $V \subseteq R[w]$ and for some $v \in V$ we have uS_wv.

Proposition 4.12 Let X be a (possibly empty) set of arbitrary modal formulas. If the logic **IL**X is complete w.r.t. ordinary Veltman semantics, it is also complete w.r.t. generalised Veltman semantics.

Proof. Suppose **IL**$\mathsf{X} \nvdash A$. Then there is an **IL**X-model $\mathfrak{M} = (W, R, S, V)$ and $w \in \mathfrak{M}$ such that $w \nVdash A$. Let $\mathfrak{M}' = gen(\mathfrak{M})$. See [70] for the proof that $\mathfrak{M}'$ really is a generalised Veltman model and that the truth values are preserved everywhere.

It remains to check if the frame $\mathfrak{F}'$ of $\mathfrak{M}'$ is an $\mathbf{IL}_{\text{set}}\mathsf{X}$-frame. By definition, this is the case if and only if for all theorems A of **IL**X we have $\mathfrak{F}' \Vdash A$. Note that for all theorems A of **IL**X we have $\mathfrak{F} \Vdash A$, where $\mathfrak{F}$ is the frame of $\mathfrak{M}$. Take an arbitrary valuation $U \subseteq W \times \mathsf{Prop}$ (Prop is the set of all propositional variables) and we claim that $(\mathfrak{F}', U) \Vdash A$. Clearly $(\mathfrak{F}, U) \Vdash A$. Now, the same construction as above (with U instead of V) results in the model $(\mathfrak{F}', U)$ (the definition of $\mathfrak{F}'$ does not depend on a valuation). And we have seen that truth values coincide, so $(\mathfrak{F}', U) \Vdash A$, as required. $\dashv$

See Subsection 4.2.4 regarding the failure of the converse of the preceding claim.

4.2.1 The logic ILM

Completeness of the logic **IL**M w.r.t. generalised Veltman semantics is an easy consequence (see Proposition 4.12) of the completeness of **IL**M w.r.t. the ordinary semantics, first proved by de Jongh and Veltman ([18]). Another proof of the same result was given by Goris and Joosten, using the construction method ([28, 36]).

The frame condition $wRxS_wyRz \Rightarrow xRz$ for M is reflected in its labelling lemma, Lemma 3.24:

R. Verbrugge determined the characteristic property $(\mathsf{M})_{\text{gen}}$ in [62]:

$$uS_wV \Rightarrow (\exists V' \subseteq V)(uS_wV' \;\&\; R[V'] \subseteq R[u]).$$

The following lemma holds for our definition of $\prec_S$ too.

Lemma 4.13 ([11], Lemma 3.7) Let w and u be some **IL**M-MCS's, and let S be a set of formulas. If $w \prec_S u$ then $w \prec_{S \cup u_\emptyset^\Box} u$.

When we combine this with the main result of the previous section we get a simple and succinct completeness proof.

Theorem 4.14 The logic **IL**M is complete w.r.t. $\mathbf{IL}_{\text{set}}$M-models.

Proof. Given Theorem 4.9, it suffices to show that for any set $\mathcal{D}$, the **IL**M-structure for $\mathcal{D}$ possesses the property $(\mathsf{M})_{\text{gen}}$: $uS_wV \Rightarrow (\exists V' \subseteq V)(uS_wV'$ & $R[V'] \subseteq R[u]$. Let $(W, R, \{S_w : w \in W\}, V)$ be the **IL**M-structure for $\mathcal{D}$.

Let uS_wV and take $V' = \{v \in V : w \prec_{u_\emptyset^\Box} v\}$. We claim uS_wV' and $R[V'] \subseteq R[u]$. Suppose $w \prec_S u$. Lemma 4.13 implies $w \prec_{S \cup u_\emptyset^\Box} u$. Since uS_wV, by Definition 4.6, there is $v \in V$ with $w \prec_{S \cup u_\emptyset^\Box} v$. So, $v \in V'$. Thus, uS_wV'.

Now let $v \in V'$ and $z \in W$ be such that vRz. Since $v \in V'$, we know $w \prec_{u_\emptyset^\Box} v$. Then for all $\Box B \in u$ we have $\Box B \in v$. Since vRz, we have $B, \Box B \in z$. So, $u \prec z$ and by Definition 4.6 uRz. ⊣

4.2.2 The logic $\mathrm{IL}\mathsf{M}_0$

Modal completeness of $\mathbf{IL}\mathsf{M}_0$ w.r.t. ordinary Veltman semantics was proved in [27] by Goris and Joosten. Certain difficulties encountered in this proof were our main motivation for using generalised Veltman semantics. We will sketch one of these difficulties and show in what way the generalised semantics overcomes it. Characteristic property $(\mathsf{M}_0)_{\text{gen}}$ (see [49]):

$$wRuRxS_wV \ \Rightarrow\ (\exists V' \subseteq V)(uS_wV' \ \&\ R[V'] \subseteq R[u])).$$

The frame condition $wRxRyS_wuRz \Rightarrow xRz$ for M_0 is reflected in its labelling lemma, Lemma 3.25, which holds for our definition of $\prec_S$ too:

Lemma 4.15 ([11], Lemma 3.9) Let w, u and x be $\mathbf{IL}\mathsf{M}_0$-MCS's, and S an arbitrary set of formulas. If $w \prec_S u \prec x$ then $w \prec_{S \cup u_\emptyset^\Box} x$.

To motivate our way of proving completeness (of $\mathbf{IL}\mathsf{M}_0$, but also in general) w.r.t. generalised Veltman semantics, let us sketch a situation for which there are clear benefits in working with generalised Veltman semantics. We do this only now because $\mathbf{IL}\mathsf{M}_0$ is sufficiently complex to display some of these benefits. Suppose we are building models step-by-step (as in the *construction method* from [27]), and worlds w, u_1, u_2 and x occur in the configuration displayed in Figure 4.1. Furthermore, suppose we need to produce an S_w-successor v of x.

With the ordinary semantics, we need to ensure that for our S_w-successor v, for each $\Box B_1 \in u_1$ and $\Box B_2 \in u_2$, we have $\Box B_1, \Box B_2 \in v$. It is not obvious that such a construction is possible. In case of $\mathbf{IL}\mathsf{M}_0$, it was successfully solved in [27] by preserving the invariant

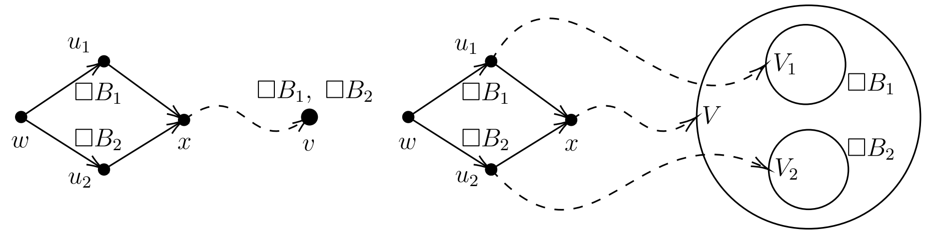


Figure 4.1: Left: extending an ordinary Veltman model. Right: extending a generalised Veltman model. Straight lines represent R-transitions, while curved lines represent S_w-transitions. Full lines represent the starting configuration, and dashed lines represent the transitions that are to be added. This figure is also taken from [50].

that sets of boxed formulas in u_i are linearly ordered. This way, finite (quasi-)models can always be extended by only taking the last u_i into consideration.

With generalised Veltman semantics, we need to produce a whole set of worlds V, but the requirements from the frame condition $wRuRxS_wV \ \Rightarrow\ (\exists V' \subseteq V)(uS_wV' \ \&\ R[V'] \subseteq R[u])$ on each particular world are less demanding. For each u_i, there has to be a corresponding $V_i \subseteq V$ with $\Box B_i$ contained (true) in every world of V_i. Lemma 4.15 gives a recipe for producing such worlds.

Theorem 4.16 The logic $\mathbf{IL}\mathsf{M}_0$ is complete w.r.t. $\mathbf{IL}_{\mathrm{set}}\mathsf{M}_0$-models.

Proof. Given Theorem 4.9, it suffices to show that for any set $\mathcal{D}$, the $\mathbf{IL}\mathsf{M}_0$-structure for $\mathcal{D}$ possesses the property $(\mathsf{M}_0)_{\mathrm{gen}}$. Let $(W, R, \{S_w : w \in W\}, \Vdash)$ be the $\mathbf{IL}\mathsf{M}_0$-structure for $\mathcal{D}$. Assume $wRuRxS_wV$ and take $V' = \{v \in V : w \prec_{u^{\Box}_{\emptyset}} v\}$. We claim that uS_wV' and $R[V'] \subseteq R[u]$. Obviously $V' \subseteq V \subseteq R[w]$. Assume $w \prec_S u$. Lemma 4.15 and $w \prec_S u \prec x$ imply $w \prec_{S\cup u^{\Box}_{\emptyset}} x$. Now xS_wV and the definition of S_w imply there is $v \in V$ such that $w \prec_{S\cup u^{\Box}_{\emptyset}} v$. Lemma 4.2 implies $w \prec_{u^{\Box}_{\emptyset}} v$. So, $v \in V'$. It remains to verify that $R[V'] \subseteq R[u]$. Let $v \in V'$ and $z \in W$ be worlds such that vRz. Since $w \prec_{u^{\Box}_{\emptyset}} v$, for all $\Box B \in u$ we have $\Box B \in v$, and since vRz, it follows that $\Box B, B \in z$. Thus, $u \prec z$ i.e. uRz. $\dashv$

4.2.3 The logic ILP

As in the case of the logic $\mathbf{IL}\mathsf{M}$, the completeness of $\mathbf{IL}\mathsf{P}$ w.r.t. the generalised semantics is an easy consequence of the completeness of $\mathbf{IL}\mathsf{P}$ w.r.t. the ordinary semantics, first proved by de Jongh and Veltman [18]. Verbrugge determined the characteristic property $(\mathsf{P})_{\mathrm{gen}}$ in [62]:

$$wRw'RuS_wV \Rightarrow (\exists V' \subseteq V)\ uS_{w'}V'.$$

The labelling lemma, Lemma 3.22, holds for our definition of $\prec_S$ too.

Lemma 4.17 ([11] Lemma 3.8) Let w, x and u be some **IL**P-MCS's, and let S and T be arbitrary sets of formulas. If $w \prec_S x \prec_T u$ then $w \prec_{S \cup x^{\boxdot}_T} u$.

Theorem 4.18 The logic **IL**P is complete w.r.t. $\mathbf{IL}_{\mathrm{set}}\mathsf{P}$-models.

Proof. Given Theorem 4.9, it suffices to show that for any set $\mathcal{D}$, the **IL**P-structure for $\mathcal{D}$ possesses the property $(\mathsf{P})_{\mathrm{gen}}$. Let $(W, R, \{S_w : w \in W\}, \Vdash)$ be the **IL**P-structure for $\mathcal{D}$. Let $wRw'RuS_wV$ and take $V' = V \cap R[w']$. We claim $uS_{w'}V'$. Let T be arbitrary such that $w' \prec_T u$. Lemma 4.17 and $w \prec_{\emptyset} w' \prec_T u$ imply $w \prec_{w'^{\boxdot}_T} u$. Now, uS_wV implies that there is a $v \in V$ with $w \prec_{w'^{\boxdot}_T} v$. Let $A \rhd \neg \bigwedge T' \in w'$ for some finite $T' \subseteq T$. Then $\neg A, \Box\neg A \in w'^{\boxdot}_T$. Lemma 4.2 and $w \prec_{w'^{\boxdot}_T} v$ imply $\neg A, \Box\neg A \in v$. Thus, $w' \prec_T v$. Finally, $V' \subseteq R[w']$ holds by assumption, thus $uS_{w'}V'$. ⊣

4.2.4 The logic ILP$_0$

The interpretability principle $\mathsf{P}_0 = A \rhd \Diamond B \to \Box(A \rhd B)$ is introduced in J. Joosten's master thesis in 1998. In [28] it is shown that the interpretability logic **IL**P_0 is incomplete w.r.t. Veltman models. Since we will show that **IL**P_0 is complete w.r.t. the generalised semantics, this is the first example of an interpretability logic complete w.r.t. the generalised semantics, but incomplete w.r.t. the ordinary semantics. Characteristic property $(\mathsf{P}_0)_{\mathrm{gen}}$ was determined in [28]. A slightly reformulated version:

$$wRxRuS_wV \ \&\ (\forall v \in V)R[v] \cap Z \neq \emptyset \ \Rightarrow\ (\exists Z' \subseteq Z)uS_xZ'.$$

The following technical lemma is almost obvious.

Lemma 4.19 Let x be an **IL**X-MCS, A a formula, and T a finite set of formulas. Let B_G be an arbitrary formula, and T_G an arbitrary finite set of formulas, for every $G \in T$. Furthermore, assume:

a) $A \rhd \bigvee_{G \in T} B_G \in x$;

b) $(\forall G \in T)\ B_G \rhd \bigvee_{H \in T_G} \neg H \in x$.

Then we have $A \rhd \bigvee_{H \in S'} \neg H \in x$, where $S' = \bigcup_{G \in T} T_G$.

Proof. Let $G \in T$. Since $T_G \subseteq S'$, clearly $\vdash \bigvee_{H \in T_G} \neg H \rhd \bigvee_{H \in S'} \neg H$. The requirement b) and the axiom (J2) imply $B_G \rhd \bigvee_{H \in S'} \neg H \in x$. Now $|T| - 1$ applications of the axiom (J3) give $\bigvee_{G \in T} B_G \rhd \bigvee_{H \in S'} \neg H \in x$. Finally, apply the requirement a) and the axiom (J2). ⊣

Next we need a labelling lemma for **IL**P_0. This is where we use the technical lemma above.

Lemma 4.20 Let w, x and u be some $\mathbf{IL}\mathsf{P_0}$-MCS's, and let S be a set of formulas. If $w \prec x \prec_S u$ then $w \prec_{x_S^\Box} u$.

Proof. Let A be an arbitrary formula. Let $T \subseteq x_S^\Box$ be a finite set such that $A \rhd \bigvee_{G\in T} \neg G \in w$. We will prove that $\neg A, \Box\neg A \in u$. If $G \in T$ $(\subseteq x_S^\Box)$, then $G = \Box\neg B_G$, for some formula B_G. Thus $A \rhd \bigvee_{G\in T} \neg\Box\neg B_G \in w$, and by easy inferences and maximal consistency: $A \rhd \bigvee_{G\in T} \Diamond B_G \in w$, and $A \rhd \Diamond \bigvee_{G\in T} B_G \in w$. Applying $\mathsf{P_0}$ gives $\Box(A \rhd \bigvee_{G\in T} B_G) \in w$. The assumption $w \prec x$ implies $A \rhd \bigvee_{G\in T} B_G \in x$. For each $G \in T$ $(\subseteq x_S^\Box)$ there is a finite subset T_G of S such that $B_G \rhd \bigvee_{H\in T_G} \neg H \in x$. Let $S' = \bigcup_{G\in T} T_G$. Clearly S' is a finite subset of S. Lemma 4.19 implies $A \rhd \bigvee_{H\in S'} \neg H \in x$. Finally, $S' \subseteq S$ and the assumption $x \prec_S u$ imply $\neg A, \Box\neg A \in u$. ⊣

The following simple observation is useful both for $\mathbf{IL}\mathsf{P_0}$ and $\mathbf{IL}\mathsf{R}$.

Lemma 4.21 Let w, x, v and z be some $\mathbf{IL}\mathsf{X}$-MCS's, and let S be a set of formulas. If $w \prec_{x_S^\Box} v \prec z$ then $x \prec_S z$.

Proof. Let S' be a finite subset of S with $A \rhd \bigvee_{G\in S'} \neg G \in x$. Then $\Box\neg A \in x_S^\Box$. Now $w \prec_{x_S^\Box} v$ and Lemma 4.2 imply $\Box\neg A \in v$. Since $v \prec z$, we have $\neg A, \Box\neg A \in z$. ⊣

Theorem 4.22 The logic $\mathbf{IL}\mathsf{P_0}$ is complete w.r.t. $\mathbf{IL}_{\rm set}\mathsf{P_0}$-frames.

Proof. Given Theorem 4.9, it suffices to show that for any set $\mathcal{D}$, the $\mathbf{IL}\mathsf{P_0}$-structure for $\mathcal{D}$ possesses the property $(\mathsf{P_0})_{\rm gen}$. Let $(W, R, \{S_w : w \in W\}, \Vdash)$ be the $\mathbf{IL}\mathsf{P_0}$-structure for $\mathcal{D}$. Assume $wRxRuS_wV$ and $R[v] \cap Z \neq \emptyset$ for each $v \in V$. We will prove that there is $Z' \subseteq Z$ such that uS_xZ'. Let S be a set of formulas such that $w \prec x \prec_S u$. Lemma 4.20 implies $w \prec_{x_S^\Box} u$. Since uS_wV, there is $v \in V$ such that $w \prec_{x_S^\Box} v$. Since $R[v] \cap Z \neq \emptyset$, choose a world $z_S \in R[v] \cap Z$. Now $w \prec_{x_S^\Box} v \prec z_S$ and Lemma 4.21 imply $x \prec_S z_S$. Put $Z' = \{z_S : S \text{ is a set of formulas such that } x \prec_S u\}$. Clearly $Z' \subseteq Z$. So, $Z' \subseteq R[x]$, and since for each set S such that $x \prec_S u$ we have $x \prec_S z_S$, it follows that uS_xZ'. ⊣

In [62] and [73] a possibility was explored of transforming a generalised Veltman model to an ordinary Veltman model, such that these two models are bisimilar (in some aptly defined sense). A natural question is whether such transformation exists if we add the requirement that characteristic properties are preserved. The example of $\mathbf{IL}\mathsf{P_0}$ shows that there are $\mathbf{IL}_{\rm set}\mathsf{P_0}$-models with no (bisimilar or otherwise) counterpart $\mathbf{IL}\mathsf{P_0}$-models.

4.2.5 The logic ILR

Completeness of $\mathbf{IL}\mathsf{R}$ w.r.t. ordinary Veltman semantics is an open problem (see [11]). In this subsection we will prove that $\mathbf{IL}\mathsf{R}$ is complete w.r.t. the generalised semantics. Characteristic property $(\mathsf{R})_{\rm gen}$ was determined in [28]. A slightly reformulated version:

$$wRxRuS_wV \Rightarrow (\forall C \in \mathcal{C}(x,u))(\exists U \subseteq V)(xS_wU \ \&\ R[U] \subseteq C),$$

where $\mathcal{C}(x,u) = \{C \subseteq R[x] : (\forall Z)(uS_xZ \Rightarrow Z \cap C \neq \emptyset)\}$ is the family of "choice sets".

The following lemma holds for our definition of $\prec_S$ too.

Lemma 4.23 ([11], Lemma 3.10) Let w, x and u be some **IL**R-MCS's, and let S and T be arbitrary sets of formulas. If $w \prec_S x \prec_T u$ then $w \prec_{S \cup x_T^\Box} u$.

Theorem 4.24 The logic **IL**R is complete w.r.t. $\mathbf{IL}_{\mathrm{set}}$R-models.

Proof. Given Theorem 4.9, it suffices to show that for any set $\mathcal{D}$, the **IL**R-structure for $\mathcal{D}$ possesses the property $(\mathsf{R})_{\mathrm{gen}}$. Let $(W, R, \{S_w : w \in W\}, \Vdash)$ be the **IL**R-structure for $\mathcal{D}$. Assume $wRxRuS_wV$ and $C \in \mathcal{C}(x,u)$. We are to show that $(\exists U \subseteq V)(xS_wU \;\&\; R[U] \subseteq C)$. We will first prove an auxiliary claim:

$$(\forall S)\Big(w \prec_S x \;\Rightarrow\; (\exists v \in V)(w \prec_{S \cup x_\emptyset^\Box} v \;\&\; R[v] \subseteq C)\Big).$$

So, let S be arbitrary such that $w \prec_S x$, and suppose (for a contradiction) that for every $v \in V$ with $w \prec_{S \cup x_\emptyset^\Box} v$, we have $R[v] \nsubseteq C$, that is, there is some $z_v \in R[v] \setminus C$. Let $Z = \{z_v : v \in V, w \prec_{S \cup x_\emptyset^\Box} v\}$. We claim that uS_xZ. Let T be arbitrary such that $x \prec_T u$, and we should prove that there exists $z \in Z$ such that $x \prec_T z$. From $w \prec_S x \prec_T u$ and Lemma 4.23 it follows that $w \prec_{S \cup x_T^\Box} u$. Since uS_wV, there is $v \in V$ with $w \prec_{S \cup x_T^\Box} v$. Now, $x_\emptyset^\Box \subseteq x_T^\Box$ and Lemma 4.2 imply $w \prec_{S \cup x_\emptyset^\Box} v$, so there is a world $z_v \in Z$ as defined earlier. Furthermore, $w \prec_{x_T^\Box} v \prec z_v$ and Lemma 4.21 imply $x \prec_T z_v$. To prove uS_xZ it remains to verify that $Z \subseteq R[x]$. Let $z_v \in Z$ be arbitrary and apply Lemma 4.2 and Lemma 4.21 as before. Now, uS_xZ and $C \in \mathcal{C}(x,u)$ imply $C \cap Z \neq \emptyset$, contradicting the definition of Z. This concludes the proof of the auxiliary claim. Let $U = \{v \in V : w \prec_{x_\emptyset^\Box} v \text{ and } R[v] \subseteq C\}$. Auxiliary claim implies $U \neq \emptyset$. If $w \prec_S x$, auxiliary claim implies there is $v \in U$ such that $w \prec_{S \cup x_\emptyset^\Box} v$ and $R[v] \subseteq C$, so $v \in U$. Thus xS_wU. It is clear that $R[U] \subseteq C$. ⊣

4.2.6 The logics ILW and ILW*

To prove that **IL**W is complete, one could try to find a sufficiently strong "labelling lemma" and use Definition 4.6 (**IL**X-structure). One candidate might be the following condition:

$$w \prec_S u \;\Rightarrow\; (\exists G \in \mathcal{D})\,\Big(w \prec_{S \cup \{\Box\neg G\}} u \;\&\; G \in u\Big),$$

where $\mathcal{D}$ is finite, closed under subformulas and such that each $w \in W$ contains A_w and $\Box\neg A_w$ for some $A_w \in \mathcal{D}$. If there is such a condition, it would greatly simplify proofs of completeness for extensions of **IL**W. Unfortunately, at the moment we do not know if such a condition can be formulated and proved.

Another approach is to use a modified version of Definition 4.6 to work with **IL**W and its extensions. This way we won't require a labelling lemma, but we lose generality in the following sense. To prove the completeness of **IL**XW, for some X, it no longer suffices to simply show that the structure defined in Definition 4.6 has the required characteristic property (when each world is an **IL**X-MCS). Instead, the characteristic property of **IL**X has to be shown to hold on the modified structure. So, to improve compatibility with proofs based on Definition 4.6, we should prove the completeness of **IL**W with a definition as similar to Definition 4.6 as possible. That is what we do in the remainder of this section. This approach turns out to be good enough for **IL**W* (**IL**WM$_0$). We didn't succeed in using it to prove the completeness of **IL**WR. However, to the best of our knowledge, **IL**WR might not be complete at all.

We use the following condition $(\mathsf{W})_{\text{gen}}$ (the "positive" version):

$$(\mathsf{W})_{\text{gen}} \ := \ uS_wV \ \Rightarrow \ (\exists V' \subseteq V) \left(uS_wV' \ \& \ R[V'] \cap S_w^{-1}[V] = \emptyset \right).$$

In what follows, **IL**WX denotes an arbitrary extension of **IL**W.

We will use the following two lemmas, which we already discussed near the end of the previous chapter (lemmas 3.28 and 3.29). Here we emphasise that they hold for our definition of $\prec_S$ too.

Lemma 4.25 ([11], Lemma 3.12) Let w be an **IL**WX-MCS, and B and C formulas such that $\neg(B \rhd C) \in w$. Then there is an **IL**WX-MCS u such that $w {\prec_{\{\Box\neg B, \neg C\}}} u$ and $B \in u$.

Lemma 4.26 ([11], Lemma 3.13) Let w and u be some **IL**WX-MCS, B and C some formulas, and S a set of formulas such that $B \rhd C \in w$, $w \prec_S u$ and $B \in u$. Then there is an **IL**WX-MCS v such that $w \prec_{S \cup \{\Box\neg B\}} v$ and $C, \Box\neg C \in v$.

Given a binary relation R, let $\dot{R}[x] = R[x] \cup \{x\}$.

When defining S_w we have to take care to make it compatible with the properties of a generalised Veltman model, in particular, the property that wRu implies $uS_w\{u\}$ and the property that $wRuRv$ implies $uS_w\{v\}$. So, if we fix w and u, we should have $uS_w\{v\}$ for all $v \in \dot{R}[u] (= R[u] \cup \{u\})$. However, because of monotonicity, we want not only $uS_w\{v\}$ in such cases, but also uS_wV for all $V \subseteq R[v]$ that contain v. This is why we add the condition (a) in the definition below (within the definition of S_w).

If the set $\dot{R}[x]$ contains maximal consistent sets (which it usually does in this section), then $\bigcup \dot{R}[x]$ is a set of formulas. If satisfaction coincides with formulas contained, then it is useful to think of $\bigcup \dot{R}[x]$ as the set of formulas B such that either B or $\Diamond B$ is satisfied in x (however, one has to be careful with such an interpretation, since we do not claim our truth lemma to hold for all formulas).

Definition 4.27 Let X be W or W*. We say that $\mathfrak{M} = (W, R, \{S_w : w \in W\}, V)$ is *the*

ILX*-structure for a set of formulas* $\mathcal{D}$ if:

$$\begin{array}{rcl} W & := & \{w : w \text{ is an } \mathbf{IL}\mathsf{X}\text{-MCS and for some } G \in \mathcal{D}, G \wedge \Box\neg G \in w\}; \\ wRu & :\Leftrightarrow & w \prec u; \\ uS_wV & :\Leftrightarrow & wRu \text{ and, } V \subseteq R[w] \text{ and, one of the following holds:} \\ & & (a)\ V \cap \dot{R}[u] \neq \emptyset; \\ & & (b)\ (\forall S)\left(w \prec_S u \Rightarrow (\exists v \in V)\left(\exists G \in \mathcal{D} \cap \bigcup \dot{R}[u]\right)\ w \prec_{S\cup\{\Box\neg G\}} v\right); \\ w \in V(p) & :\Leftrightarrow & p \in w. \end{array}$$

With this definition, we can now prove a truth lemma.

Lemma 4.28 Let X be W or W*. The **IL**X-structure $\mathfrak{M}$ for $\mathcal{D}$ is a generalised Veltman model. Furthermore, the following holds:

$$\mathfrak{M}, w \Vdash G \quad \text{if and only if} \quad G \in w,$$

for each $G \in \mathcal{D}$ and $w \in W$.

Proof. Let us first verify that the **IL**X-structure $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ for $\mathcal{D}$ is a generalised Veltman model. All the properties, except for quasi-transitivity, have easy proofs (see the proof of Lemma 4.8). Let us prove the quasi-transitivity. Assume uS_wV, and vS_wU_v for all $v \in V$. Put $U = \bigcup_{v\in V} U_v$. We claim that uS_wU. Clearly $U \subseteq R[w]$. To prove uS_wU we will distinguish the cases (a) and (b) from the definition of the relation S_w for uS_wV. In the case (a), we have $v_0 \in V$ for some $v_0 \in \dot{R}[u]$. We will next distinguish two cases from the definition of $v_0S_wU_{v_0}$. In the case (aa) we have $x \in U_{v_0}$ for some $x \in \dot{R}[v_0]$. Since $v_0 \in \dot{R}[u]$, we then have $x \in \dot{R}[u]$. Since $x \in U_{v_0} \subseteq U$, then $U \cap \dot{R}[u] \neq \emptyset$. So, we have uS_wU, as required. In the case (ab) we have:

$$(\forall S)(w \prec_S v_0 \Rightarrow (\exists x \in U_{v_0})(\exists G \in \mathcal{D} \cap \bigcup \dot{R}[v_0])\ w \prec_{S\cup\{\Box\neg G\}} x).$$

To prove uS_wU in this case, we will use the case (b) from the definition of the relation S_w. Assume $w \prec_S u$. Then we have $w \prec_S u \prec v_0$ or $w \prec_S u = v_0$. Either way, possibly using Lemma 4.2, we have $w \prec_S v_0$, and so there are $x \in U_{v_0}$ and $G \in \mathcal{D} \cap \bigcup \dot{R}[v_0]$ with $w \prec_{S\cup\{\Box\neg G\}} x$. Since uRv_0 or $u = v_0$, we have $\dot{R}[v_0] \subseteq \dot{R}[u]$. So, the claim follows. In the case (b), we have:

$$(\forall S)(w \prec_S u \Rightarrow (\exists v \in V)(\exists G \in \mathcal{D} \cap \bigcup \dot{R}[u])\ w \prec_{S\cup\{\Box\neg G\}} v).$$

To prove uS_wU we will use the case (b) from the definition of the relation S_w. Assume $w \prec_S u$. Then there are $v_0 \in V$ and $G \in \mathcal{D} \cap \bigcup \dot{R}[u]$ such that $w \prec_{S\cup\{\Box\neg G\}} v_0$. From $v_0 \in V$ it follows that $v_0S_wU_{v_0}$. We will next distinguish between the possible cases in the definition of $v_0S_wU_{v_0}$. In the first case (ba) we have $U_{v_0} \cap \dot{R}[v_0] \neq \emptyset$, i.e. there is

$x \in U_{v_0} \cap \dot{R}[v_0]$. Then $w \prec_{S\cup\{\Box\neg G\}} v_0 = x$ or $w \prec_{S\cup\{\Box\neg G\}} v_0 \prec x$. In both cases (possibly using Lemma 4.2) we have $w \prec_{S\cup\{\Box\neg G\}} x$. In the case (bb):

$$(\forall S')(w \prec_{S'} v_0 \Rightarrow (\exists x \in U_{v_0})(\exists G' \in \mathcal{D} \cap \bigcup \dot{R}[v_0])\ w \prec_{S'\cup\{\Box\neg G'\}} x).$$

From $w \prec_{S\cup\{\Box\neg G\}} v_0$ it follows that there are some $x \in U_{v_0}$ and $G' \in \mathcal{D} \cap \bigcup \dot{R}[v_0]$ such that $w \prec_{S\cup\{\Box\neg G, \Box\neg G'\}} x$. Lemma 4.2 implies $w \prec_{S\cup\{\Box\neg G\}} x$, as required. We claim that for each formula $G \in \mathcal{D}$ and each world $w \in W$ the following holds:

$$\mathfrak{M}, w \Vdash G \ \text{ if and only if } \ G \in w.$$

The claim is proved by induction on the complexity of G. The only non-trivial case is when $G = B \rhd C$. Assume $B \rhd C \in w$, wRu and $u \Vdash B$. Induction hypothesis implies $B \in u$. We claim that $uS_w[C]_w$. Clearly $[C]_w \subseteq R[w]$. Assume $w \prec_S u$. Lemma 4.26 implies that there is an **IL**X-MCS v with $w \prec_{S\cup\{\Box\neg B\}} v$ and $C, \Box\neg C \in v$ (thus $v \in W$). Since $C \in v$, the induction hypothesis implies $v \Vdash C$. Since $w \prec v$, i.e. wRv, then $v \in [C]_w$. Now, $B \in \mathcal{D}$ and $B \in u$ imply $B \in \mathcal{D} \cap \bigcup \dot{R}[u]$. Thus, $uS_w[C]_w$ holds by the clause (b) from the definition. To prove the converse, assume $B \rhd C \notin w$. Since w is an **IL**X-MCS, $\neg(B \rhd C) \in w$. Lemma 4.25 implies there is u with $w \prec_{\{\Box\neg B, \neg C\}} u$ and $B \in u$. Lemma 4.2 implies $\Box\neg B \in u$. So, $B \wedge \Box\neg B \in u$; thus $u \in W$. The induction hypothesis implies $u \Vdash B$. Let $V \subseteq R[w]$ be such that uS_wV. We will find a world $v \in V$ such that $w \prec_{\{\neg C\}} v$. We will distinguish the cases (a) and (b) from the definition of the relation S_w. Consider the case (a). Let v be an arbitrary node in $V \cap \dot{R}[u]$. If $v = u$, clearly $w \prec_{\{\Box\neg B, \neg C\}} v$. If uRv, then we have $w \prec_{\{\Box\neg B, \neg C\}} u \prec v$. Lemma 4.2 implies $w \prec_{\{\Box\neg B, \neg C\}} v$. Consider the case (b). From $w \prec_{\{\Box\neg B, \neg C\}} u$ and the definition of S_w it follows that there is $v \in V$ and a formula $D \in \mathcal{D}$ such that $w \prec_{\{\Box\neg B, \neg C, \Box\neg D\}} v$. In both cases we have $w \prec_{\{\neg C\}} v$; thus $C \notin v$. Induction hypothesis implies $v \nVdash C$; whence $V \nVdash C$, as required. $\dashv$

This lemma brings us one step away from a completeness proof.

Theorem 4.29 The logic **IL**W is complete w.r.t. $\mathbf{IL}_{\text{set}}$W-models.

Proof. In the light of Lemma 4.28, it suffices to show that the **IL**W-structure $\mathfrak{M}$ for $\mathcal{D}$ possesses the property $(\mathsf{W})_{\text{gen}}$. Recall the characteristic property $(\mathsf{W})_{\text{gen}}$:

$$uS_wV \Rightarrow (\exists V' \subseteq V)(uS_wV' \ \&\ R[V'] \cap S_w^{-1}[V] = \emptyset).$$

Suppose for a contradiction that there are w, u and V such that:

$$uS_wV \ \&\ (\forall V' \subseteq V)(uS_wV' \Rightarrow R[V'] \cap S_w^{-1}[V] \neq \emptyset). \tag{4.1}$$

Let $\mathcal{V}$ denote all such sets V (we keep w and u fixed). Let $n = 2^{|\mathcal{D}|}$. Fix any enumeration

$\mathcal{D}_0, \ldots, \mathcal{D}_{n-1}$ of $\mathcal{P}(\mathcal{D})$ that satisfies $\mathcal{D}_0 = \emptyset$. We define a new relation S^i_w for all $0 \leq i < n$, $y \in W$ and $U \subseteq W$ as follows:

$$yS^i_wU \iff yS_wU,\ \mathcal{D}_i \subseteq \bigcup \dot{R}[y],\ U \subseteq \left[\bigvee_{G \in \mathcal{D}_i} \Box\neg G\right]_w .$$

Let $y \in W$ and $U \subseteq R[w]$ be arbitrary. Let us prove that yS_wU implies the following:

$$(\exists U' \subseteq U)(\exists i < n)\ yS^i_wU'. \tag{4.2}$$

If yS_wU holds by (a) from the definition of S_w, the set $U \cap \dot{R}[y]$ is non-empty. Pick arbitrary $z \in U \cap \dot{R}[y]$ and put $U' = \{z\}$. We have either $wRyRz$ or $y = z$. If $wRyRz$, we have $yS_w\{z\}$. Otherwise, $y = z$. Now quasi-reflexivity implies $yS_w\{z\}$. Since $y \in W$, there is a formula $G \in \mathcal{D}$ such that $G \wedge \Box\neg G \in y$. Fix $i < n$ such that $\mathcal{D}_i = \{G\}$. Clearly $\mathcal{D}_i \subseteq \bigcup \dot{R}[y]$. Since $z \in U$ and yS_wU, clearly $U' \subseteq R[w]$. Since $y = z$ or yRz, we also have $\Box\neg G \in z$. Truth lemma implies $U' \Vdash \Box\neg G$; since if zRt, $G \notin t$, (truth lemma is applied here) $t \nVdash G$, so $z \Vdash \Box\neg G$. Thus $U' \subseteq [\Box\neg G]_w$, and yS^i_wU'. If yS_wU holds by (b) from the definition of S_w, take:

$$U' = \{z \in U : (\exists G \in \mathcal{D} \cap \bigcup \dot{R}[y])\ w \prec_{\{\Box\neg G\}} z\};$$

$$\mathcal{D}_i = \{G \in \mathcal{D} \cap \bigcup \dot{R}[y] : (\exists z \in U)\ w \prec_{\{\Box\neg G\}} z\}.$$

In other words, U' is the image of the mapping that is implicitly present in the definition of the relation S_w (clause (b)): for each S, pick a world v_S (to be included in U'), and a formula G_S (to be included in $\mathcal{D}_i$). Let $m < n$ be maximal such that there are $U \in \mathcal{V}$ and $U' \subseteq U$ with the following properties:

(i) $(\forall x \in U)[(\exists y \in R[x])(\exists Z \subseteq U)(\exists i \leq m)\ yS^i_wZ \Rightarrow x \notin U']$;

(ii) $(\forall x \in W)(xS_wU \Rightarrow xS_wU')$.

Since $\mathcal{D}_0 = \emptyset$, we have $[\bigvee_{G \in \mathcal{D}_0} \Box\neg G]_w = [\bot]_w = \emptyset$. So there are no $Z \subseteq [\bigvee_{G \in \mathcal{D}_0} \Box\neg G]_w$ such that yS_wZ for some $y \in W$. So, if we take $m = 0$ and $U' = U$ for any $U \in \mathcal{V}$, (i) and (ii) are trivially satisfied. Since n is finite and conditions (i) and (ii) are satisfied for at least one value m, there must be a maximal $m < n$ with the required properties. Let us first prove that $m < n - 1$. Assume the opposite, that is (since $m < n$), $m = n - 1$. Then there are $U \in \mathcal{V}$ and $U' \subseteq U$ such that the conditions (i) and (ii) are satisfied for $m = n - 1$. Since $U \in \mathcal{V}$, we have uS_wU. The condition (ii) implies uS_wU'. Now $U \in \mathcal{V}$, $U' \subseteq U$ and uS_wU' imply $R[U'] \cap S^{-1}_w[U] \neq \emptyset$. Thus there are $x \in U'$ and $y \in R[x]$ such that yS_wU. Now (ii) implies yS_wU'. The earlier remark (4.2) implies that there is $Z \subseteq U'$ and $i < n$ such that yS^i_wZ. Since $m = n - 1$, it follows that $i \leq m$. The condition (i) implies $x \notin U'$, a contradiction. Thus, $m < n - 1$. Let us now prove that m is, contrary to the assumption, not maximal, by showing that $m+1$ satisfies (i) and (ii). Let $U \in \mathcal{V}$ and $U' \subseteq U$ be some sets such that the conditions (i) and (ii) are satisfied for m. Denote:

$$Y = \{x \in U' : (\exists y \in R[x])(\exists Z \subseteq U')\ yS_w^{m+1}Z\}.$$

Let us prove that $m+1$ also satisfies (i) and (ii) with U' instead of U, and $U' \setminus Y$ instead of U'. We should first show that $U' \in \mathcal{V}$. So, suppose that $uS_wT \subseteq U'$. Now, $T \subseteq U' \subseteq U$ and $U \in \mathcal{V}$ imply that there are some $v \in T$ and $z \in R[v]$ such that zS_wU. The property (ii) for m (with sets U and U') implies zS_wU'. So, $R[T] \cap S_w^{-1}[U'] \neq \emptyset$, as required. Now let us verify the property (i) for the newly defined sets (U' and $U' \setminus Y$). Let $x \in U', y \in R[x], Z \subseteq U', i \leq m+1$ be arbitrary such that yS_w^iZ. If $i \leq m$, the property (i) for m implies $x \notin U'$, so in particular, $x \notin U' \setminus Y$. If $i = m+1$, then $x \in Y$. Thus $x \notin U' \setminus Y$ and the condition (i) is satisfied.

It remains to prove (ii). Take arbitrary $x \in W$ such that xS_wU'. For every $y \in Y$, the definition of Y implies the existence of some $z_y \in R[y]$ and $U_y \subseteq U'$ such that $z_yS_w^{m+1}U_y$. From the definition of the relation S_w^{m+1} we have $\mathcal{D}_{m+1} \subseteq \bigcup \dot{R}[z_y]$. Now, yRz_y and the truth lemma imply $y \Vdash \Diamond G$, for each $G \in \mathcal{D}_{m+1}$. From the definition of the relation S_w^{m+1} and $z_yS_w^{m+1}U_y$ we have $U_y \subseteq [\bigvee_{G \in \mathcal{D}_{m+1}} \Box\neg G]_w$. So, the following holds:

$$Y \Vdash \bigwedge_{G \in \mathcal{D}_{m+1}} \Diamond G \quad \text{and} \quad U_y \Vdash \bigvee_{G \in \mathcal{D}_{m+1}} \Box\neg G,$$

for all $y \in Y$. Thus, $U_y \cap Y = \emptyset$, for every $y \in Y$. For every $y \in U' \setminus Y$ put $U_y = \{y\}$. Again, $U_y \cap Y = \emptyset$. Note that $\bigcup_{y \in U'} U_y = U' \setminus Y$. Now xS_wU' and quasi-transitivity imply $xS_wU' \setminus Y$. The fact that (i) and (ii) hold for $m+1$ contradicts the maximality of m. $\dashv$

We could have proven completeness of **IL**W through simpler arguments if the completeness of **IL**W was our ultimate goal, and one way is through the following proposition. However, we need the format of the argument above when proving completeness of extensions of **IL**W.

Proposition 4.30 Let $\mathfrak{M} = (W, R, \{S_w \in W\}, \Vdash)$ be a Veltman model with the property (W). Then $gen(\mathfrak{M}) = (W, R, \{S'_w \in W\}, \Vdash)$ is a generalised Veltman model with the property $(\mathsf{W})_{\mathrm{gen}}$.

Proof. Assume the opposite, i.e. $gen(\mathfrak{M})$ has the property $\overline{(\mathsf{W})_{\mathrm{gen}}}$. We will obtain an infinite chain $z_0S_wx_0Rz_1S_wx_1R\ldots$, which contradicts (W). Choose a world w, a non–empty set $X \subseteq R[w]$ and $z_0 \in W$ such that $z_0S'_wX$ and:

$$(\forall V \subseteq X)\Big(z_0S'_wV \Rightarrow (\exists v \in V)(R[v] \cap (S'_w)^{-1}[X] \neq \emptyset)\Big).$$

We have defined the first element, z_0, of our sequence. We will define the remainder of the sequence recursively. Suppose we have defined worlds $z_0, x_0, z_1, x_1, \ldots$ up to some world z_i for $i \in \omega$ such that the following is true:

$$(\forall V \subseteq X)\Big(z_i S'_w V \Rightarrow (\exists v \in V)(R[v] \cap (S'_w)^{-1}[X] \neq \emptyset)\Big).$$

We will now construct worlds x_i and z_{i+1}. Note that the world z_0 satisfies this property.

The fact that $z_i S'_w X$ and the definition of model $gen(\mathfrak{M})$ imply $z_i S_w x_i$ for some $x_i \in X$. Hence $z_i S'_w \{x_i\}$, so the property that by assumption holds for z_i implies that there is z_{i+1} such that $x_i R z_{i+1}$ and $z_{i+1} S'_w X$.

Assume there is a set $V \subseteq X$ such that:

$$z_{i+1} S'_w V \;\&\; (\forall v \in V)(R[v] \cap (S'_w)^{-1}[X] = \emptyset).$$

Since wRx_iRz_{i+1}, we have $x_i S'_w \{z_{i+1}\}$. Now, $z_i S'_w \{x_i\}$, $x_i S'_w \{z_{i+1}\}$, $z_{i+1} S'_w V$ and quasi-transitivity of the relation S'_w imply $z_i S'_w V$. But this, together with the fact that $(\forall v \in V)(R[v] \cap (S'_w)^{-1}[X] = \emptyset)$, contradicts the property that, by assumption, holds for z_i. Hence, for the world z_{i+1} the following holds:

$$(\forall V \subseteq X)\Big(z_{i+1} S'_w V \Rightarrow (\exists v \in V)(R[v] \cap Z \neq \emptyset)\Big).$$

$\dashv$

Goris and Joosten proved in [27] the completeness of **IL**W^* (recall that this is equivalent to **IL**$\mathsf{WM_0}$) w.r.t. ordinary Veltman semantics. One way to obtain completeness w.r.t. generalised Veltman semantics would be to use Proposition 4.12. We proceed to prove completeness directly, without referring to ordinary semantics. The benefit is, as always, in the possibility that this approach might be used for the extension of **IL**W (and **IL**W^*) which may not be complete with respect to ordinary Veltman semantics.

Theorem 4.31 The logic **IL**W^* is complete w.r.t. $\mathbf{IL}_{\text{set}}\mathsf{W}^*$-models.

Proof. With Lemma 4.28, it suffices to prove that the **IL**$\mathsf{W}*$-structure for $\mathcal{D}$ possesses the properties $(\mathsf{W})_{\text{gen}}$ and $(\mathsf{M_0})_{\text{gen}}$, for each appropriate $\mathcal{D}$. So, let $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ be the **IL**W^*-structure for $\mathcal{D}$. Theorem 4.29 shows that the model $\mathfrak{M}$ possesses the property $(\mathsf{W})_{\text{gen}}$. It remains to show that it possesses the property $(\mathsf{M_0})_{\text{gen}}$. Assume $wRuRxS_wV$. We claim that there is $V' \subseteq V$ such that uS_wV' and $R[V'] \subseteq R[u]$. First, consider the case when xS_wV holds by the clause (a) from the definition of S_w. So there is $v \in V$ such that $x = v$ or xRv. In both cases, $wRuRv$, and so $uS_w\{v\}$. It is clear that $R[v] \subseteq R[x] \subseteq R[u]$. So it suffices to take $V' = \{v\}$. Otherwise, xS_wV holds by the clause (b). Take $V' = \{v \in V : w \prec_{u^\Box_\emptyset} v\}$. Clearly, $V' \subseteq V \subseteq R[w]$. Assume $w \prec_S u$. Now $w \prec_S u \prec x$ and Lemma 4.15 imply $w \prec_{S \cup u^\Box_\emptyset} x$. The definition of xS_wV (clause (b)) implies there is $G \in \mathcal{D} \cap \bigcup \dot{R}[x]$ (so $G \in \mathcal{D} \cap \bigcup \dot{R}[u]$) and $v \in V$ such that $w \prec_{S \cup u^\Box_\emptyset \cup \{\Box\neg G\}} v$, thus also $v \in V'$. In particular, $w \prec_{S \cup \{\Box\neg G\}} v$. Since S was arbitrary, uS_wV'. It remains to verify that $R[V'] \subseteq R[u]$. Assume $V' \ni vRz$. Since $w \prec_{u^{\Box}_{\emptyset}} v$, for all $\Box B \in u$ we have $\Box B \in v$, and since vRz, it follows that $\Box B, B \in z$. Thus, $u \prec z$ i.e. uRz. $\dashv$

In [49] it is shown that **IL**W* possesses the finite model property w.r.t. generalised Veltman semantics. To show decidability, (stronger) completeness w.r.t. ordinary Veltman models was used in [49]. However, we observe that Theorem 4.31 above suffices for the mere purpose of decidability.

4.2.7 The logic ILWR

In previous subsections we saw that the completeness of **IL**R can be proven using **IL**R-structures, and that the completeness of **IL**W can be proven using **IL**W-structures. These two types of structures have non-trivial differences (they differ by more than just the notion of **IL**X-consistency used). However, we saw that **IL**W*-structures have the same form as **IL**W-structures. So, one may hope to prove completeness of **IL**WR with the help of the same form of structures.

Unfortunately, it seems that **IL**WR-structures, if by an **IL**WR-structure we mean an **IL**W-structure with the notion of **IL**W-consistency replaced with that of **IL**WR-consistency, does not possess the characteristic property $(\mathsf{R})_{\mathrm{gen}}$. In the remainder of this chapter (based on [26]) we call the type of the problem that emerges “the label iteration problem”. We will demonstrate how to overcome this problem for a simpler logic. As for the logic **IL**WR itself, we return to it in the closing chapter of the thesis. We decided to postpone the discussion because (1) we do not obtain an answer to the main problem of whether **IL**WR is complete or not; (2) we need some arithmetical background for that chapter; and (3) the discussion will be lengthy, and is more naturally presented as a separate chapter.

4.3 Generalising ILX-structures

The labelling that was considered so far was concerned with two or three worlds at a time. Due to the transitivity of R, labelling longer sequences often simplifies to labelling pairs or triples of worlds.

In this section, we first show that labelling sequences in **IL**R-models indeed reduces to labelling triples of worlds. The completeness of the logic **IL**R w.r.t. the ordinary Veltman semantics is still an open problem. The fact that labels for this logic are compatible with transitive closures makes our labelling a good candidate for the step-by-step completeness proofs such as the *construction method* [27].

In the remainder of the section we deal with logics whose labelling does not trivially reduce to labelling pairs or triples of worlds. At the moment, the only logics falling into this category that we know of are various extensions of **IL**W. An example is **IL**WR, which may also be the most interesting example since it is the simplest logic among those whose (in)completeness status is currently open.

There is an easily identifiable problem in taking transitive closures when working with assuringness. Suppose we are working in **IL**R. Let us recall that for **IL**R we have that $\Gamma{\prec_S}\Delta{\prec_T}\Delta'$ implies $\Gamma{\prec_{S\cup\Delta^\Box_T}}\Delta'$ (Lemma 3.26).

Consider the following situation that might occur while iteratively building a model in a step-by-step completeness proof: $x \prec_U \Gamma \prec_S \Delta \prec_T y$, where U, S and T are arbitrary sets of formulas, while x, Γ, Δ, y are MCS's and at the same time the worlds in the model we are building. If we wish to compute the label between x and y it does make a difference whether we first compute the label for the transitive transition between x and Δ or the label for the transitive transition between Γ and y. In the first case we get $U \cup \Gamma^\Box_S \cup \Delta^\Box_T$ as the label between x and y, and otherwise we get $U \cup \Gamma^\Box_{S\cup\Delta^\Box_T}$. Lemma 4.32 implies that the following is the case:

$$U \cup \Gamma^\Box_{S\cup\Delta^\Box_T} \subseteq U \cup \Gamma^\Box_S \cup \Delta^\Box_T.$$

This determines which way should the closure procedure proceed when faced with a choice, i.e. we should go with the first choice since it results in a more informative label.

Lemma 4.32 For logics containing R we have that $\Gamma{\prec_S}\Delta$ implies $\Gamma^\Box_{S\cup\Delta^\Box_T} \subseteq \Delta^\Box_T$.

Proof. Consider $\Box\neg A \in \Gamma^\Box_{S\cup\Delta^\Box_T}$, that is, for some $S_i \in S$ and $\Box\neg B_j \in \Delta^\Box_T$, $A \rhd \bigvee\neg S_i \vee \bigvee\neg\Box\neg B_j \in \Gamma$. By R, $\neg(A \rhd \bigvee B_j) \rhd \bigvee\neg S_i \in \Gamma$, whence by $\Gamma{\prec_S}\Delta$, we get $A \rhd \bigvee B_j \in \Delta$. But for each B_j there is $T_{jk} \in T$ with $B_j \rhd \bigvee\neg T_{jk} \in \Delta$, whence $A \rhd \bigvee\neg T_{jk} \in \Delta$ and $\Box\neg A \in \Delta^\Box_T$. ⊣

4.3.1 Motivation for labelling systems

We will now present an issue concerning labelling in **IL**WR. Both **IL**W [19] and **IL**R [50] are known to be complete, but this question remains open for **IL**WR.

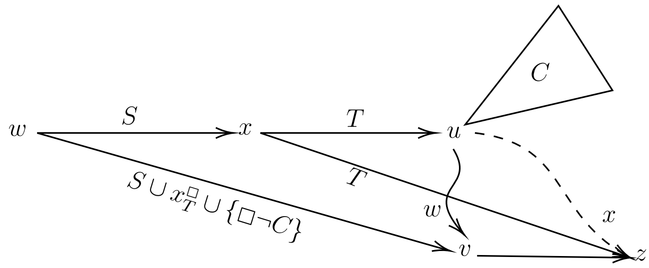


Figure 4.2: Labels for **IL**WR

We first discuss this problem, and then see how a more elaborate labelling system can help. At the moment we do not know if the labelling systems will lead to a full completeness proof of **IL**WR (see the final chapter of this thesis).

The plan for the remainder of the section, and indeed the chapter, is to illustrate *labelling systems* and how they help. To accomplish this we will work with **IL**P, another logic exhibiting the same issue (if we wish to prove completeness with respect to **IL**W-structure-like objects). With **IL**P we can give a full completeness proof together with a to-the-point presentation on how to deal with logics with non-trivial labelling of sequences. Thus, the above mentioned elaborate labelling systems should at least be an ingredient, if not the whole solution, in proving the more interesting completeness results (such as, possibly, the one for **IL**WR).

Suppose[2] we are building a model step-by-step (as in the *construction method* [27]) and we have $A \rhd B \in w \prec_S x \prec_T u \ni A$. So, we need to find some v with $B \in v$ and a sufficiently strong label for wRv; and then declare uS_wv. Using the labelling lemmas for W and R, it is easy to find v with $w \prec_{S \cup x_T^{\Box} \cup \{\Box\neg C\}} v$ for some C contained either in u or in a world R-accessible from u. Let us for the moment suppose that any such v fits our purposes.

Now, assume that at some later point during the construction, a world z appears with vRz. By the frame condition of the principle R, we should have uS_xz. If we were building an **IL**R-model (and not an **IL**WR-model), we would have to ensure just that z has the same assuringness as u with respect to x, that is, xRz should be labelled with T. Since we are building an **IL**WR-model and in order to ensure the frame condition for W, in addition to that we are to find a formula C' with $x \prec_{T \cup \{\Box\neg C'\}} z$. An obvious candidate for C' is C. However, from $w \prec_{S \cup x_T^{\Box} \cup \{\Box\neg C\}} v \prec z$ we only get $x \prec_T z$ (Lemma 4.21), and what we would like is to have $x \prec_{T \cup \{\Box\neg C\}} z$. Let us refer to this phenomenon as *the problem of label iteration*.

One way to try to solve this problem is to simply *require* $\Box\neg C$ to appear at the right

[2]This paragraph describes the situation represented in Figure 4.2. The triangle in the top-right corner represents the set $\dot{R}[u]$. Some world in this set has to contain the formula C.

place in the original label, i.e., instead of asking for $w \prec_{S \cup x^{\Box}_{T} \cup \{\Box\neg C\}} v$, we ask for

$$w \prec_{S \cup x^{\Box}_{T \cup \{\Box\neg C\}} \cup \{\Box\neg C\}} v. \tag{4.3}$$

If we are proving completeness w.r.t. generalised semantics using the approach from earlier sections, this means that we should add a new condition in the definition of S_w (Definition 4.6). However, similar to how the original condition concerning two worlds requires us to add the new condition concerning three worlds that we just described, this condition itself requires us to add another condition, this time concerning four worlds. Let us illustrate this.

Suppose we chose the world v and occur the following situation later in the construction process (see Figure 4.2):

$$A \rhd B \in w \prec_S x \prec_T u \ni A \quad \text{and} \quad v \prec z.$$

We would like to show $uS_x z$. With the new condition added in the definition of S_w, we now have to show that if $x \prec_{S'} x' \prec_{T'} u$, then the world v satisfies $x \prec_{S' \cup \{\Box\neg A\} \cup x'^{\Box}_{T' \cup \{\Box\neg A\}}} v \ni B$. It would be convenient if we were able to prove

$$w \prec_{x^{\Box}_{S' \cup \{\Box\neg A\} \cup x'^{\Box}_{T' \cup \{\Box\neg A\}}} \cup \{\Box\neg A\}} v, \tag{4.4}$$

since from this we can conclude $x \prec_{S' \cup \{\Box\neg A\} \cup x'^{\Box}_{T' \cup \{\Box\neg A\}}} v \ni B$. However, to be able to conclude (4.4) we need to have a new case in the definition of S_w, one that concerns not just w, x and u; but w, x, x' and u. So, the choice of v that fits earlier requirements might not be good enough. Analogous reasoning applies for longer sequences of worlds, i.e. any finite amount of requirements will not suffice.

It turns out the problem of label iteration, that, as we just saw, occurs with **IL**WR, also occurs when trying to prove that **IL**P is complete w.r.t. the class of generalised **IL**P-frames where an additional requirement which ensures $(\mathsf{W})_{\mathrm{gen}}$ is present.[3] In the remainder of this section we will give a detailed exposition on how to handle this problem in the case of **IL**P. The same general approach should be useful for any other extension of **IL**W that exhibits the problem of label iteration.

[3] $(\mathsf{W})_{\mathrm{gen}}$ is the following condition: $uS_wV \Longrightarrow (\exists V' \subseteq V)uS_wV' \,\&\, R[V'] \cap S_w^{-1}[V] = \emptyset$.
The requirement we mention is that whenever $w \prec_S u$ and we are making an S_w-successor v of u, that $w \prec_{S \cup \{\Box\neg B\}} v$ for some $B \in \mathcal{D} \cap \bigcup \dot{R}[u]$ where $\dot{R}[u] = R[u] \cup \{u\}$. Since it is well known that **IL**P, which is a complete logic [18], contains **IL**W (see e.g. [66]), we already know that **IL**P is complete w.r.t. the class of generalised **IL**P-frames that satisfy $(\mathsf{W})_{\mathrm{gen}}$. We do not, however, know in general if the models obtained by the standard completeness argument also satisfy this specific requirement (which is, at least a priori, stronger than $(\mathsf{W})_{\mathrm{gen}}$).

4.3.2 Test case: ILWP-structures

We will now introduce the labelling system for **IL**P and prove the completeness of **IL**P w.r.t. the class of generalised **IL**P-frames where an additional requirement which ensures $(\mathsf{W})_{\rm gen}$ is present.

$(\mathsf{P})_{\rm gen}$:

$$wRw'RuS_wV \Rightarrow (\exists V' \subseteq V)\ uS_{w'}V'.$$

Recall the labelling lemma for **IL**P (Lemma 3.22)

$$w \prec_S x \prec_T u \Rightarrow w \prec_{S \cup x^{\boxdot}_T} u.$$

The actual labelling that we use is an iterated generalisation of this property. Thus, instead of defining labels between pairs of MCS's, we consider tuples of MCS's with labels between them: $w_n \prec_{S_n} w_{n-1} \prec_{S_{n-1}} \cdots \prec_{S_1} w_0$. We wish to define labels for **IL**P similar to the ones for **IL**WR between w and v in (4.3) and (4.4). We will first define these labels, and then prove the appropriate labelling lemma.

Definition 4.33 For $n \in \omega \backslash \{0\}$, let $\{w_0, \ldots, w_n\}$ be a finite sequence of **IL**P-MCS's, let $\{S_1, \ldots, S_n\}$ be a finite sequence of sets of formulas and B be a formula. We recursively define n sets of formulas, one for every $j \in \{1, \ldots, n\}$:

$$Q(\{w_0, \ldots, w_n\}, \{S_1, \ldots, S_n\}, B, j).$$

Usually the MCS's $\{w_0, \ldots, w_n\}$ and the sets of formulas $\{S_1, \ldots, S_n\}$ will be clear from the context, so we will write $Q_j(B)$ for $Q(\{w_0, \ldots, w_n\}, \{S_1, \ldots, S_n\}, B, j)$. We now recursively define the elements of our sequence:

$$\begin{aligned} Q_1(B) &:= S_1 \cup \{\Box\neg B\}; \\ Q_{j+1}(B) &:= S_{j+1} \cup \{\Box\neg B\} \cup {w_j}^{\boxdot}_{Q_j(B)}. \end{aligned}$$

Note that the preceding definition amounts to the following:

$$Q_j(B) = S_j \cup \{\Box\neg B\} \cup {w_{j-1}}^{\boxdot}_{S_{j-1} \cup \{\Box\neg B\} \cup {w_{j-2}}^{\boxdot}_{S_{j-2} \cup \{\Box\neg B\} \cup \ldots_{\ldots \cup {w_1}^{\boxdot}_{S_1 \cup \{\Box\neg B\}}}}}.$$

Lemma 4.34 Let $n \in \omega \backslash \{0\}$ be arbitrary, $\{w_0, \ldots, w_n\}$ be a finite sequence of **IL**P-MCS's, $\{S_1, \ldots, S_n\}$ a finite sequence of sets of formulas and $B \rhd C$ a formula such that:

$$B \rhd C \in w_n \prec_{S_n} w_{n-1} \prec_{S_{n-1}} \cdots \prec_{S_1} w_0 \ni B.$$

Then there is an **IL**P-MCS v such that $w_n \prec_{Q_n(B)} v$ and $C, \Box\neg C \in v$.

Proof. We prove the claim by induction on n. In the base case we are to find v such that $w_1 \prec_{S_1 \cup \{\Box\neg B\}} v$. But this is just Lemma 3.29.

Let us prove the claim for $n+1$. Fix MCS's $\{w_0, \dots, w_{n+1}\}$, labels $\{S_1, \dots, S_{n+1}\}$ and a formula $B \rhd C$. Assume

$$B \rhd C \in w_{n+1} \prec_{S_{n+1}} w_n \prec_{S_n} \cdots \prec_{S_1} w_0 \ni B.$$

The goal is to find v with $w_{n+1} \prec_{Q_{n+1}(B)} v \ni C, \Box\neg C$, i.e.

$$w_{n+1} \prec_{S_{n+1} \cup \{\Box\neg B\} \cup {w_n}^{\boxdot}_{Q_n(B)}} v \ni C, \Box\neg C.$$

From $w_{n+1} \prec w_n$ and the axiom P we have $B \rhd C \in w_n$. By the induction hypothesis, there is v with $w_n \prec_{Q_n(B)} v \ni C, \Box\neg C$. From $w_{n+1} \prec_{S_{n+1}} w_n \prec_{Q_n(B)} v$ and the labelling lemma for $\mathbf{IL}\mathsf{P}$ (Lemma 3.22) we have:

$$w_{n+1} \prec_{S_{n+1} \cup {w_n}^{\boxdot}_{Q_n(B)}} v.$$

Since $\{\Box\neg B\} \subseteq Q_n(B) \subseteq {w_n}^{\boxdot}_{Q_n(B)}$, we have $S_{n+1} \cup {w_n}^{\boxdot}_{Q_n(B)} = S_{n+1} \cup \{\Box\neg B\} \cup {w_n}^{\boxdot}_{Q_n(B)}$. $\dashv$

Note that the last line shows that a simpler definition of $Q_{j+1}(B)$ would suffice: $Q_{j+1}(B) := S_{j+1} \cup {w_j}^{\boxdot}_{Q_j(B)}$ instead of $Q_{j+1}(B) := S_{j+1} \cup \{\Box\neg B\} \cup {w_j}^{\boxdot}_{Q_j(B)}$. However, the purpose of this section is to introduce a method for dealing with arbitrary extensions of $\mathbf{IL}\mathsf{W}$. We do not think it is likely that such a simplification could be made in the case of more interesting logics, such as $\mathbf{IL}\mathsf{WR}$.

Recall that in this chapter, the set $\mathcal{D}$ is always assumed to be a finite set of formulas closed under taking subformulas and single negations, and $\top \in \mathcal{D}$.

Next we define the structures w.r.t. which we later prove completeness. Note that in the definition below, worlds are sets of formulas. Because of this, the operation $\bigcup \dot{R}[u]$ makes sense and defines a set of formulas.

Definition 4.35 We say that $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ is the $\mathbf{IL}\mathsf{WP}$-structure for a set of formulas $\mathcal{D}$ if:

- $W = \{w : w \text{ is an } \mathbf{IL}\mathsf{P}\text{-MCS and for some } B \in \mathcal{D},\ B \wedge \Box\neg B \in w\}$;
- $wRu \Leftrightarrow w \prec u$;
- $uS_wV \Leftrightarrow wRu$ and $V \subseteq R[w]$ and, moreover, one of the following holds:
  (a) $V \cap \dot{R}[u] \neq \emptyset$;
  (b) we have for all $n \in \omega \setminus \{0\}$, all $\{w_0, \dots, w_n\}$, and all $\{S_1, \dots, S_n\}$:
  $$w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u \Rightarrow (\exists v \in V)(\exists B \in \mathcal{D} \cap \bigcup \dot{R}[u])\ w \prec_{Q_n(B)} v;$$
- $w \Vdash p \Leftrightarrow p \in w$.

Lemma 4.36 The **IL**WP-structure $\mathfrak{M}$ for $\mathcal{D}$ is a generalised Veltman model. Furthermore, the following holds for each $w \in W$ and $G \in \mathcal{D}$:

$$\mathfrak{M}, w \Vdash G \text{ if and only if } G \in w,$$

Proof. Let us first verify that the **IL**WP-structure $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ for $\mathcal{D}$ is a generalised Veltman model. All the properties, except for quasi-transitivity, have easy proofs (see Lemma 4.28).

Let us prove quasi-transitivity. Thus, we assume uS_wV, and vS_wU_v for all $v \in V$. We put $U = \bigcup_{v \in V} U_v$ and claim that uS_wU. Clearly $U \subseteq R[w]$. To prove uS_wU we will distinguish cases from the definition of the relation S_w for uS_wV.

In Case (a), there exists an MCS $v_0 \in V$ for some $v_0 \in \dot{R}[u]$. We will next distinguish two Cases from the definition of $v_0 S_w U_{v_0}$.

In Case (aa) we can find $x \in U_{v_0}$ for some $x \in \dot{R}[v_0]$. Since $v_0 \in \dot{R}[u]$, also $x \in \dot{R}[u]$. And since $x \in U_{v_0} \subseteq U$, we have $U \cap \dot{R}[u] \neq \emptyset$. So, we have uS_wU as required.

In Case (ab):

$$\begin{aligned}&\text{For all } n \in \omega \backslash \{0\}, \text{ all } \{w_0, \ldots, w_n\}, \text{ and all } \{S_1, \ldots, S_n\} \text{ we have:} \\ &w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = v_0 \Rightarrow (\exists x \in U_{v_0})(\exists B \in \mathcal{D} \cap \bigcup \dot{R}[v_0])\ w \prec_{Q_n(B)} x.\end{aligned} \tag{4.5}$$

To prove uS_wU in this case, we will use Case (b) from the definition of the relation S_w. Let $n \in \omega \backslash \{0\}$ be arbitrary and let $\{w_0, \ldots, w_n\}$ and $\{S_1, \ldots, S_n\}$ be arbitrary such that $w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u$. If $u = v_0$, applying (4.5) with the worlds $\{w_0, \ldots, w_n\}$ and the labels $\{S_1, \ldots, S_n\}$ produces the required $x \in U_{v_0}$ and $B \in \mathcal{D} \cap \bigcup \dot{R}[v_0]$. Otherwise, i.e. if uRv_0, let $w'_0 = v_0$, $w'_{i+1} = w_i$, $S'_1 = \emptyset$, $S'_{i+1} = S_i$ and apply the formula above with $n+1$, the sequence $\{w'_0, \ldots, w'_{n+1}\}$ and the labels $\{S'_1, \ldots, S'_{n+1}\}$. This gives us a world $x \in U_{v_0}$ and a formula $B \in \mathcal{D} \cap \bigcup \dot{R}[v_0]$ with:

$$w \prec_{S_n \cup \{\Box\neg B\} \cup {w_{n-1}}^{\boxdot}_{S_{n-1} \cup \ldots \ldots {w_1}^{\boxdot}_{S_1 \cup \{\Box\neg B\} \cup u^{\boxdot}_{\emptyset \cup \{\Box\neg B\}}}}} x.$$

Weakening this fact by Lemma 3.8 with removing $u^{\boxdot}_{\emptyset \cup \{\Box\neg B\}}$, we have the required property. Since uRv_0 or $u = v_0$, we have $\dot{R}[v_0] \subseteq \dot{R}[u]$. Thus, we can reuse B for this S_w transition.

In Case (b):

$$\begin{aligned}&\text{For all } n \in \omega \backslash \{0\}, \text{ all } \{w_0, \ldots, w_n\}, \text{ and all } \{S_1, \ldots, S_n\} \text{ we have:} \\ &w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u \Rightarrow (\exists v \in V)(\exists B \in \mathcal{D} \cap \bigcup \dot{R}[u])\ w \prec_{Q_n(B)} v.\end{aligned}$$

To prove uS_wU we will use Case (b) from the definition of the relation S_w. So, let

$n \in \omega\backslash\{0\}$ be arbitrary and let $\{w_0, \ldots, w_n\}$ and $\{S_1, \ldots, S_n\}$ be arbitrary such that $w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u$.

By the assumption of this case, there are $v_0 \in V$ and $B \in \mathcal{D} \cap \bigcup \dot{R}[u]$ such that $w \prec_{Q_n(B)} v_0$. From $v_0 \in V$ we have $v_0 S_w U_{v_0}$. We will next distinguish the possible cases in the definition for $v_0 S_w U_{v_0}$.

In the first Case (ba) we have $U_{v_0} \cap \dot{R}[v_0] \neq \emptyset$, i.e. there is $x \in U_{v_0}$ such that either $v_0 = x$ or $v_0 R x$. In both cases we have $w \prec_{Q_n(B)} x$.

In Case (bb), we have (Case (b) for $v_0 S_w U_{v_0}$ applied to $n = 1$ and $S_1 = Q_n(B)$) that there are some $x \in U_{v_0}$ and $B' \in \mathcal{D} \cap \bigcup \dot{R}[v_0]$ such that $w \prec_{Q_n(B) \cup \{\Box\neg B'\}} x$. By weakening, $w \prec_{Q_n(B)} x$, as required.

We claim that for each formula $G \in \mathcal{D}$ and each world $w \in W$ the following holds:

$$\mathfrak{M}, w \Vdash G \quad \text{if and only if} \quad G \in w.$$

The proof is by induction on the complexity of G. The only non-trivial case is when $G = B \rhd C$.

Assume $B \rhd C \in w$, wRu and $u \Vdash B$. Induction hypothesis implies $B \in u$. We claim that $uS_w[C]_w$ by Case (b) from the definition of S_w. Clearly wRu and $[C]_w \subseteq R[w]$.

Fix $n \in \omega\backslash\{0\}$, $\{w_0, \ldots, w_n\}$ and $\{S_1, \ldots, S_n\}$. Assume $w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u$. Since $B \rhd C \in w_n$ and $B \in w_0$, Lemma 4.36 implies that there is an **IL**P-MCS v with $w_n \prec_{Q_n(B)} v$ and $C, \Box\neg C \in v$ (thus $v \in W$). Since $C \in v$, the induction hypothesis implies $v \Vdash C$. Since $w \prec v$, i.e. wRv, then $v \in [C]_w$. Finally, $B \in \mathcal{D}$ and $B \in u$ imply $B \in \mathcal{D} \cap \bigcup \dot{R}[u]$.

To prove the converse, assume $B \rhd C \notin w$. Since w is an **IL**P-MCS, $\neg(B \rhd C) \in w$. Lemma 3.28 implies there is u with $w \prec_{\{\Box\neg B, \neg C\}} u$ and $B \in u$. Since $w \prec_{\{\Box\neg B\}} u$, we have in particular that $\Box\neg B \in u$. So, $u \in W$. The induction hypothesis implies $u \Vdash B$. Let $V \subseteq R[w]$ be such that uS_wV. We will find a world $v \in V$ such that $w \prec_{\{\neg C\}} v$. We will distinguish Cases (a) and (b) from the definition of the relation S_w. Consider Case (a). Let v be an arbitrary world in $V \cap \dot{R}[u]$. If $v = u$, clearly $w \prec_{\{\Box\neg B, \neg C\}} v$. If uRv, then we have $w \prec_{\{\Box\neg B, \neg C\}} u \prec v$. This implies $w \prec_{\{\Box\neg B, \neg C\}} v$. Consider Case (b). From $w \prec_{\{\Box\neg B, \neg C\}} u$ and the definition of S_w it follows that there is $v \in V$ such that (for some formula D) $w \prec_{\{\Box\neg B, \neg C, \Box\neg D\}} v$. In both cases we have $w \prec_{\{\neg C\}} v$; thus $C \notin v$. Induction hypothesis implies $v \nVdash C$; whence $V \nVdash C$, as required. ⊣

Theorem 4.37 **IL**P is complete w.r.t. the class of all generalised Veltman frames satisfying $(\mathsf{P})_{\text{gen}}$. In particular, **IL**P is complete w.r.t. the class of **IL**WP-structures generated by all adequate sets $\mathcal{D}$.

Proof. In the light of Lemma 4.36, it suffices to show that the **IL**WP-structure $\mathfrak{M}$ for $\mathcal{D}$

possesses the property $(\mathsf{P})_{\text{gen}}$.[4]

Let us prove $(\mathsf{P})_{\text{gen}}$. Let $wRw'RuS_wV$ and take $V' = V \cap R[w']$. We claim $uS_{w'}V'$.

We distinguish two possible cases for uS_wV. If it holds by Case (a), there is $v \in V$ such that either $u = v$ or uRv. In both cases $w'Rv$. Let $U = \{v\}$. Clearly $U \subseteq V$. Since $w'RuRv$, $uS_{w'}\{v\}$, i.e. $uS_{w'}U$. The remainder of the proof deals with the case when uS_wV holds by Case (b) from the definition of S_w.

Fix $n \in \omega\backslash\{0\}$, the worlds $\{w_0, \ldots, w_n\}$ and the labels $\{S_1, \ldots, S_n\}$. Assume $w' = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u$. We have $w \prec_{\emptyset} w_n \prec_{S_n} \cdots \prec_{S_1} w_0$. Now the definition of uS_wV implies there is $v \in V$ with:

$$w \prec_{\emptyset \cup \{\Box\neg B\} \cup w_n{}^{\boxdot}_{Q_n(B)}} v.$$

We claim that $w_n \prec_{Q_n(B)} v$. Assume $A \rhd \bigvee \neg F_i \in w_n$ with $F_i \in Q_n(B)$ (we are to show that $\neg A, \Box\neg A \in v$). Clearly $\neg A, \Box\neg A \in w_n{}^{\boxdot}_{Q_n(B)}$. Since $a \prec_S b$ implies $S \subseteq b$, we have $\neg A, \Box\neg A \in v$. ⊣

So, we have a strategy to tackle less well-behaved logics. In the final chapter of the thesis we will see what happens when we apply this approach to **IL**WR.

[4]In general, for example with the logic **IL**WR, we would want to verify if $(\mathsf{W})_{\text{gen}}$ holds. The proof would be the same as the proof of Theorem 4.29. In this case it is a consequence of $(\mathsf{P})_{\text{gen}}$.

Chapter 5

Decidability

The content of this chapter is based on the published papers [49] and [50].

Both papers [49] and [50] use tools introduced in an earlier paper [53]. Throughout these three papers several inconvenient choices stacked up, culminating in an awkward notion of an "adequate set $\Gamma_{\mathcal{D}}$ for an appropriate set $\mathcal{D}$" in [50]. Unfortunately such inconveniences were unavoidable if we wanted, as we did, to use results from the previously published papers without modifying the original proofs. We will use this chapter as an opportunity to fix these issues. We will first prove a slightly modified version of the key result of [53]. We will then use this modified framework to present the results of [49] and [50].

Introduction

For **IL**, **IL**M, **IL**P and **IL**W, the original completeness proofs were proofs of completeness w.r.t. the appropriate finite models [18], [19]. For these logics, the finite model property (FMP) w.r.t. ordinary Veltman semantics, as well as their decidability, is thus immediate (and completeness and the FMP w.r.t. generalised Veltman semantics are easily shown to follow from these results). These completeness proofs use *truncated* maximal consistent sets, that is, sets that are maximal consistent with respect to the so-called *adequate* set. The principal requirement of *adequacy* is that the set is finite. The exact requirements vary with the logic at hand. Already with **IL**M, defining adequacy is not trivial (see [18]).

For more complex logics, not much is known about the FMP w.r.t. ordinary Veltman semantics. The filtration method can be used with generalised models to obtain finite models. This approach was successfully used to prove the FMP of **IL**M_0 and **IL**W^* w.r.t. generalised Veltman semantics [53], [49], and similarly with **IL**P_0 and **IL**R [50].

A drawback of this approach is in that the FMP w.r.t. ordinary Veltman semantics does not follow from the FMP w.r.t. generalised Veltman semantics. So, if that is what we are interested it, we do not provide an easy way for such results. However, decidability can be obtained from the FMP w.r.t. either semantics (unless the logic in question is incomplete w.r.t. ordinary Veltman semantics, as is the case with $\mathbf{IL}\mathsf{P}_0$, in which case we actually need the FMP w.r.t. generalised semantics). At the moment it is not clear whether the choice of semantics would affect our ability to produce results regarding computational complexity of provability and consistency of the logic at hand (see the next chapter).

A filtration is often generated by logical equivalence over some appropriate set of formulas. Here we use bisimilarity instead, i.e. we merge two worlds if they are bisimilar according to at least one bisimulation. We will later see that such construction makes sense. Applying this construction yields finite models.

5.1 Preliminaries

Let us overview basic notions and results of [53]. As we announced, we will not follow the content presented there faithfully; but we will give proofs of our statements whenever we diverge from the original papers.

A note on notation: given some relation $\mathcal{R}$, in this chapter we write $\mathcal{R}\big[x\big]$ to denote the set $\{y : xRy\}$. In this chapter we often need to write $\mathcal{R}\Big[[x]\Big]$ and wish to avoid writing $\mathcal{R}[[x]]$; so we increased the font size for the outer brackets.

Let A be a formula. If A equals $\neg B$ for some B, then $\sim A$ is B, otherwise $\sim A$ is $\neg B$. This is the "single negation" operation which we already met in the chapter concerning completeness.

We need a notion of *adequacy*, i.e. when is a set of formulas Γ "adequate".[1] Our filtration will start with a (possibly infinite) model and an adequate set we are interested in. Based on these choices obtain a finite model. Already in **GL**, the set $\{\Diamond p, \Diamond\Diamond p, \Diamond\Diamond\Diamond p, \dots\}$ has no finite models. Thus, the desired notion of equivalence has to be restricted somehow, and this is why we need adequacy. The notion of semantic equivalence between the starting model and the finite model we obtain later will be precisely the equivalence w.r.t. this adequate set. Notice that we choose the adequate set beforehand, and this set affects the content of the obtained finite model. Our operation is an operation on *models*, and not *frames*. In general this fact may be undesirable since dependence on a valuation may lead to the loss of the appropriate frame condition (characteristic property). As it turns out, the construction we use preserves the characteristic property in all known cases despite its dependence on the forcing relation.

Visser [66] defined the notion of bisimulation between Veltman models. Vrgoč and Vuković [69] extended this definition to generalised Veltman models. Here we explicate that the notion of a bisimulation depends on a set of propositional variables, which need

[1] Note that this is a different notion of *adequacy* than the one used for completeness proofs in [18], [19], and [27]. In this chapter we deal with, and only with, semantics. This eliminates most or all of the hassle usually present when one has to take syntax into account.

not be the set of all propositional variables.[2]

Definition 5.1 ([69]) Let Prop be a (possibly finite) subset of the set of all propositional variables. A bisimulation between generalised Veltman models $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ and $\mathfrak{M}' = (W', R', \{S'_{w'} : w' \in W'\}, \Vdash)$ w.r.t. Prop is a non-empty relation $Z \subseteq W \times W'$ such that:

(at) if wZw', then $w \Vdash p$ if and only if $w' \Vdash p$, for all $p \in \mathsf{Prop}$;

(forth) if wZw' and wRu, then there is $u' \in W'$ such that $w'R'u'$, uZu' and for all $V' \subseteq W'$ such that $u'S'_{w'}V'$ there is $V \subseteq W$ such that uS_wV and for all $v \in V$ there is $v' \in V'$ with vZv';

(back) if wZw' and $w'R'u'$, then there is $u \in W$ such that wRu, uZu' and for all $V \subseteq W$ such that uS_wV there is $V' \subseteq W'$ such that $u'S'_{w'}V'$ and for all $v' \in V'$ there is $v \in V$ with vZv'.

The contents of Prop will usually be left unspecified, since we expect its value to be fixed and constant in all contexts. Ultimately we aim to use bisimulations with Prop equalling the set of all propositional variables occurring in Γ.

Definition 5.2 Let $\mathfrak{M}$ and $\mathfrak{M}'$ be arbitrary generalised Veltman models. We write $w \equiv_n^{\mathsf{Prop}} w'$ if $w \in W$ and $w' \in W'$, and for all modal formulas A whose modal depth is at most n and whose propositional variables are contained in the set Prop we have

$$w \Vdash A \text{ if and only if } w' \Vdash A.$$

If Prop is unspecified, we assume quantification over all propositional variables; similarly if n is unspecified, we assume quantification over formulas of all modal depths.

Given a generalised Veltman model $\mathfrak{M}$, the union of all bisimulations on $\mathfrak{M}$, denoted by $\sim$, is the largest bisimulation on $\mathfrak{M}$, and $\sim$ is an equivalence relation [69]:

Lemma 5.3 ([69]) Let $\mathfrak{M}$, $\mathfrak{M}'$ and $\mathfrak{M}''$ be generalised Veltman models.

(1) If $w \in W$ and $w' \in W'$ are bisimilar w.r.t. some set Prop, then $w \equiv^{\mathsf{Prop}} w'$.

(2) The identity $\{(w, w) : w \in W\} \subseteq W \times W$ is a bisimulation.

(3) The inverse of a bisimulation between $\mathfrak{M}$ and $\mathfrak{M}'$ is a bisimulation between $\mathfrak{M}'$ and $\mathfrak{M}$.

(4) The composition of bisimulations $Z \subseteq W \times W'$ and $Z' \subseteq W' \times W''$ is a bisimulation between $\mathfrak{M}$ and $\mathfrak{M}''$.

[2] On the topic of bisimulations, it may be interesting that in [14], Čačić and Vrgoč defined the notion of a game for Veltman models and proved that a winning strategy for the defender in such a game is equivalent to picking out a bisimulation between two models.

(5) The union of a family of bisimulations between $\mathfrak{M}$ and $\mathfrak{M}'$ is also a bisimulation between $\mathfrak{M}$ and $\mathfrak{M}'$. Thus there exists the largest bisimulation between models $\mathfrak{M}$ and $\mathfrak{M}'$.

Fix any set of propositional variables Prop. Fact (5) of the preceding lemma implies that there is the largest bisimulation $\sim$ on $\mathfrak{M}$. Facts (2), (3) and (4) of the previous lemma imply that $\sim$ is an equivalence relation, while (1) tells us that bisimilarity implies modal equivalence. From now on, we will only need this largest bisimulation $\sim$. Again, we won't make the dependence $\mathsf{Prop} \mapsto \sim$ explicit since we never work with more than one set of propositional variables in any context.

A $\sim$-equivalence class of $w \in W$ will be denoted by $[w]$. For any set of worlds V, put $\widetilde{V} = \{[w] : w \in V\}$.

Definition 5.4 A *filtration* of $\mathfrak{M}$ through Γ, $\sim$ is any generalised Veltman model $\widetilde{\mathfrak{M}} = (\widetilde{W}, \widetilde{R}, \{\widetilde{S}_{[w]} : w \in W\}, \Vdash)$ such that for all $w \in W$ and $A \in \Gamma$ we have $w \Vdash A$ if and only if $[w] \Vdash A$ (we denote both forcing relations as $\Vdash$, as there is no risk of confusion).

Before describing in what way is our particular filtration constructed, let us first introduce a tool we will use to prove finiteness.

Definition 5.5 An *n-bisimulation* between generalised Veltman models $(W, R, \{S_w : w \in W\}, \Vdash)$ and $(W', R', \{S'_{w'} : w' \in W'\}, \Vdash)$ w.r.t. a set Prop is any sequence $Z_n \subseteq \cdots \subseteq Z_0 \subseteq W \times W'$:

(at) if wZ_0w' then $w \Vdash p$ if and only if $w' \Vdash p$ for all $p \in \mathsf{Prop}$;

(forth) if wZ_nw' and wRu, then there exists $u' \in R'\big[w'\big]$ with $uZ_{n-1}u'$ and for all $V' \in S'_{w'}\big[u'\big]$ there is $V \in S_w\big[u\big]$ such that for all $v \in V$ there is $v' \in V'$ with $vZ_{n-1}v'$;

(back) if wZ_nw' and $w'R'u'$, then there exists $u \in R\big[w\big]$ such that $uZ_{n-1}u'$ and for all $V \in S_w\big[u\big]$ there is $V' \in S'_{w'}\big[u'\big]$ such that for all $v' \in V'$ there is $v \in V$ with $vZ_{n-1}v'$.

We will use the same notion of adequacy for filtrations that we use for completeness (Definition 4.5).

Note 5.6 Note that the definition of adequacy we use for filtrations is not compatible with published papers [53], [49], and [50]. The way results are proven in those papers requires much more involved definitions of adequacy. The incompatibility affects most of the remaining content in this chapter too, so we will not stress particular incompatibilities in the remainder of the chapter.

If we restrict the class of models for which our filtration method is applicable to a special sort of models, and we will call these models *maximal* models, then the results of

[53], [49] and [50] can be proved in a slightly more succinct form. In particular, with this kind of models we only need to apply the filtration once in order to obtain a finite model. We already met such models in the previous chapter; all our completeness results were proven with respect to this restricted class of models (see Definition 4.6 and Definition 4.27).

Definition 5.7 Let Γ be an adequate set and $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ a generalised model. We say that $\mathfrak{M}$ is *maximal* w.r.t. Γ if for each $w \in W$ there is $A_w \in \Gamma$ such that $w \Vdash A_w$ and $R[w] \Vdash \neg A_w$.

Lemma 5.8 ([53], Lemma 3.1.2) Let $\mathfrak{M}$ be a generalised Veltman model, Prop a finite set of propositional variables and $w, w' \in \mathfrak{M}$. Then w and w' are n-bisimilar w.r.t. Prop if and only if $w \equiv_n^{\mathsf{Prop}} w'$.

Note that given some finite adequate set Γ and a subset Prop of the set of all propositional variables, the height of a maximal model $\mathfrak{M}$ is bounded by $|\Gamma|$. Due to this, worlds in $\mathfrak{M}$ are $|\Gamma|$-bisimilar w.r.t. Prop if and only if they are bisimilar w.r.t. Prop. Thus, worlds w and w' in $\mathfrak{M}$ are bisimilar w.r.t. Prop if and only if $w \equiv_{|\Gamma|}^{\mathsf{Prop}} w'$.

The following lemma combines the key results of [53] (Lemma 2.3, Theorem 2.4., Theorem 3.2).

Lemma 5.9 Let Γ be an adequate set and $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ a model that is maximal w.r.t. Γ. Let $\sim$ denote the largest bisimulation on $\mathfrak{M}$. Define:[3]

(1) $[w]\,\widetilde{R}\big[u\big]$ if and only if for some $w' \in [w]$ and $u' \in [u]$ we have $w'Ru'$.

(2) $[u]\,\widetilde{S}_{[w]}\widetilde{V}$ if and only if $[w]\,\widetilde{R}\big[u\big]$, $\widetilde{V} \subseteq \widetilde{R}\big[[w]\big]$, and for all $w' \in [w]$ and $u' \in [u]$ such that $w'Ru'$ we have $u'S_{w'}V'$ for some V' such that $\widetilde{V'} \subseteq \widetilde{V}$;

(3) for all propositional variables $p \in \Gamma$ put $[w] \Vdash p$ if and only if $w \Vdash p$, and for all propositional variables $q \notin \Gamma$ put $[w] \nVdash q$ for all $[w] \in \widetilde{W}$.

Then $\widetilde{\mathfrak{M}} = (\widetilde{W}, \widetilde{R}, \{\widetilde{S}_{[w]} : w \in W\}, \Vdash)$ is a model and a filtration of $\mathfrak{M}$ through $\Gamma, \sim$. Furthermore, $\widetilde{\mathfrak{M}}$ is maximal w.r.t. Γ and finite.

Proof. We first prove that $\widetilde{M}$ is a model.

1. It is easy to see that $\widetilde{R}$ is a binary relation on $\widetilde{W}$, and $\widetilde{S}_{[w]} \subseteq \widetilde{W} \times 2^{\widetilde{W}} \setminus \{\emptyset\}$ for all $w \in W$.

2. Transitivity of $\widetilde{R}$. Assume $[w]\,\widetilde{R}\,[u]\,\widetilde{R}\,[v]$. For some $w' \in [w]$, $u', u'' \in [u]$ and $v'' \in [v]$ we have $w'Ru' \sim u''Rv''$. Thus, there is $v' \sim v''$ with $u'Rv'$. By the transitivity of R we have $w'Rv'$, so $[w]\,\widetilde{R}\,[v]$.

3. Converse well-foundedness of $\widetilde{R}$. Assume $[w_1]\,\widetilde{R}\,[w_2]\,\widetilde{R}\,[w_3]\,\widetilde{R}\ldots$ We prove by induction that for every w_i there is $w'_i \sim w_i$ and $w'_1Rw'_2R\ldots Rw'_i$. Base case is $i = 1$, and

[3] The set of worlds of the new model is $\widetilde{W}$, that is, the set $\{[w] : w \in W\}$. Unlike $\widetilde{W}$, relations $\widetilde{R}$ and $\widetilde{S}_{[w]}$ are not the product of applying the (previously introduced) operation $V \mapsto \widetilde{V}$, but rather completely new entities which we now define.

here we can let $w_1' = w_1$. Suppose the claim holds for all values strictly smaller than i, and let us prove the claim for i. Since $w_{i-1}' \sim w_{i-1}$ and $[w_{i-1}]\,R\,[w_i]$, there is some $w_i' \sim w_i$ such that $w_{i-1}'Rw_i'$. By induction hypothesis we know that $w_1'Rw_2'R\ldots Rw_{i-1}'$. Thus, $w_1'Rw_2'R\ldots Rw_{i-1}'Rw_i'$, as required.

4. Quasi-reflexivity of $\widetilde{S}_{[w]}$, for all $w \in W$. Assume $[w]\,\widetilde{R}\,[u]$. To prove $[u]\,\widetilde{S}_{[w]}\{[u]\}$ let $w' \in [w]$ and $u' \in [u]$ be arbitrary such that $w'Ru'$. By the quasi-reflexivity of $S_{w'}$, we have $u'S_{w'}\{u'\}$, and clearly $\widetilde{\{u'\}} \subseteq \{[u]\}$ (the two sets are equal).

5. Quasi-transitivity of $\widetilde{S}_{[w]}$, for all $w \in W$. Assume $[u]\,\widetilde{S}_{[w]}\widetilde{V}$ and for all $v \in V$, $[v]\,\widetilde{S}_{[w]}\widetilde{Z_v}$. We claim that $[u]\,\widetilde{S}_{[w]}\widetilde{\bigcup_{v\in V} Z_v}$. So, take $w' \in [w]$ and $u' \in [u]$ such that $w'Ru'$. Since $[u]\,\widetilde{S}_{[w]}\widetilde{V}$, there is $V_{w',u'}$ such that $u'S_{w'}V_{w',u'}$ and $\widetilde{V_{w',u'}} \subseteq \widetilde{V}$. Let $v' \in V_{w',u'}$ be arbitrary, and let $v \in V$ be the element such that $v' \sim v$. Since $[v]\,\widetilde{S}_{[w]}\widetilde{Z_v}$, there is $T_{w',u',v'}$ such that $v'S_{w'}T_{w',u',v'}$ and $\widetilde{T_{w',u',v'}} \subseteq \widetilde{Z_v}$. Define $T_{w',u'} = \bigcup_{v'\in V_{w',u'}} T_{w',u',v'}$. By the quasi-transitivity of $S_{w'}$, we have $u'S_{w'}T_{w',u'}$. Since $v' \in V_{w',u'}$ was arbitrary in the definition of $T_{w',u',v'}$, clearly $\widetilde{T_{w',u'}} \subseteq \widetilde{\bigcup_{v\in V} Z_v}$.

6. The property that $[w]\,\widetilde{R}\,[u]\,\widetilde{R}\,[v]$ implies $[u]\,\widetilde{S}_{[w]}\{[v]\}$. Let $w' \in [w]$, $u', u'' \in [u]$ and $v'' \in [v]$ be the worlds such that $w'Ru' \sim u''Rv''$. Then there is $v' \sim v''$ such that $u'Rv'$. So, $u'S_{w'}\{v'\}$. Clearly $\widetilde{\{v'\}} \subseteq \{[v]\}$.

7. Monotonicity is immediate.

8. The forcing relation $\Vdash$ is well-defined since it does not depend on a representative of the class $[w]$.

Next we prove that the model $\widetilde{\mathfrak{M}}$ is a filtration of $\mathfrak{M}$ through $\Gamma, \sim$.

We need to check if all truth values coincide, i.e. $w \Vdash A$ if and only if $[w] \Vdash A$. We prove this by induction on the complexity, and as usual we focus on the formulas of the form $A \rhd B$.

Assume $w \nVdash A \rhd B$. Then there is $[u] \in [A]_w$ such that if uS_wV, then $V \nVdash B$.

Let $\widetilde{V}$ be arbitrary such that $[u]\,\widetilde{S}_{[w]}\widetilde{V}$. Then uS_wV' for some $\widetilde{V'} \subseteq \widetilde{V}$. Since $V' \nVdash B$, we get $\widetilde{V'} \nVdash B$ by the induction hypothesis. Therefore, $\widetilde{V} \nVdash B$.

For the other direction, assume $w \Vdash A \rhd B$. Assume $[w]\,\widetilde{R}\,[u] \Vdash A$. We construct $\widetilde{V}$ such that $[u]\,\widetilde{S}_{[w]}\widetilde{V} \Vdash B$. Let $w' \in [w]$ and $u' \in [u]$ be arbitrary such that wRu. Since $w' \sim w$, $w' \Vdash A \rhd B$, and therefore for some $V(w', u')$ we have $u'S_{w'}V(w', u') \Vdash B$. Put $V := \bigcup_{w'\in[w],u'\in[u],w'Ru'} V(w', u')$. By the induction hypothesis, $\widetilde{V} \Vdash B$. To obtain $[u]\,\widetilde{S}_{[w]}\widetilde{V} \Vdash B$ it remains to show that $\widetilde{V} \subseteq R([w])$. Obviously, if $[v] \in \widetilde{V}$, then for some w', u', and v' we have $u'S_{w'}V(w', u') \ni v'$, for some $v' \sim v$. Clearly $w'Rv'$, so $[w]\,\widetilde{R}\,[v]$, as required.

Now we prove that the model $\widetilde{\mathfrak{M}}$ is maximal w.r.t. Γ.

Suppose that for some $w \in W$ there is no $A \in \Gamma$ such that $[w] \Vdash A$ and $\widetilde{R}\big[[w]\big] \Vdash \neg A$. Since $\mathfrak{M}$ is maximal, there is $A_w \in \Gamma$ such that $w \Vdash A_w$ and $R[w] \Vdash \neg A_w$. Since $\widetilde{\mathfrak{M}}$ is a filtration, $[w] \Vdash A_w$. Since $\widetilde{R}\big[[w]\big] \nVdash \neg A_w$ by assumption, there must be u such that $[w]\,\widetilde{R}\,[u] \Vdash A_w$. Since $\widetilde{\mathfrak{M}}$ is a filtration, $u \Vdash A_w$. This contradicts the fact that

$R[w] \Vdash \neg A_w$.

Finally, we prove that the model $\widetilde{\mathfrak{M}}$ is finite.

Recall that worlds w and w' in a maximal model are bisimilar w.r.t. Prop if and only if $w \equiv^{\mathsf{Prop}}_{|\Gamma|} w'$. We have just seen that $\widetilde{\mathfrak{M}}$ is maximal w.r.t. Γ. Thus, every world $[w] \in \widetilde{\mathfrak{M}}$ corresponds to an $\equiv^{\mathsf{Prop}}_{|\Gamma|}$-class. There are only finitely many formulas of modal depth bounded by $|\Gamma|$ and containing a finite number of variables, up to equivalence. Thus, the number of worlds in $\widetilde{\mathfrak{M}}$ is finite.

$\dashv$

Lemma 5.9 implies that **IL** has the FMP w.r.t. generalised Veltman semantics. To prove that a specific extension has the FMP, it remains to show that filtration preserves its characteristic property.

Since we are going to use **IL**X-structures (see Definition 4.6 and Definition 4.27) as the starting models $\mathfrak{M}$, we can use the fact that they are maximal (see Chapter 4).

5.2 The finite model property of ILW and ILW*

In this section we prove that if a generalised Veltman model $\mathfrak{M}$ possesses the property $(\mathsf{W})_{\mathrm{gen}}$ then the filtration $\widetilde{\mathfrak{M}}$ also possesses the property $(\mathsf{W})_{\mathrm{gen}}$. As a result we obtain not only the finite model property of **IL**W, but also the finite model property of **IL**W* (when combining with the results of [53]).

The results of this section together with the completeness of **IL**W w.r.t. generalised Veltman models, imply decidability of **IL**W. We will discuss this in more detail in the next section; here we only deal with the finite model property.

Note that de Jongh and Veltman [19] already proved the completeness of the system **IL**W w.r.t. finite Veltman models, which implies the finite model property of **IL**W. However, the ultimate goal here is to prove the finite model property of **IL**W*. To obtain this result we cannot reuse the existing proofs that **IL**W has the finite model property. Since **IL**W* = **IL**WM$_0$ (see [68]) and we know that the filtration preserves the property $(\mathsf{M_0})_{\mathrm{gen}}$, we need to show that it preserves $(\mathsf{W})_{\mathrm{gen}}$ too. For this reason the well-known fact that **IL**W has the finite model property is not directly applicable.

Lemma 5.10 Let Γ be an adequate set of formulas. Let $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ be a generalised Veltman model that is maximal w.r.t. Γ and also satisfies $(\mathsf{W})_{\mathrm{gen}}$. Let $\sim$ be the largest bisimulation on $\mathfrak{M}$. Then the generalised model $\widetilde{\mathfrak{M}} = (\widetilde{W}, \widetilde{R}, \{\widetilde{S}_{[w]} : w \in W\}, \Vdash)$ (Lemma 5.9) satisfies the condition $(\mathsf{W})_{\mathrm{gen}}$.

Proof. Let $w \in W$ be a world and X a non–empty set such that $\widetilde{X} \subseteq \widetilde{R}\big[[w]\big]$. For each world $u \in [w]$ we define sets X_u and Z_u as follows:

$$X_u = \{x \in \bigcup \widetilde{X} \mid uRx\} \ \text{ and } \ Z_u = \{z \in \bigcup \widetilde{S}^{-1}_{[w]}\big[\widetilde{X}\big] \mid uRz\}.$$

First we prove that for each $u \in [w]$ the set X_u is non–empty. Let $x \in \bigcup \widetilde{X}$ be a world. By assumption, we have $\widetilde{X} \subseteq \widetilde{R}\big[[w]\big]$, so in particular $[w]\,\widetilde{R}\,[x]$. The definition of the relation $\widetilde{R}$ implies there are worlds $w' \in [w]$ and $x' \in [x]$ such that $w'Rx'$ holds. Now $u \sim w'$, $w'Rx'$ and the property (back) of the bisimulation $\sim$ imply that there exists a world $u' \in W$ such that $x' \sim u'$ and uRu'. Obviously we have $u' \in \bigcup \widetilde{X}$, and thus also $u' \in X_u$. Hence, the set X_u is non–empty. The proof that the set Z_u is non–empty for every $u \in [w]$ is completely analogous.

We intend to apply the property $(\mathsf{W})_{\mathrm{gen}}$ to all $u \in [w]$, to the set X_u and all $y \in Z_u$.

First we prove that for every world $u \in [w]$ we have $Z_u \subseteq S_u^{-1}\big[X_u\big]$. Consider any world $u \in [w]$. Assume that there exists a world $z \in Z_u$ such that $z \notin S_u^{-1}\big[X_u\big]$ holds, i.e. zS_uX_u does not hold. By the definition of the set Z_u we have uRz. Since $z \in \bigcup \widetilde{S}_{[w]}^{-1}\big[\widetilde{X}\big]$ we also have $[z]\,\widetilde{S}_{[w]}\widetilde{X}$. Now uRz, $u \in [w]$ and the definition of $\widetilde{S}_{[w]}$ imply that there exists a set V such that zS_uV and $\widetilde{V} \subseteq \widetilde{X}$. We now prove that $V \subseteq X_u$. Let x be an element of the set V. From $\widetilde{V} \subseteq \widetilde{X}$ we have in particular that $[x] \in \widetilde{X}$, and thus also $x \in \bigcup \widetilde{X}$. From zS_uV we have in particular that $V \subseteq R\big[u\big]$, so uRx. Hence, $x \in X_u$.

Now zS_uV, $V \subseteq X_u \subseteq R\big[u\big]$, and the monotonicity of the relation S_u imply zS_uX_u, which contradicts the assumption. We have now proved that for each $u \in [w]$ we have $Z_u \subseteq S_u^{-1}\big[X_u\big]$, i.e. for each $y \in Z_u$ we have yS_uX_u.

By applying the assumption that the model $\mathfrak{M}$ has the property $(\mathsf{W})_{\mathrm{gen}}$ to any $u \in [w]$, X_u and any $y \in Z_u$ we get:

$$(5) \qquad (\exists V_{y,u} \subseteq X_u)\Big(yS_uV_{y,u} \;\&\; (\forall v \in V_{y,u})(R\big[v\big] \cap S_u^{-1}\big[X_u\big] = \emptyset)\Big).$$

Recall that we have chosen a world $w \in W$ and a set $X \subseteq W$ such that $\widetilde{X} \subseteq \widetilde{R}\big[[w]\big]$. So, to prove that the model $\widetilde{\mathfrak{M}}$ possesses the property $(\mathsf{W})_{\mathrm{gen}}$, we will prove the following:

$$\begin{aligned}(\forall\,[z] \in \widetilde{W})\Big(\,[z]\,\widetilde{S}_{[w]}\widetilde{X} \;\;\Rightarrow\;\; & (\exists \widetilde{V_z} \subseteq \widetilde{X})([z]\,\widetilde{S}_{[w]}\widetilde{V_z} \;\& \\ & (\forall v \in V_z)(\widetilde{R}\big[[v]\big] \cap \widetilde{S}_{[w]}^{-1}\big[\widetilde{X}\big] = \emptyset))\Big).\end{aligned}$$

In order to prove the preceding claim, consider any world $z \in \bigcup \widetilde{W}$ with $[z]\,\widetilde{S}_{[w]}\widetilde{X}$. For every $u \in [w]$ let $Y_{z,u} = \{y \in Z_u \mid y \sim z\}$. We first prove that for each $u \in [w]$ the set $Y_{z,u}$ is non–empty. Since $[z]\,\widetilde{S}_{[w]}\widetilde{X}$, we have in particular that $[w]\,\widetilde{R}\,[z]$. By the definition of the relation $\widetilde{R}$ it follows that there are worlds $w' \in [w]$ and $z' \in [z]$ such that $w'Rz'$. Facts $u \sim w'$, $w'Rz'$, and the property (back) of the bisimulation $\sim$ imply that there exists a world $u' \in W$ such that $u' \sim z'$ and uRu'. Obviously we have $u' \in [z]$, and thus $u' \in Y_{z,u}$, so the set $Y_{z,u}$ is non–empty for every $u \in [w]$.

In order to define the set $\widetilde{V_z}$ with the desired properties we first define certain sets for every world $u \in [w]$. Consider any worlds $u \in [w]$ and $y \in Y_{z,u}$. From the definition of the set $Y_{z,u}$ we have $y \in Z_u$. From the fact labelled with (5) we have that there exists a set $V_{y,u} \subseteq X_u$ with the following property:

$$(6) \qquad yS_uV_{y,u} \ \& \ (\forall v \in V_{y,u})(R\big[v\big] \cap S_u^{-1}\big[X_u\big] = \emptyset).$$

Now we define the set V_z as follows:

$$V_z = \bigcup_{u\in[w],\ y\in Y_{z,u}} V_{y,u}.$$

We now proceed to prove that the set $\widetilde{V_z}$ has the desired property, i.e. that the following holds:

$$\widetilde{V_z} \subseteq \widetilde{X} \ \& \ [z]\,\widetilde{S}_{[w]}\widetilde{V_z} \ \& \ (\forall v \in V_z)(\widetilde{R}\big[[v]\big] \cap \widetilde{S}_{[w]}^{-1}\big[\widetilde{X}\big] = \emptyset).$$

We have $V_{y,u} \subseteq X_u$ for every $u \in [w]$ and $y \in Y_{z,u}$; therefore $\widetilde{V_z} \subseteq \bigcup_{u\in[w]} \widetilde{X_u} \subseteq \widetilde{X}$.

We now prove that $[z]\,\widetilde{S}_{[w]}\widetilde{V_z}$. Consider any worlds u and y such that $u \in [w]$, $y \in [z]$, and uRy. From the definition of the set $Y_{z,u}$ it follows that $y \in Y_{z,u}$. From the fact labelled with (6) we have in particular that $yS_uV_{y,u}$. From the definition of the set V_z we have $\widetilde{V_{y,u}} \subseteq \widetilde{V_z}$. We have already shown that $[w]\,\widetilde{R}\,[z]$. Since $\widetilde{V_z} \subseteq \widetilde{X}$, and $\widetilde{X} \subseteq \widetilde{R}\big[[w]\big]$ by the assumption, we also have $\widetilde{V_z} \subseteq \widetilde{R}\big[[w]\big]$. From the definition of $\widetilde{S}_{[w]}$ we now have that $[z]\,\widetilde{S}_{[w]}\widetilde{V_z}$.

It remains to show that $\widetilde{R}\big[[v]\big] \cap \widetilde{S}_{[w]}^{-1}\big[\widetilde{X}\big] = \emptyset$, for every $v \in V_z$. Suppose the contrary. Then there are worlds $v \in V_z$ and $s \in W$ such that $[s] \in \widetilde{R}\big[[v]\big] \cap \widetilde{S}_{[w]}^{-1}\big[\widetilde{X}\big]$. From here it follows in particular that $[v]\,\widetilde{R}\big[s\big]$ and $[s]\,\widetilde{S}_{[w]}\widetilde{X}$. The fact that $v \in V_z$ and the definition of the set V_z imply that there are worlds $u \in [w]$ and $y \in Y_{z,u}$ such that $v \in V_{y,u}$. From the fact that $[v]\,\widetilde{R}\big[s\big]$, and by the definition of the relation $\widetilde{R}$, it follows that there are worlds $v' \in [v]$ and $s' \in [s]$ such that $v'Rs'$. Now $v \sim v'$, $v'Rs'$ and the property (back) of the bisimulation $\sim$ imply that there exists a world s'' such that vRs'' and $s'' \sim s'$. From $s'' \sim s' \sim s$ and $[s]\,\widetilde{S}_{[w]}\widetilde{X}$ it follows that $s'' \in \bigcup \widetilde{S}_{[w]}^{-1}\big[\widetilde{X}\big]$. From the fact labelled with (6) we know that $yS_uV_{y,u}$, and so in particular we have $V_{y,u} \subseteq R\big[u\big]$. Since we have $v \in V_{y,u}$, we also have $v \in R\big[u\big]$, i.e. uRv. Now uRv and vRs'', and the transitivity of the relation R imply that uRs''. Hence $s'' \in \bigcup \widetilde{S}_{[w]}^{-1}\big[\widetilde{X}\big]$ and uRs''. By the definition of the set Z_u we have $s'' \in Z_u$. This contradicts the fact labelled with (6), i.e. the fact that $R\big[v\big] \cap Z_u = \emptyset$. ⊣

Corollary 5.11 The logic **IL**W^* has the finite model property with respect to generalised Veltman models which satisfy conditions $(\mathsf{W})_{\text{gen}}$ and $(\mathsf{M}_0)_{\text{gen}}$.

5.3 Decidability of ILW and ILW^*

In [53] it was proved that **IL**M_0 has the finite model property with respect to generalised Veltman models which satisfy the condition $(\mathsf{M}_0)_{\text{gen}}$. Together with the completeness of **IL**$\mathsf{M_0}$ with respect to (maximal) generalised Veltman models, this suffices to prove that **IL**$\mathsf{M_0}$ is decidable, by a standard argument (cf. [4], p. 341):

- the set of theorems of **IL**$\mathsf{M_0}$ is recursively enumerable,
- the set (up to isomorphism) of finite generalised Veltman models with the property $(\mathsf{M_0})_{\mathrm{gen}}$ is recursively enumerable,
- we construct an algorithm which simultaneously enumerates theorems of **IL**$\mathsf{M_0}$, which are compared to a given formula A, and generalised Veltman models with the property $(\mathsf{M_0})_{\mathrm{gen}}$, on which the truth of $\neg A$ is tested.

The finite model property implies that the algorithm will either find the generalised Veltman model with the property $(\mathsf{M_0})_{\mathrm{gen}}$ in which $\neg A$ is satisfied, or establish that the formula A is a theorem of **IL**$\mathsf{M_0}$, in finitely many steps.

By an analogous argument, the finite model property and completeness of **IL**W^* w.r.t. generalised semantics implies the decidability of **IL**W^*. Thus, we have proved the following theorem.

Theorem 5.12 The systems **IL**$\mathsf{M_0}$, **IL**W and **IL**W^* are decidable.

5.4 The FMP and decidability for IL$\mathsf{P_0}$ and ILR

In this final section we repeat our construction, this time for **IL**$\mathsf{P_0}$ and **IL**R. A small difference is that here we don't have completeness w.r.t. ordinary semantics (**IL**$\mathsf{P_0}$), or completeness w.r.t. generalised semantics is still an open question (**IL**R). For present purposes this simply means that there is only one way of proving decidability: through generalised semantics. So, here we rely on results of Chapter 4.

Lemma 5.13 Let Γ be an adequate set of formulas. Let $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ be a generalised Veltman model that is maximal w.r.t. Γ and also satisfies $(\mathsf{P_0})_{\mathrm{gen}}$. Let $\sim$ be the largest bisimulation on $\mathfrak{M}$. Then the generalised model $\widetilde{\mathfrak{M}} = (\widetilde{W}, \widetilde{R}, \{\widetilde{S}_{[w]} : w \in W\}, \Vdash)$ (Lemma 5.9) satisfies the condition $(\mathsf{P_0})_{\mathrm{gen}}$.

Proof. Assume $[w]\,\widetilde{R}\,[x]\,\widetilde{R}\,[u]\,\widetilde{S}_{[w]}V$ and $\widetilde{R}\Big[[v]\Big] \cap Z \neq \emptyset$ for each $[v] \in V$. We claim that there exists $Z' \subseteq Z$ such that $[u]\,\widetilde{S}_{[x]}Z'$.

Since $[w]\,\widetilde{R}\,[x]$, there are $w_0 \in [w]$ and $x_0 \in [x]$ such that $w_0 R x_0$. Let $x' \in [x]$ and $u' \in [u]$ be any worlds such that $x' R u'$. The condition (back) implies that there is a world $u_{x',u'}$ such that $x_0 R u_{x',u'}$ and $u_{x',u'} \sim_{\mathfrak{M}} u'$. Now, $[u]\,\widetilde{S}_{[w]}V$, $u_{x',u'} \in [u]$ and $w_0 R u_{x',u'}$ imply there is a set $V_{x',u'}$ such that $u_{x',u'} S_{w_0} V_{x',u'}$ and $\widetilde{V_{x',u'}} \subseteq V$. Since $\widetilde{R}\Big[[v]\Big] \cap Z \neq \emptyset$ for each $[v] \in V$, we have $\widetilde{R}\Big[[v]\Big] \cap Z \neq \emptyset$ for each $v \in V_{x',u'}$. For each $v \in V_{x',u'}$, choose a world z_v such that $[z_v] \in \widetilde{R}\Big[[v]\Big] \cap Z$. Now $[v]\,\widetilde{R}\,[z_v]$ implies that there are some $v' \in [v]$ and

$z'_v \in [z_v]$ such that $v'Rz'_v$. Applying (back), we can find a world z''_v such that vRz''_v and $z'_v \sim z''_v$. Put $Z_{x',u'} = \{z''_v : v \in V_{x',u'}\}$. Note that we have $R\big[v\big] \cap Z_{x',u'} \neq \emptyset$ for each $v \in V_{x',u'}$.

Applying $(\mathsf{P}_0)_{\text{gen}}$ gives $u_{x',u'}S_{x_0}Z'_{x',u'}$ for some $Z'_{x',u'} \subseteq Z_{x',u'}$. Clearly $\widetilde{Z'_{x',u'}} \subseteq \widetilde{Z_{x',u'}} \subseteq Z$. Continuing our first application of (back), there is a set $Z''_{x',u'}$ such that $u'S_{x'}Z''_{x',u'}$, and for each $z'' \in Z''_{x',u'}$ there is $z' \in Z'_{x',u'}$ such that $z' \sim z''$. This implies $\widetilde{Z''_{x',u'}} \subseteq \widetilde{Z'_{x',u'}}$. Let $T = \bigcup_{x' \in [x], u' \in [u], x'Ru'} Z''_{x',u'}$ and $Z' = \widetilde{T}$. It is easy to see that $Z' \subseteq Z$ and $Z' \subseteq \widetilde{R}\big[[x]\big]$. We have $u'S_{x'}Z''_{x',u'}$ with $\widetilde{Z''_{x',u'}} \subseteq Z'$ for all $x' \in [x]$ and $u' \in [u]$ with $x'Ru'$. Thus, $[u]\,\widetilde{S}_{[x]}Z'$. ⊣

Corollary 5.14 $\mathbf{IL}\mathsf{P}_0$ is decidable.

Proof. Since $\mathbf{IL}\mathsf{P}_0$ is complete, it remains to show that it has the finite model property. Let $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ be the $\mathbf{IL}\mathsf{P}_0$-structure for an appropriate $\mathcal{D}$, and apply Lemma 5.13. As the resulting model $\widetilde{\mathfrak{M}}$ itself also satisfies the conditions of Lemma 5.13, we can apply Lemma 5.13 once more, and by Lemma 5.9 obtain a finite model. ⊣

Let us prove the same for $\mathbf{IL}\mathsf{R}$.

Lemma 5.15 Let Γ be an adequate set of formulas. Let $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ be a generalised Veltman model that is maximal w.r.t. Γ and also satisfies $(\mathsf{R})_{\text{gen}}$. Let $\sim$ be the largest bisimulation on $\mathfrak{M}$. Then the generalised model $\widetilde{\mathfrak{M}} = (\widetilde{W}, \widetilde{R}, \{\widetilde{S}_{[w]} : w \in W\}, \Vdash)$ (Lemma 5.9) satisfies the condition $(\mathsf{R})_{\text{gen}}$.

Proof. Assume $[w]\,\widetilde{R}\,[x]\,\widetilde{R}\,[u]\,\widetilde{S}_{[w]}V$, and let $C \in \mathcal{C}([x], [u])$ be an arbitrary choice set. We are to prove that there is a set U such that $\widetilde{U} \subseteq V$, $[x]\,\widetilde{S}_{[w]}\widetilde{U}$ and $\widetilde{R}\big[\widetilde{U}\big] \subseteq C$.

Put $C_{x'} = \{c \in R\big[x'\big] : [c] \in C\}$ for all $x' \in [x]$.

Let us first prove that for some $x_0 \in [x], u_0 \in [u]$ with x_0Ru_0 we have $C_{x_0} \in \mathcal{C}(x_0, u_0)$. Suppose not. Then for all $x' \in [x], u' \in [u]$ with $x'Ru'$, there is a set $Z_{x',u'}$ such that $u'S_{x'}Z_{x',u'}$ with $Z_{x',u'} \cap C_{x'} = \emptyset$. Put $Z = \bigcup_{x' \in [x], u' \in [u], x'Ru'} Z_{x',u'}$. Thus $\widetilde{Z} \subseteq \widetilde{R}\big[[x]\big]$. Thus $[u]\,\widetilde{S}_{[x]}\widetilde{Z}$. Since $C \in \mathcal{C}([x], [u])$, there is $z \in Z$ such that $[z] \in C \cap \widetilde{Z}$. Thus $z \in Z_{x',u'}$ for some $x' \in [x], u' \in [u]$ and $x'Ru'$. The definition of $C_{x'}$ implies $z \in C_{x'}$. Thus, $Z_{x',u'} \cap C_{x'} \neq \emptyset$, a contradiction.

Now we claim that for all $y \in [x]$ there is $u_y \sim u_0$ with yRu_y and $C_y \in \mathcal{C}(y, u_y)$. Since $y \sim x_0$ and x_0Ru_0, the (back) condition implies that there is a world u_y such that $u_y \sim u_0$ and yRu_y (and other properties that we will return to later). We will show that $C_y \in \mathcal{C}(y, u_y)$. Let Z' be such that u_yS_yZ', and we are to prove that $C_y \cap Z' \neq \emptyset$. The earlier instance of (back) condition for u_y further implies that there is a set Z with $u_0S_{x_0}Z$, and for all $z \in Z$ there is $z' \in Z'$ with $z \sim z'$. Let $z \in Z \cap C_{x_0}$ be an arbitrary element (which exists because, as we proved, C_{x_0} is a choice set). Then there is $z' \in Z'$ such that $z' \sim z$. Since $[z] \in C$, i.e. $[z'] \in C$, we have $z' \in C_y$. In particular, $Z' \cap C_y \neq \emptyset$. Thus, $C_y \in \mathcal{C}(y, u_y)$.

Let us prove that there is a set U such that $\widetilde{U} \subseteq V$, $[x]\,\widetilde{S}_{[w]}\widetilde{U}$ and $\widetilde{R}\Big[\widetilde{U}\Big] \subseteq C$. Let $w' \in [w]$ and $y \in [x]$ be such that $w'Ry$. Since $[u]\,\widetilde{S}_{[w]}V$, there is a set $V_{w',y}$ such that $u_y S_{w'} V_{w',y}$ and $\widetilde{V_{w',y}} \subseteq V$. Applying $(\mathsf{R})_{\text{gen}}$ with C_y, there is $U_{w',y} \subseteq V_{w',y}$ such that $yS_{w'}U_{w',y}$ and $R\Big[U_{w',y}\Big] \subseteq C_y$. Let $U = \bigcup_{w' \in [w], y \in [x], w'Ry} U_{w',y}$. Clearly $\widetilde{U} \subseteq V$. Thus $\widetilde{U} \subseteq \widetilde{R}\Big[[w]\Big]$. The definition of $\widetilde{S}_{[w]}$ implies $[x]\,\widetilde{S}_{[w]}\widetilde{U}$.

It remains to verify that $\widetilde{R}\Big[\widetilde{U}\Big] \subseteq C$. Let $t \in U$ and $z \in W$ be such that $[t]\,\widetilde{R}\,[z]$. Then we have $t \in U_{w',y}$ for some $w' \in [w]$ and $y \in [x]$. Since $[t]\,\widetilde{R}\,[z]$, there are $t' \in [t]$ and $z' \in [z]$ with $t'Rz'$. The condition (forth) implies that there is z'' such that tRz'' and $z' \sim z''$. Since $R\Big[U_{w',y}\Big] \subseteq C_y$ and $z'' \in R\Big[U_{w',y}\Big]$, we have $z'' \in C_y$. The definition of C_y implies $[z''] \in C$, or equivalently, $[z] \in C$. $\dashv$

Corollary 5.16 **IL**R is decidable.

Future work

The "natural order of things" regarding the exploration of basic modal properties of interpretability logics is to first obtain completeness result(s), then see if they are decidable, and finally classify them in terms of computational complexity. The queue of logics waiting for their decidability to be determined is currently empty: all interpretability logics known to be complete w.r.t. at least one semantics are also known to be decidable. Taking a step back, the best known candidates for a complete logic, whose completeness is still an open question, are logics whose axioms are among those in two recently introduced series of arithmetically sound principles [29]. Another candidate is the logic **IL**W_ω which we explore in Chapter 8.

Chapter 6

Complexity

In this chapter, which is for the most part self-contained, we prove the PSPACE-completeness of **IL**, **IL**W and **IL**P.

The first part of this chapter (the part that concerns the logic **IL**) is based on the published paper [47].

Introduction

Computational complexity of modal logics was first studied by Ladner [43]. Various tableau-based methods were used in proofs of PSPACE-decidability of a number of modal logics (like **K**, **K4**, **S4** etc; see [43] and [58]).

Chagrov and Rybakov [16] prove the PSPACE-completeness of the closed fragments of modal systems **K** and **K4** (and in fact any logic **L** such that $\mathbf{K} \subseteq \mathbf{L} \subseteq \mathbf{K4}$), while for logics **Grz** and **GL** they establish PSPACE-completeness of their one-variable fragments.

Shapirovsky [55] proved the PSPACE-decidability of propositional polymodal provability logic **GLP**. PSPACE-completeness of the closed fragment of the system **GLP** is proved by Pakhomov in [51].

In this chapter we explore complexity of interpretability logics. Bou and Joosten proved in [6] that the decidability problem for the closed fragment of **IL** is PSPACE-hard.[1] This implies in particular that **IL** is PSPACE-hard. The fact that **IL** is PSPACE-hard also follows from the fact that already **GL** is PSPACE-hard, and **IL** conservatively extends **GL**.

In [63] (the final two chapters) another interesting topic regarding complexity and interpretability is studied: feasible interpretability. In that version of interpretability (which is a type of *axioms interpretability*), the length of a proof of an axiom's translation is polynomially bounded. In this chapter we will study classical theorems interpretability. The modal interpretability logic of feasible (and, of course, classical) interpretability of **PA** is **IL**M (as shown in [63]). Unfortunately, we did not succeed in determining the

[1] Čačić and Vuković [15] proved normal forms exist for a wide class of closed **IL** formulas. Čačić and Kovač [13] quantified asymptotically how wide those classes are asymptotically.

complexity class of **ILM**. See the concluding sections for some comments regarding our attempts.

We first consider the complexity problem for the interpretability logic **IL** and prove that it is PSPACE-complete [47]. Our constructions can be seen as generalisations of the constructions by Boolos presented in [5] (Chapter 10). If we restrict our work to **GL**, the resulting method is very similar to the one given by Boolos, up to the terminology. Our method can also be seen as extending the method presented in [55], of proving PSPACE-completeness (monomodal case), and other similar proofs where one constructs a tree-like structure in a space-efficient manner.

Since this method extends the methods available for **GL**, let us briefly describe the way we can efficiently check **GL**-satisfiability, without going into too much detail. A very brief description is that for **GL** a certain depth-first search through irreflexive transitive trees of depth bounded by the complexity of a formula suffices [64]. We will not present any particular algorithm in full (see e.g. [5] or [55]). Suppose we start with some set of **GL**-formulas $\Delta \subseteq \Gamma$ where Γ is finite and closed under subformulas. Let us suppose that in the language of **GL** there is only one modal operator: $\Diamond$ (this choice makes our description shorter). If Δ is satisfiable, there are some $\mathfrak{M}$ and $w \in \mathfrak{M}$ such that $w \Vdash \Delta$. In general, Δ needn't fully determine the modal theory of the world that satisfies Δ. For example, if $\Delta = \{p \vee q\}$ where p and q are propositional variables, we can take a model satisfying p or a model satisfying q to demonstrate that Δ is satisfiable. So the first step is to determine what other formulas should be satisfied in the world (or one of the worlds) where Δ is satisfied.[2] So, the satisfiability of Δ is reduced to the satisfiability of some set Σ which fully determines the truth of formulas contained in Γ. For example, if $w \Vdash \Delta$, then the set $\{A \in \Gamma : w \Vdash A\}$ is a good candidate for Σ. We can iterate through all the sets Σ such that $\Delta \subseteq \Sigma \subseteq \Gamma$, and check if they are satisfiable. We next describe how to perform this check for a particular set Σ such that $\Delta \subseteq \Sigma \subseteq \Gamma$. Fix some such Σ. Let us first describe what does it mean for some set Q to be propositionally satisfiable. We first describe an operation on the set Q. Assign to every formula X of the form $\Diamond A$ (or $A \rhd B$ in case we are talking about propositional satisfiability in the context of interpretability logics) that appears (either as such, or as a subformula) in Q a fresh propositional letter p_X. Now uniformly substitute all occurrences of formulas X of the form $\Diamond A$ or $A \rhd B$ that appear in Q with the assigned variables p_X.

We say the set Q is *propositionally satisfiable* if the thus obtained set of (propositional) formulas is satisfiable.

First we check if Σ is propositionally satisfiable. Continuing the description of how to check **GL**-satisfiability, if Σ is propositionally satisfiable, we look at all $\Diamond$-formulas

[2]Strictly speaking this is not a necessary step; we could get by with just specifying which propositional variables and formulas of the form $\Diamond B$ should be true, while ensuring the choice is coherent with the contents of Δ.

$\Diamond A_1, \ldots, \Diamond A_k$ in Σ. For each $\Diamond A_i$, we need to check if A_i is satisfiable. For this, we can use our **GL**-satisfiability algorithm again; this time with a set containing A_i. If checks return a positive answer for all i (i.e. for all i there are some $\mathfrak{M}_i$ and $w_i \in \mathfrak{M}_i$ with $w_i \Vdash A_i$), we can make a model for Σ by prepending a fresh world w to the disjoint union of models $\mathfrak{M}_i$. Clearly $w \Vdash \Diamond A_1, \ldots, \Diamond A_k$. In this case we stop our search: we found a satisfiable set $\Sigma \supseteq \Delta$. Otherwise, we try the next set Σ, $\Delta \subseteq \Sigma \subseteq \Gamma$. If we don't find a satisfiable superset Σ of the set Δ, the algorithm determines that Δ is not satisfiable.

There are two important aspects we did not specify in the preceding description of how to check the **GL**-satisfiability of a given set of formulas. First, there is the issue of $\Diamond$-formulas that are not present in Σ. We must ensure these formulas are false in the remainder of the model. A simple way to do this is to parameterise the algorithm with a set of "banned formulas" (which is, just like Δ, a subset of Γ). Returning to the appropriate step of our description, before checking if A_i is satisfiable for $\Diamond A_i \in \Sigma$ we look at the set of all B_j such that $\Diamond B_j \in \Gamma \setminus \Sigma$. Such formulas B_j must not be the satisfied in the model the algorithm builds for A_i (because we later wish to obtain the truth lemma, so the set of the contained formulas must correspond to the set of true formulas). So, we add all such formulas B_j to the set of "banned formulas" when we recursively call the algorithm to check if A_i is satisfiable.

Another aspect we did not cover is termination. How do we ensure the recursion stops? A way to achieve this is to utilise the fact that $\mathbf{GL} \vdash \Diamond A \leftrightarrow \Diamond(A \land \Box\neg A)$. So, instead of trying to build a model for A_i, we try and build a model for A_i (a model $\mathfrak{M}_i$ with $w_i \in \mathfrak{M}_i$ such that $w_i \Vdash A_i$) where for all worlds x except for w_i we have $x \nVdash A_i$. This property can be ensured with another parameter: "delayed banned formulas". To put it shortly, we need to tell the next call which formula within Δ is the formula A_i.

In this chapter we work with ordinary Veltman models. Our methods extend the approach used for **GL**, which we sketched above. Generally speaking, there are two new conceptual issues that have to be dealt with. One is that we have to deal with S_w relations (which did not exist in **GL**-models), so the concept of a model defined by a particular run of an algorithm is more complex than is the case with **GL**. With **GL** it sufficed to let the accessibility relation R equal the execution tree of a successful run, where each node represents a (successful) check if a particular set of formulas is satisfiable. Another but related issue is that the structure of the model can no longer be made dependent solely on the formulas we'd like to satisfy in its root (together with the sets of "banned" and "delayed banned" formulas). That is, in addition to formulas we'd like to be shown (un)satisfiable, there is additional information to carry between recursive calls of our algorithm.

6.1 Preliminaries

Recall that in this thesis, unless stated otherwise, we assume the language of interpretability logics contains only the symbols $\bot$, $\to$, $\rhd$, and countably many symbols for propositional variables. The results of this section can be easily extended to include other symbols such as $\Box$. In particular, we know that $\Box A$ is equivalent to $(A \to \bot) \rhd \bot$, so given a PSPACE algorithm P for the basic language we can construct a PSPACE algorithm P^+ for the extended language: it consists of a linear-space preprocessing stage together with P.

A *rooted Veltman model* $(\mathfrak{M}, w)$ is a pair consisting of a Veltman model $\mathfrak{M} = (W, R, \{S_x : x \in W\})$ and a world w such that all other worlds are R-accessible from w; we say that $(\mathfrak{M}, w)$ is a model of a formula B (a set of formulas Φ) if $\mathfrak{M}, w \Vdash B$ ($\mathfrak{M}, w \Vdash B$, for each $B \in \Phi$).

For a Veltman model $\mathfrak{M}$ and a world x, the *rooted submodel generated by* x is the rooted model $(\mathfrak{N}, x)$, where $\mathfrak{N}$ is the restriction of $\mathfrak{M}$ to the set of all worlds that are either x itself or are R-accessible from x.

Let us denote by $\mathrm{Sub}(A)$ the set of subformulas of a formula A. For a given formula A we define $\Gamma = \mathrm{Sub}(A) \cup \{\bot\}$. We assume that the formula A and the corresponding set Γ are fixed and available in all contexts. In other words, we will assume that these objects are available to our algorithm even though we don't mention them explicitly as input parameters.

In the paper [47] the set Γ also contained negations of subformulas of A. Here we change the approach slightly. We add two additional inputs to our algorithm: $\mathcal{B} \subseteq \Gamma$ which stands for *banned* formulas, and $\mathcal{D} \subseteq \Gamma$ which stands for *delayed banned* formulas. The original algorithm ([47]) solved the problem of satisfiability, i.e. "given some set $\Delta \subseteq \Gamma$, is there a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \Delta$?" The new algorithm presented here solves a slightly more general problem: given some sets $\Delta, \mathcal{B}, \mathcal{D} \subseteq \Gamma$, is there a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \Delta$, such that $\dot{R}[w] \Vdash \neg\mathcal{B}$, and $R[w] \Vdash \neg\mathcal{D}$. Clearly, the difference is only superficial since the old algorithm can answer this question too, given $\Delta \cup \neg\mathcal{B} \cup \Box\neg\mathcal{B} \cup \Box\neg\mathcal{D}$ as input.

Because of these newly added inputs (compared to **IL**), we do not need additional negated formulas in the set Γ. We believe this approach is more elegant than the one used in our published paper.

Given $S \subseteq \Gamma$, we define $\mathrm{Full}(S) := S \cup \{\neg B : B \in \Gamma \setminus S\}$. We will say that $S \subseteq \Gamma$ is *Boolean satisfiable* (w.r.t. Γ) if $\mathrm{Full}(S)$ is a propositionally satisfiable set (see the introduction for a precise description of *propositional satisfiability*). In general, it can happen that $\mathrm{Full}(S) \nsubseteq \Gamma$, but this won't be an issue. Let us briefly describe the purpose of $\mathrm{Full}(S)$. When the algorithm attempts the construction of a world satisfying Δ, it first has to decide the truth values of formulas not determined by Δ. A way to implement this choice is to iterate through all subsets S of Γ, and try to make exactly the formulas of S satisfied. This implies we want all B for $B \in \Gamma \setminus S$ to be falsified.

6.2 The logic IL

We will present a PSPACE-algorithm that given a modal formula A checks whether there is a rooted Veltman model $(\mathfrak{M}, w)$ of A (and additionally, given parameters $\mathcal{B}$ and $\mathcal{D}$, the model has to satisfy $\dot{R}[w] \Vdash \neg\mathcal{B}$, and $R[w] \Vdash \neg\mathcal{D}$). In order to prove that **IL** is in PSPACE, given a modal formula A we will apply this algorithm to the set $\{A\}$, with the sets $\mathcal{B} = \emptyset$ and $\mathcal{D} = \{A\}$ (alternatively, we can take both sets $\mathcal{B}$ and $\mathcal{D}$ to be empty, which might result in a slightly larger model). The execution with these parameters will take a polynomial amount of space in $|A|$, thus demonstrating that **IL** is in PSPACE.

We will present our algorithm as the main Algorithm (1) and supplementary algorithms (2) and (3) that can make recursive calls of each other and return either a positive or a negative answer ((1) makes only calls of (2), (2) makes only calls of (3), and (3) makes only calls of (1)). First we will give a full description of the computation process (the algorithms (1), (2) and (3)) and specify what we are computing, but we will prove our claims about what we are computing only later.

First we pick the formula A whose satisfiability we're testing.

Note 6.1 In the remainder of the section we assume that the formula A and the corresponding set $\Gamma = \mathrm{Sub}(A) \cup \{\bot\}$ are fixed. Thus, all our statements and algorithms are implicitly parametrised with A.

Algorithm (1) takes sets $\Delta, \mathcal{B}, \mathcal{D} \subseteq \Gamma$ as its input. Algorithm (2) takes sets $\Delta, \mathcal{D} \subseteq \Gamma$ as its input. Algorithm (3) takes as input a formula $C \rhd D \in \Gamma$, a set $\Delta \subseteq \Gamma \setminus \{C \rhd D\}$ of formulas of the form $E \rhd G$; and a set $\mathcal{D} \subseteq \Gamma$. All three algorithms return a single value which is either a positive or a negative answer (i.e. `yes` or `no`). When we need to refer to an input parameter X of an algorithm (i) we will sometimes write $X^{(i)}$ if "X" is otherwise ambiguous in the given context.

Algorithm (1) computes whether there is a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \Delta$, such that $\dot{R}[w] \Vdash \neg\mathcal{B}$ and $R[w] \Vdash \neg\mathcal{D}$. It does this by enumerating all (if any) Boolean satisfiable extensions $\Delta' \supseteq \Delta$ such that $\mathrm{Full}(\Delta') \cup \neg\mathcal{B}$ is propositionally satisfiable. Algorithm (1) returns a positive answer if and only if for at least one such extension Δ', Algorithm (2) returns a positive answer with $\Delta^{(2)} = \Delta'$, and $\mathcal{D}^{(2)} = \mathcal{B} \cup \mathcal{D}$.

Algorithm (2) takes sets $\Delta, \mathcal{D} \subseteq \Gamma$ as input, where Δ is assumed to be Boolean satisfiable. Algorithm (2) computes whether there is a rooted model $(\mathfrak{M}, w)$ such that

$w \Vdash \mathrm{Full}(\Delta)$ and $R[w] \Vdash \neg\mathcal{D}$. This is accomplished by examining the sets Δ^+ and Δ^-:

$$\Delta^+ = \{C \rhd D \in \Gamma : C \rhd D \in \Delta\}; \tag{6.1}$$

$$\Delta^- = \{C \rhd D \in \Gamma : C \rhd D \notin \Delta\}. \tag{6.2}$$

For each formula $C \rhd D \in \Delta^-$, we check whether there is a rooted Veltman model $(\mathfrak{M}_{C\rhd D}, w_{C\rhd D})$ of $\{\neg(C \rhd D)\} \cup \Delta^+$, such that $R[w_{C\rhd D}] \Vdash \neg\mathcal{D}$. This is done through the call of Algorithm (3) which is parametrised with $(C \rhd D)^{(3)} = C \rhd D$, $\Delta^{(3)} = \Delta^+$, and $\mathcal{D}^{(3)} = \mathcal{D}$. Algorithm (2) returns a positive answer if and only if all these checks are positive.

Let us give a brief informal description of what Algorithm (3) computes. Suppose we are constructing a rooted model $(\mathfrak{M}, w)$ of Δ where $C \rhd D \in \Delta^-$. Algorithm (3) is trying to build the part of $R[w]$ that should witness the falsity of $C \rhd D$, i.e. there should be a world satisfying C without an S_w transition to a world satisfying D. Let us denote the aforementioned part of $R[w]$ (that witnesses the falsity of $C \rhd D$) as U. Algorithm (3) first chooses which formulas in Γ will be satisfied somewhere in U. We can denote the set of such formulas as P ("positive") and let $N = \Gamma \setminus P$. The formulas in the set N are the ones we want to be false everywhere in U. Later we will make U closed under S_w. So in particular, if $E \rhd G$ is to be true in w, either we have a world satisfying G somewhere in U, or E must be false everywhere in U (if E was true somewhere in U, we would need G true somewhere in U).

Let us resume with the formal description of Algorithm (3). First we need an auxiliary notion.

We will say that a pair of sets of formulas (N, P) is a $(\Delta, C \rhd D)$-pair if:

- $N, P \subseteq \Gamma$;
- $C \in P$, $D \in N$, $\bot \notin P$;
- for each $E \rhd G \in \Delta$, either $E \in N$ or $G \in P$.

Algorithm (3) takes a single formula $C \rhd D \in \Gamma$ as input, a set $\Delta \subseteq \Gamma \setminus \{C \rhd D\}$ of formulas of the form $E \rhd G$; and a set of formulas $\mathcal{D} \subseteq \Gamma$. We return a positive answer if there is a $(\Delta, C \rhd D)$-pair (N, P) such that for every $G \in P$ there is a rooted Veltman model $(\mathfrak{M}_G, w_G)$ of G such that:

1. $\dot{R}[w_G] \Vdash \neg\mathcal{D}, \neg N$;
2. $R[w_G] \Vdash \neg G$.

Note that these checks can be computed with Algorithm (1) if we provide the following input values: $\Delta^{(1)} = \{G\}$, $\mathcal{B}^{(1)} = \mathcal{D} \cup N$, and $\mathcal{D}^{(1)} = \{G\}$.

We will now proceed to prove that this algorithm has the required properties. To do so, let us first verify that algorithms (1), (2) and (3) do what we described.

Lemma 6.2 Given $\Delta, \mathcal{B}, \mathcal{D} \subseteq \Gamma$, the following statements are equivalent:

1. there exists a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \Delta$, such that $\dot{R}[w] \Vdash \neg\mathcal{B}$ and $R[w] \Vdash \neg\mathcal{D}$;

2. there is a set $\Delta' \subseteq \Gamma$ and a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \text{Full}(\Delta')$, such that $R[w] \Vdash \neg\mathcal{B}, \neg\mathcal{D}$, the set Δ' is a Boolean satisfiable extension of Δ, and $\text{Full}(\Delta') \cup \neg\mathcal{B}$ is propositionally satisfiable.

Proof. For (1.) to (2.), note that if there is a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \Delta$, such that $\dot{R}[w] \Vdash \neg\mathcal{B}$ and $R[w] \Vdash \neg\mathcal{D}$, the same structure is a model of $\text{Full}(\Delta')$ where $\Delta' = \{A \in \Gamma : w \Vdash A\}$. This extension is satisfiable, and in particular propositionally satisfiable, with $\neg\mathcal{B}$. Conditions involving $R[w]$ are clearly preserved.

For the other direction, we reuse the model whose existence is known. The only non-obvious property to check is whether $w \Vdash \neg\mathcal{B}$. Suppose $w \Vdash B$ and $\neg B \in \neg\mathcal{B}$ for some formula $B \in \Gamma$. Since Δ' is Boolean satisfiable, $B \in \Delta'$ (otherwise $\neg B \in \text{Full}(\Delta')$ and then $w \Vdash \neg B$). Since $w \Vdash \Delta'$, it follows that $w \Vdash \neg B$, a contradiction. $\dashv$

Lemma 6.3 Let Δ be a Boolean satisfiable subset of Γ, and $\mathcal{B}, \mathcal{D} \subseteq \Gamma$. The sets Δ^+ and Δ^- are given by (6.1) and (6.2). The following are equivalent:

1. there exists a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \text{Full}(\Delta)$ and $R[w] \Vdash \neg\mathcal{D}$;

2. for all $C \rhd D \in \Delta^-$, there is a rooted Veltman model $(\mathfrak{M}_{C\rhd D}, w_{C\rhd D})$ of $\{\neg(C \rhd D)\} \cup \Delta^+$, such that $R[w_{C\rhd D}] \Vdash \neg\mathcal{D}$.

Proof. First assume that there is a rooted Veltman model $(\mathfrak{M}, w)$ of $\text{Full}(\Delta)$. It is easy to see that we can put $(\mathfrak{M}_{C\rhd D}, w_{C\rhd D}) = (\mathfrak{M}, w)$, for each $C \rhd D \in \Delta^-$.

In the other direction, suppose we have rooted Veltman models $(\mathfrak{M}_{C\rhd D}, w_{C\rhd D})$ with the described properties for each $C \rhd D \in \Delta^-$. Denote $(W_{C\rhd D}, R_{C\rhd D}, \{S_x^{C\rhd D} : x \in W_{C\rhd D}\}, \Vdash_{C\rhd D}) = \mathfrak{M}_{C\rhd D}$. In order to construct a rooted Veltman model $(\mathfrak{M}, w)$ of Δ, we take the disjoint union of the models $\mathfrak{M}_{C\rhd D}$, and then "merge" the worlds $w_{C\rhd D}$ in one world w. More formally, we suppose the models $\mathfrak{M}_{C\rhd D}$ are disjoint, the world w is fresh and then define:

$$W = \{w\} \cup \bigcup_{C\rhd D \in \Delta^-} W_{C\rhd D} \setminus \{w_{C\rhd D}\};$$

$$R = \{(w, x) \in W^2 : x \neq w\} \cup \bigcup_{C\rhd D \in \Delta^-} R_{C\rhd D} \cap W^2;$$

$$S_w = \bigcup_{C\rhd D \in \Delta^-} S_{w_{C\rhd D}};$$

for $x \in W_{C\rhd D} \setminus \{w_{C\rhd D}\},\ S_x = S_x^{C\rhd D}$.

In the last line, the superscript $C \rhd D$ ($S_x^{C \rhd D}$) is added since otherwise it would be ambiguous if we are referring to one of the initial models $\mathfrak{M}_{C \rhd D}$ or the final model $\mathfrak{M}$. We omit this superscript if there is no danger of confusion (e.g. the world $w_{C \rhd D}$ exists only in $\mathfrak{M}_{C \rhd D}$, so we will write $S_{w_{C \rhd D}}$ instead of $S^{C \rhd D}_{w_{C \rhd D}}$.) We put satisfaction of proposition variables in w according to the set Full(Δ).

It is easy to prove by induction on the complexity of a formula $B \in \Gamma$ that we have the following: $\mathfrak{M}, w \Vdash B$ if and only if $B \in \Delta$. Let us consider the case $B = C \rhd D$. Suppose $\mathfrak{M}, w \Vdash C \rhd D$. Assume $C \rhd D \notin \Delta$ for a contradiction. Clearly $C \rhd D \in \Delta^-$. By assumption, $C \rhd D$ fails in $\mathfrak{M}_{C \rhd D}$, so there is some x, such that $w_{C \rhd D} R_{C \rhd D} x$, with $\mathfrak{M}_{C \rhd D}, x \Vdash C$ and for no y such that $x S_{w_{C \rhd D}} y$ do we have $\mathfrak{M}_{C \rhd D}, y \Vdash D$. By the construction of the model $\mathfrak{M}$ we have $\mathfrak{M}, x \Vdash C$ and for no y such that $x S_w y$ do we have $\mathfrak{M}, y \Vdash D$. Since wRx and the set $S_w[x]$ equals $S_{w_{C \rhd D}}[x]$, this contradicts the assumption that $\mathfrak{M}, w \Vdash C \rhd D$.

In the other direction, suppose $C \rhd D \in \Delta$ and wRx. Since $C \rhd D \in \Delta$, we have $C \rhd D \in \Delta^+$. Since wRx, by construction, x is in $\mathfrak{M}_{E \rhd G}$ for exactly one formula $E \rhd G$. Since $\mathfrak{M}_{E \rhd G}$ is a model of Δ^+, there is y such that $x S_{w_{E \rhd G}} y$ and $\mathfrak{M}_{E \rhd G}, y \Vdash D$. But y is included in $\mathfrak{M}$, and we have $x S_w y \Vdash D$. Thus $w \Vdash C \rhd D$. $\dashv$

Lemma 6.4 Let $\Delta, \mathcal{D} \subseteq \Gamma$ where Δ is a set of formulas of the form $E \rhd G$, and assume $C \rhd D \in \Gamma \setminus \Delta$. The following are equivalent:

1. there exists a rooted model $(\mathfrak{M}, w)$ of $\{\neg(C \rhd D)\} \cup \Delta$, such that $R[w] \Vdash \neg\mathcal{D}$;

2. there is a $(\Delta, C \rhd D)$-pair (N, P) such that for every $G \in P$ there is a rooted Veltman model $(\mathfrak{M}_G, w_G)$ of G such that:

   (a) $\dot{R}[w_G] \Vdash \neg\mathcal{D}, \neg N$;

   (b) $R[w_G] \Vdash \neg G$.

Proof. (1.) to (2.) Fix the model $(\mathfrak{M}, w)$ of $\{\neg(C \rhd D)\} \cup \Delta$ such that $R[w] \Vdash \neg\mathcal{D}$, and we are to find the required $(\Delta, C \rhd D)$-pair (N, P). Since $w \nVdash C \rhd D$, there is a world $w_{C \rhd D}$ with $w R w_{C \rhd D} \Vdash C$ and for all x, if $w_{C \rhd D} S_w x$, then $x \nVdash D$. W.l.o.g. we can assume that $w_{C \rhd D}$ is an R-maximal[3] world with this property. In particular, if $w_{C \rhd D} R x$, then $x \nVdash C$ (otherwise, since $w R w_{C \rhd D}$ implies $S_w[x] \subseteq S_w[w_{C \rhd D}]$, the world $w_{C \rhd D}$ would not be R-maximal in the set in question). Let $P = \{G \in \Gamma : [G]_w \cap S_w[w_{C \rhd D}] \neq \emptyset\}$ and $N = \Gamma \setminus P$.

For every $G \in P$ let $w_{C \rhd D, G}$ denote any world that is R-maximal in the set $[G]_w \cap S_w[w_{C \rhd D}]$. Put $(\mathfrak{M}_G, w_G) := (\mathfrak{M}[w_{C \rhd D, G}], w_{C \rhd D, G})$ where $\mathfrak{M}[w_{C \rhd D, G}]$ is the submodel generated by $w_{C \rhd D, G}$.

[3]A world x is R-maximal in a set S if x is in S and no R-successor of x is in S.

It is easy to check we have Properties (2a) and (2b). For instance, let us first verify $\dot{R}[w_{C \rhd D,G}] \Vdash \neg N$. Assume otherwise, i.e. for some $E \in N$ and $x \in \dot{R}[w_{C \rhd D,G}]$ we have $x \Vdash E$. Since $E \in N$, we have $[E]_w \cap S_w[w_{C \rhd D}] = \emptyset$. Since $x \in \dot{R}[w_{C \rhd D,G}]$, certainly $x \in S_w[w_{C \rhd D,G}] \subseteq S_w[w_{C \rhd D}]$, a contradiction.

In the other direction we know there is a $(\Delta, C \rhd D)$-pair (N, P) such that for every $G \in P$ there is a rooted Veltman model $(\mathfrak{M}_G, w_G)$ of G such that (2a) and (2b) hold. We construct $\mathfrak{M}$ as follows. First take the disjoint union U of the models $(\mathfrak{M}_G, w_G)$ for $G \in P$. Prepend a new world w as an R-predecessor of every world in U. Make all the worlds other than w pairwise S_w-accessible. The forcing relation of the new model should inherit the ones from the building blocks, and for w it can be chosen arbitrarily.

It remains to prove that $w \Vdash \neg(C \rhd D), \Delta$ and $R[w] \Vdash \neg\mathcal{D}$. It is easy to see that $R[w] \Vdash \neg\mathcal{D}$. To see that $w \Vdash \neg(C \rhd D)$, note that $wRw_C \Vdash C$ (since $C \in P$) and that, since $D \in N$, all the models $\mathfrak{M}_G$ validate $\neg D$. Finally, let us check if $w \Vdash \Delta$. Let $E \rhd G \in \Delta$ be an arbitrary formula, and $x \in R[w]$ an arbitrary world such that $x \Vdash E$. Since $x \Vdash E$, we have $G \in P$ (otherwise $E \in N$, contradicting $wRx \Vdash E$). Thus, the world w_G exists and by the definition of S_w, we have xS_ww_G.

$\dashv$

Recall that we fixed a formula A which we would wish to decide if it has an **IL**-model or not. We also fixed the corresponding set $\Gamma = \mathrm{Sub}(A) \cup \{\bot\}$ (see Note 6.1).

Theorem 6.5 The logic **IL** is PSPACE-decidable.

Proof. First we show that the recursion depth is bounded by $|\Gamma| + 1$. Suppose that for some initial input parameter $\Delta \subseteq \Gamma$ (with $\mathcal{B} = \mathcal{D} = \emptyset$) we have a chain of calls

$$c_0^{(1)}, c_0^{(2)}, c_0^{(3)}, c_1^{(1)}, c_1^{(2)}, c_1^{(3)}, \dots, c_n^{(1)}, c_n^{(2)}, c_n^{(3)}, c_{n+1}^{(1)}, c_{n+1}^{(2)}, c_{n+1}^{(3)}$$

of the Algorithms (1), (2) and (3) where $n = |\Gamma|$. It is possible that $c_{n+1}^{(3)}$ makes further calls of Algorithm (1), at the moment we are not assuming anything with regard to that. Let G_i for $0 \leq i \leq n$ be the formula from the set P in $c_i^{(3)}$, for which $c_i^{(3)}$ made the call $c_{i+1}^{(1)}$.

Let the index $i \leq n-1$ be arbitrary. The sets $\mathcal{D}$ in $c_{i+1}^{(1)}$, $c_{i+1}^{(2)}$ and $c_{i+1}^{(3)}$ contain G_i, so the set Δ in $c_{i+2}^{(2)}$ cannot contain G_i. Since we never remove elements from the sets $\mathcal{D}$ in subsequent calls, for $j \geq i+2$ we have that Δ in $c_j^{(2)}$ cannot contain G_i. Since Δ in $c_j^{(2)}$ is a superset of Δ in $c_j^{(1)}$, and the latter contains G_{j-1}, we have $G_j \neq G_i$ for $j \geq i+1$. Thus, the set $\{G_0, G_1, \dots, G_n\} \subseteq \Gamma$ contains $n+1$ different elements, which is impossible since $|\Gamma| = n$. This contradicts the existence of our chain, which is of length $n+2$.

Thus, our algorithm terminates. Next we see that the space reserved by our algorithm is at most polynomial in $|A|$ in any given instant. To see this, we first note that the recursion depth has an upper bound which is linear in $|A|$. At any given moment we keep only one branch in memory. So, it remains to convince ourselves that each particular execution of Algorithms (1), (2) and (3) takes at most a polynomial amount of space.

Algorithm (1) iterates through all Σ such that $\Delta \subseteq \Sigma \subseteq \Gamma$. This iteration requires n bits, and thus is linear in $|A|$. For each Σ we check if it is Boolean satisfiable (which is a PTIME preprocessing stage followed by an NP propositional satisfiability check). Finally, Algorithm (1) makes a call of Algorithm (2).

Algorithm (2) is computationally very simple; it makes a small number of calls of Algorithm (3). The number of calls made is bounded by $|\Gamma|$, but since we are not making these (or any other) calls in parallel, even this detail is irrelevant.

Algorithm (3) iterates through all candidates (N, P) for a $(\Delta, C \rhd D)$-pair. Thus, it suffices to iterate through pairs of subsets of Γ (this requires $2n$ bits), and then perform a simple PTIME check of all the conditions ($C \in P$, $D \in N$, $\bot \notin P$, for all $E \rhd G \in \Delta$ either $E \in N$ or $G \in P$). If some pair (N, P) passes these checks, we call Algorithm (1) for this pair. If one of the checks failed or Algorithm (1) returned a negative answer, we proceed to the next candidate pair.

We see that a particular execution of Algorithms (1), (2) and (3) takes a polynomial (in fact, linear) amount of space. So the whole Algorithm (1) (together with the recursive calls) is at most quadratic in $|A|$, and thus it belongs to PSPACE.

Now the only thing left to verify is that the descriptions of what we are computing in our algorithm are indeed correct. Since we already know that our computation terminates, it is enough to show that our algorithm works locally correct, i.e. that assuming that further calls do what we describe that they are doing, the call under consideration also computes what we want it to compute. Formally, we prove the correctness of descriptions by induction on the depth of the recursion (the base case are terminal calls, i.e. leaf calls in the execution tree).

The fact that the description of Algorithm (1) is correct follows from Lemma 6.2, that the description of Algorithm (2) is correct follows from Lemma 6.3, and that the description of Algorithm (3) is correct follows from Lemma 6.4.

Thus, there is a model of A if and only if Algorithm (1) returns a positive answer given the input $\Delta = \{A\}$ and $\mathcal{B} = \mathcal{D} = \emptyset$. $\dashv$

Bou and Joosten [6] proved that the decidability problem for the closed fragment of **IL** is PSPACE-hard. Together with the previous theorem, this implies the following.

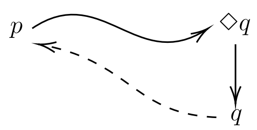
Corollary 6.6 The validity and decidability problems of the logic **IL** are PSPACE-complete.

6.3 The logic ILW

Let us now extend results of the previous section to the logic **IL**W. The characteristic property, (W), is the following:

for all $w \in W$, the relation $S_w \circ R$ is converse well-founded.

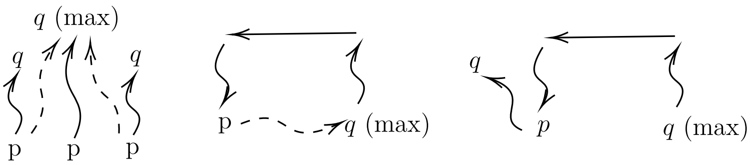
Assume we wish to know if the following set is **IL**W-satisfiable: $\{\Diamond p, p \rhd \Diamond q \vee \Diamond r, q \rhd p\}$. The algorithm for **IL** would have to produce a witness for, among other formulas, the formula $\Diamond q \vee \Diamond r$. To do this it has to iterate through propositionally satisfiable sets containing $\Diamond q \vee \Diamond r$. Disregarding the irrelevant formulas for the moment, there are two options to go with: either $\Diamond q$ or $\Diamond r$. Which option is picked depends on the order the algorithm uses for iteration. If $\Diamond q$ comes first, it will be picked and the final model might look like this (all worlds are assumed to be R-accessible from w):



And this is clearly not a good choice: we have an infinite $(R \circ S_w)$-loop, so this model is not an **IL**W-model. Thus, we have to build a different algorithm that will prevent $(R \circ S_w)$-loops. However, we have to allow certain non-problematic and non-trivial S_w-loops. Assume we want to satisfy $\Diamond p$, $p \rhd q$ and $q \rhd p$. We might build a model like this one:

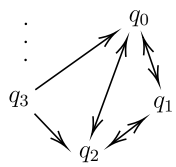
Unlike the previous model, this model is fine, and might even be the only solution (e.g. if **IL**W $\vdash \neg(p \wedge q)$). So, preventing $(R \circ S_w)$-loops cannot be reduced to simply preventing S_w-loops; we really have to take the relation R into account. This time we can't make S_w total as we did for **IL** (see Lemma 6.4).

We first show (Lemma 6.7) that **IL**W-satisfiable formulas have a particular kind of uniform models. For example, we can assume that if $p \rhd q \in \Delta^+$ and we have a number of R-successors of w that satisfy p, then there is a particular R-successor of w that satisfies q which is both S_w-accessible from all the aforementioned worlds that satisfy p, and also $(S_w \circ R \circ S_w)$-maximal in the set of worlds satisfying q. This is shown in the leftmost picture below. The other two images sketch the main part of the proof of Lemma 6.7. This is the part where we show that any model can be transformed to a model with the desired properties, but that we do not lose the characteristic property (W) in the process. It boils down to the fact that if there is a loop introduced in the process (caused by the addition of an S_w-transition, pictured dashed), then the world we selected as supposedly $(S_w \circ R \circ S_w)$-maximal actually had an $(S_w \circ R \circ S_w)$-successor (see the proof of Lemma 6.7).



Once the Lemma 6.7 convinces us that there are uniformly-structured models for all **IL**W-satisfiable formulas, we can limit our search space to that class of models.

The main modification of our algorithm will be the part that for each $C \rhd D \in \Delta^-$ iterates through (S_w-) "visibility" graphs of witnesses of formulas from the "P" component of an $(\Delta, C \rhd D)$-pair (N, P). This graph tells us which witness is allowed to S_w-access which other witness. With this information, we will avoid introducing loops, since every node will be aware of formulas that would cause a loop to appear. For example, if $P = \{q_0, q_1, q_2, q_3, \dots\}$:



On this graph the existence of an arrow $p \to q$ means that the witness of p can S_w-access the witness of q. The non-existence, however, means that the witness of p can't S_w-access *any* world satisfying q (not only the one particular chosen witness).

Let A be the formula whose satisfiability we are interested in, and $\Gamma := \mathrm{Sub}(A) \cup \{\bot\}$. As in the previous section, A and Γ are fixed throughout the remainder of this section (see Note 6.1). In this section, "model" will always mean "**IL**W-model".

Given a set of formulas Γ, a model $\mathfrak{M}$ and a world $x \in \mathfrak{M}$, denote

$$I_x = \{C \rhd D \in \Gamma : x \Vdash C \rhd D\};$$
$$N_x = \{C \rhd D \in \Gamma : x \nVdash C \rhd D\}.$$

These sets, of course, depend not only on x but also on Γ, however the notation doesn't explicate this since the choice of Γ is fixed ($\Gamma = \mathrm{Sub}(A) \cup \{\bot\}$), as was the case with **IL**.

Lemma 6.7 If $(\mathfrak{M}, x)$ is a rooted **IL**W-model, then there is another rooted **IL**W-model $(\mathfrak{N}, y)$ such that:

1. $(\mathfrak{M}, x) \equiv_\Gamma (\mathfrak{N}, y)$ and $\{B \in \Gamma : R[x] \Vdash B\} = \{B \in \Gamma : R[y] \Vdash B\}$;

2. for all $C \rhd D \in N_x$ there is a set $P_{C \rhd D} \subseteq \Gamma$ containing C such that for every $G \in P_{C \rhd D}$ there is a world $y_{C \rhd D, G}$, and:

   (a) $P_{C \rhd D} = \{B \in \Gamma : [B]_y \cap S_y[y_{C \rhd D, C}] \neq \emptyset\}$;

(b) $y_{C\rhd D,G} \Vdash G$;

(c) $y_{C\rhd D,G}$ is $(S_y \circ R \circ S_y)$-maximal in the set $[G]_y$;

(d) whenever for some $H \in P_{C\rhd D}$ and E such that $E \rhd G \in I_x$ we have $y_{C\rhd D,H} S_y z \Vdash E$, we also have $z S_y y_{C\rhd D,G}$;

(e) $S_y[y_{C\rhd D,G}] \Vdash \neg D$.

Sometimes we will refer to the worlds $y_{C\rhd D,C}$ as $y_{C\rhd D}$.

Proof. Note that the portions of $R[x]$ witnessing falsity of different $C \rhd D \in N_x$ in the rooted model $(\mathfrak{M}, x)$ can be assumed to be mutually disjoint w.r.t. the relation S. More precisely, if $C_i \rhd D_i \in N_x$, we can assume that there are corresponding worlds x_i such that $x_i \Vdash C_i$, such that $S_x[x_i] \Vdash \neg D_i$ and the sets $S_z[x_i] \cup S_z^{-1}[x_i]$ and $S_z[x_j] \cup S_z^{-1}[x_j]$ are mutually disjoint (for all $z \in \mathfrak{M}$) when $i \neq j$. A way to achieve this is to replicate $R[x]$ in $|N_x|$ copies and pick each x_i from a different copy of the original $R[x]$.

We will construct $\mathfrak{N}$ by modifying $\mathfrak{M}$. Using the assumption of disjointness that was just described, and the fact that our modifications will happen completely inside the sets $S_x[x_i]$, we only need to show that we can achieve the statement of this lemma for a fixed $C \rhd D \in N_x$.

So, fix $C \rhd D \in N_x$. Since $x \nVdash C \rhd D$, the set $V := [C]_x \setminus S_x^{-1}[[D]_x]$ is non-empty. At least one $u \in V$ is $(S_x \circ R \circ S_x)$-maximal in V, otherwise we have an $(R \circ S_x)$-loop[4] infinitely revisiting V. Pick any such $(S_x \circ R \circ S_x)$-maximal $u \in V$ and let $x_{C\rhd D} := u$. We claim that $x_{C\rhd D}$ is also $(S_x \circ R \circ S_x)$-maximal in $[C]_x$. Assume $x_{C\rhd D}(S_x \circ R \circ S_x)v \in [C]_x$ (thus also $x_{C\rhd D} S_x v$). Suppose that $v \notin V$, i.e., $v S_x z$ for some $z \Vdash D$. By the transitivity of S_x we have $x_{C\rhd D} S_x z$ which implies $x_{C\rhd D} \notin V$, a contradiction. Thus, $v \in V$. But this contradicts $(S_x \circ R \circ S_x)$-maximality of $x_{C\rhd D}$ w.r.t. V.

Let $P_{C\rhd D} = \{B \in \Gamma : [B]_x \cap S_x[x_{C\rhd D}] \neq \emptyset\}$ and $x_{C\rhd D,C} := x_{C\rhd D}$. We will now define $x_{C\rhd D,G}$ for all other $G \in P_{C\rhd D}$ ($G = C$ has just been dealt with). First put $V := S_x[x_{C\rhd D}] \cap [G]_x$. Clearly this set is not empty. As before, pick $u \in V$ that is $(S_x \circ R \circ S_x)$-maximal in V and let $x_{C\rhd D,G} := u$. Since $(S_x \circ R \circ S_x)[x_{C\rhd D,G}] \subseteq S_x[x_{C\rhd D,G}] \subseteq S_x[x_{C\rhd D}]$, clearly $x_{C\rhd D,G}$ is also $(S_x \circ R \circ S_x)$-maximal in $[G]_x$.

Thus, if we take $y := x$, we have properties (2a), (2b), and (2c), and it is easy to see we have (2e) too. We would now like to ensure the property (2d): whenever there is $H \in P_{C\rhd D}$ such that $x_{C\rhd D,H} S_x z \Vdash E$ with some $E \rhd G \in I_x$, we then have $z S_x x_{C\rhd D,G}$. We will do this by iterating through the set $Z = \{z \in \bigcup_{E\rhd G \in I_x} [E]_x : x_{C\rhd D} S_x z\}$ and including $(z, x_{C\rhd D,G})$ in S_x. The set Z, and in fact the set $S_x[x_{C\rhd D}]$ too, clearly remains constant in this process. This process may invalidate the transitivity of S_x and also the property

[4]By an "$(R \circ S_x)$-loop" we really mean an infinite $(R \circ S_x)$-chain. Unless stated otherwise, we do not assume that there is any repetition going on, despite the name. However, the results of this section could be formulated in terms of finite models (for example, by replacing every occurrence of "model" with "finite model"), in which case the loop would, of course, imply there is some repetition.

(2c). Furthermore, it is not obvious that it preserves converse well-foundedness of $R \circ S_x$. Let us first show the converse well-foundedness of $R \circ (S_x)^+$.[5] After that, we will close S_x under transitivity, and converse well-foundedness of $R \circ S_x$ will be immediate.

Suppose for a contradiction that we obtain an $(R \circ (S_x)^+)$-loop at some point during the iterative process. We had converse well-foundedness before inserting some pair, denote it as $(z, x_{C \rhd D,G})$, in S_x. Thus, the loop we obtained must contain occurrences of $zS_x x_{C \rhd D,G}$ between infinitely many occurrences of R-transitions. We will first show that a loop of a specific form must exist.

In the aforementioned $(R \circ (S_x)^+)$-loop, take the last occurrence of $zS_x x_{C \rhd D,G}$ that appears before some arbitrarily chosen R-transition bRa. The relation S_x is transitive on the segment of S_x-transitions between $x_{C \rhd D,G}$ and b, so $x_{C \rhd D,G} S_x b$. Similarly, take the first occurrence of $zS_x x_{C \rhd D,G}$ after bRa. Since transitivity holds on the segment of S_x-transitions between a and z, we have $aS_x z$. Thus, for some $a, b \in S_x[x_{C \rhd D}]$, a loop of this form must exist:

$$a \ S_x \ z \ S_x \ x_{C \rhd D,G} \ S_x \ b \ R \ a.$$

Now, since

$$x_{C \rhd D,G} \ S_x \ b \ R \ a \ S_x \ z \Vdash E,$$

for some E such that $E \rhd G \in I_x$, we have in fact

$$x_{C \rhd D,G} \ S_x \ b \ R \ a \ S_x \ z \ S_x \ u \Vdash G,$$

for some $u \in [G]_x$ that was present in the model before extending S_x with $(z, x_{C \rhd D,G})$. This contradicts $(S_x \circ R \circ S_x)$-maximality of $x_{C \rhd D,G}$ w.r.t. $[G]_x$. Thus, there are no $(R \circ (S_x)^+)$-loops. It remains to close S_x under transitivity.

Worlds of the form $x_{C \rhd D,G}$ may lose their $(S_x \circ R \circ S_x)$-maximality (property (2c)) in this iterative process once we start adding S_x-transitions for another formula $G' \in P_{C \rhd D}$. This can be remedied easily. Let $\overline{x_{C \rhd D,G}}$ denote the world currently denoted by $x_{C \rhd D,G}$, and now redefine $x_{C \rhd D,G}$ to be the $(S_x \circ R \circ S_x)$-maximal world in the set $\{z : \overline{x_{C \rhd D,G}} S_x z \Vdash G\}$.

Since $\overline{x_{C \rhd D,G}}$ aggregates paths from all worlds requiring an S_x-successor satisfying G, and $\overline{x_{C \rhd D,G}} S_x x_{C \rhd D,G}$, we have all the old properties, with no need to extend S_x. Since we did not change the model itself, we do not have to check if the properties (2a)–(2e) are preserved.

We let the model $\mathfrak{N}$ be the model thus obtained, and $y := x$. Finally, we should check $(\mathfrak{M}, x) \equiv_\Gamma (\mathfrak{N}, y)$ and $\{B \in \Gamma : R[x] \Vdash B\} = \{B \in \Gamma : R[y] \Vdash B\}$, where $(\mathfrak{M}, x)$ now stands for the state of the model before our transformations took place. First note that the second claim holds trivially, since our transformation preserves R and S_z for $z \neq x$. So it suffices to verify that x and y agree on propositional variables $p \in \Gamma$ (which by

[5]Given a binary relation Q, we denote the transitive closure of Q with Q^+.

definition they do) and formulas of the form $C \rhd D \in \Gamma$. Since we extend S_x in a way that does not enlarge $S_x[x_{C \rhd D}]$, clearly if $x \nVdash C \rhd D$, also $y \nVdash C \rhd D$. If $x \Vdash C \rhd D$, since we do not erase any S_x-transition, also $y \Vdash C \rhd D$. ⊣

Let us now present the algorithm. We use an approach similar to the one for the logic **IL**. In particular, we have three algorithms, which we call Algorithms (1), (2) and (3), again.

We need one more ingredient for **IL**W. We will say that a binary relation $\mathcal{G}$ is a *visibility graph* for a $(\Delta, C \rhd D)$-pair (N, P) if $\mathcal{G} \subseteq P^2$ is transitive and reflexive on P. Let $\overline{\mathcal{G}} := \mathcal{G} \cap \mathcal{G}^{-1}$. Note that due to the reflexivity, every $G \in P$ is contained in the equivalence class $[G] \in P/\overline{\mathcal{G}}$.

Algorithms (1) and (2) are defined as before, so we skip their definitions (replacing each implicit or explicit occurrence of "**IL**" with "**IL**W"). Algorithm (3) takes a single formula $C \rhd D \in \Gamma$ as input, a set $\Delta \subseteq \Gamma \setminus \{C \rhd D\}$ of formulas of the form $E \rhd G$; and a set of formulas $\mathcal{D} \subseteq \Gamma$. It returns a positive answer if for some $(\Delta, C \rhd D)$-pair (N, P) and a visibility graph $\mathcal{G}$, for every $G \in P$ there is a rooted Veltman model $(\mathfrak{M}_G, w_G)$ of G such that:

1. $\dot{R}[w_G] \Vdash \neg\mathcal{D}, \neg N, \neg\{H : H \rhd J \in \Delta \text{ and } (G, J) \notin \mathcal{G}\}$;
2. $R[w_G] \Vdash \neg G, \neg\{H : H \rhd J \in \Delta \text{ and } [G]_{\overline{\mathcal{G}}} = [J]_{\overline{\mathcal{G}}}\}$.

Note that these checks can be computed with Algorithm (1).

We will now proceed to prove that algorithms (1), (2) and (3) are correct.

Lemma 6.8 Given $\Delta, \mathcal{B}, \mathcal{D} \subseteq \Gamma$, the following statements are equivalent:

1. there exists a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \Delta$, such that $\dot{R}[w] \Vdash \neg\mathcal{B}$ and $R[w] \Vdash \neg\mathcal{D}$;
2. there is a set $\Delta' \subseteq \Gamma$ and a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \mathrm{Full}(\Delta')$, such that $R[w] \Vdash \neg\mathcal{B}, \neg\mathcal{D}$, the set Δ' is a Boolean satisfiable extension of Δ, and $\mathrm{Full}(\Delta') \cup \neg\mathcal{B}$ is propositionally satisfiable.

Proof. See the proof of Lemma 6.2. Since we use the same model $(\mathfrak{M}, w)$ in both directions of the proof, we do not need to check if the characteristic property (W) is preserved. ⊣

The sets Δ^+ and Δ^- are given by (6.1) and (6.2), as before.

Lemma 6.9 Let Δ be a Boolean satisfiable subset of Γ, and $\mathcal{B}, \mathcal{D} \subseteq \Gamma$. The following statements are equivalent:

1. there exists a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \mathrm{Full}(\Delta)$ and $R[w] \Vdash \neg\mathcal{D}$;

2. for all $C \rhd D \in \Delta^-$, there is a rooted Veltman model $(\mathfrak{M}_{C \rhd D}, w_{C \rhd D})$ of $\{\neg(C \rhd D)\} \cup \Delta^+$, such that $R[w_{C \rhd D}] \Vdash \neg\mathcal{D}$.

Proof. The proof of this lemma is essentially the same as the proof of Lemma 6.3. The (1.)-to-(2.) direction is exactly the same, and since we are taking generated submodels of the model $(\mathfrak{M}, w)$ we do not need to check if the characteristic property is preserved in the newly defined models.

In the other direction, we should check whether the resulting model $(\mathfrak{M}, w)$ satisfies converse well-foundedness of $R \circ S_w$. However, every $(R \circ S_w)$-path that exists in the joint model $(\mathfrak{M}, w)$ is already present in one of the initial models $(\mathfrak{M}_{C \rhd D}, w_{C \rhd D})$ as an $(R \circ S_{w_{C \rhd D}})$-path. Thus, no $(R \circ S_w)$-loops can occur in this construction. $\dashv$

Lemma 6.10 Let $\Delta, \mathcal{D} \subseteq \Gamma$ where Δ is a set of formulas of the form $E \rhd G$, and assume $C \rhd D \in \Gamma \setminus \Delta$. The following statements are equivalent:

1. there exists a rooted model $(\mathfrak{M}, w)$ of $\{\neg(C \rhd D)\} \cup \Delta$, such that $R[w] \Vdash \neg\mathcal{D}$;

2. there is a $(\Delta, C \rhd D)$-pair (N, P) and a visibility graph $\mathcal{G}$ for it such that for every $G \in P$ there is a rooted Veltman model $(\mathfrak{M}_G, w_G)$ of G such that:

   (a) $\dot{R}[w_G] \Vdash \neg\mathcal{D}, \neg N, \neg\{H : H \rhd J \in \Delta \text{ and } (G, J) \notin \mathcal{G}\}$;

   (b) $R[w_G] \Vdash \neg G, \neg\{H : H \rhd J \in \Delta \text{ and } [G]_{\overline{\mathcal{G}}} = [J]_{\overline{\mathcal{G}}}\}$.

Proof. (1.) to (2.). Assume w.l.o.g. that the model $(\mathfrak{M}, w)$ has properties of the model $(\mathfrak{N}, y)$ described in Lemma 6.7. In particular, we assume the existence of worlds $w_{C \rhd D}$ and $w_{C \rhd D, G}$ as described there. To show (2), let (N, P) be the $(\Delta, C \rhd D)$-pair defined as follows: $P = \{B \in \Gamma : [B]_w \cap S_w[w_{C \rhd D}] \neq \emptyset\}$ and $N = \Gamma \setminus P$. Note that this coincides with $P_{C \rhd D}$ in Lemma 6.7.

Let us define the visibility graph $\mathcal{G} \subseteq P^2$ as follows. Given $G_1, G_2 \in P$, put $(G_1, G_2) \in \mathcal{G}$ if and only if $w_{C \rhd D, G_1} S_w w_{C \rhd D, G_2}$. The reflexivity and transitivity of $\mathcal{G}$ follow from the reflexivity and transitivity of S_w.

Let $G \in P$ and put $(\mathfrak{M}_G, w_G) := (\mathfrak{M}[w_{C \rhd D, G}], w_{C \rhd D, G})$ where $\mathfrak{M}[w_{C \rhd D, G}]$ is the submodel generated by $w_{C \rhd D, G}$. Also let $(W_G, R_G, \{S^G_x : x \in W_G\}, \Vdash_G) = \mathfrak{M}_G$.

Portions of properties (2a) and (2b) that appeared already in Lemma 6.4 ($\dot{R}[w_G] \Vdash \neg\mathcal{D}, \neg N$ and $R[w_G] \Vdash \neg G$) are proved exactly the same as before.

Let us verify $\dot{R}[w_G] \Vdash \neg\{H : H \rhd J \in \Delta \text{ and } (G, J) \notin \mathcal{G}\}$. Assume otherwise, i.e. for some $x \in \dot{R}[w_G]$ we have $x \Vdash H$ for some $H \rhd J \in \Delta$ and $(G, J) \notin \mathcal{G}$. Properties in Lemma 6.7 imply $x S_w w_{C \rhd D, J}$. Thus, $w_{C \rhd D, G} S_w w_{C \rhd D, J}$, contradicting $(G, J) \notin \mathcal{G}$.

Finally, let us verify that $R[w_G] \Vdash \neg\{H : H \rhd J \in \Delta \text{ and } [G]_{\overline{\mathcal{G}}} = [J]_{\overline{\mathcal{G}}}\}$. Assume otherwise, i.e. for some $x \in R[w_G]$ we have $x \Vdash H$ for some $H \rhd J \in \Delta$ and $[G]_{\overline{\mathcal{G}}} = [J]_{\overline{\mathcal{G}}}$. Again, Lemma 6.7 implies $x S_w w_{C \rhd D, J}$. Since $[G]_{\overline{\mathcal{G}}} = [J]_{\overline{\mathcal{G}}}$, we have $w_{C \rhd D, J} S_w w_{C \rhd D, G}$. Thus, $w_{C \rhd D, G} R x S_w w_{C \rhd D, G}$, contradicting the converse well-foundedness of $R \circ S_w$ in $\mathfrak{M}$.

(2.) to (1.). We construct $\mathfrak{M}$ as follows. Let $(W_G, R_G, \{S_x^G : x \in W_G\}, \Vdash_G) = \mathfrak{M}_G$ for all $G \in P$. First take the disjoint union U of the models $(\mathfrak{M}_G, w_G)$ for $G \in P$. Prepend a new world w as a predecessor of every world in U. Let

$$Q = \{(x, w_{G_2}) \in U^2 : (G_1, G_2) \in \mathcal{G}, x \in \mathfrak{M}_{G_1}, x \Vdash E, E \rhd G_2 \in \Delta\}$$

Take $S_w = (=_U \cup \bigcup_{G \in P} R_G \cup Q)^+$. This concludes the construction of $\mathfrak{M}$. Let us first verify that $\mathfrak{M}$ is an **IL**W-model.

The only non-trivial property is the converse well-foundedness of $R \circ S_w$. Suppose for a contradiction that there is an $(R \circ S_w)$-loop. Since Γ is finite, for some $G \in P$ the world w_G must occur more than once in the loop. If we fix two occurrences of w_G in the loop and look at the part of the chain bounded by these occurrences, we obtained another (although not necessarily different) $(R \circ S_w)$-loop, but this time we are guaranteed to have only a finite amount of pairwise distinct transitions. Furthermore, it is easy to see that our $(R \circ S_w)$-loop induces some $(R \circ Q^+)$-loop (remove each S_w-transition that is an $=_U$-transition and collapse all consecutive R-transitions into a single R-transition).

Note that by the definition of Q, whenever xQy, if G_x and G_y are indices of the models containing x and y (resp.), we have $(G_x, G_y) \in \mathcal{G}$. In particular this means that the formulas G_x, where x is any world from the $(R \circ Q^+)$-loop, all belong to the same equivalence class of $\mathcal{G}$. Recall that our loop contains only finitely many distinct Q and R transitions. Fix some transition xRy that appears infinitely often in the loop.

Since x is the target of some Q-transition (an $(R \circ Q^+)$-loop cannot contain two consecutive R-moves), we have $x = w_H$ for some $H \in P$. Since $xRyQz$ for some z, there is some $E \rhd G \in \Delta$ such that $(H, G) \in \mathcal{G}$ and $y \Vdash E$. However, since $[G]_{\mathcal{G}} = [H]_{\mathcal{G}}$, we have $R[w_H] \Vdash \neg E$ (see (2b)). This contradicts $w_H = xRy \Vdash E$.

It remains to prove that $w \Vdash \neg(C \rhd D), \Delta$ and $R[w] \Vdash \neg\mathcal{D}$. To see that $w \Vdash \neg(C \rhd D)$, note that $wRw_C \Vdash C$ and that, since $D \in N$, all the models $\mathfrak{M}_G$ validate $\neg D$. Similarly, $R[w] \Vdash \neg\mathcal{D}$. Let $E \rhd G \in \Delta$ be an arbitrary formula, and $x \in R[w]$ an arbitrary world such that $x \Vdash E$. Assume that $x \in \mathfrak{M}_H$. Since $x \Vdash E$, we have $(H, G) \in \mathcal{G}$ (see (2a)). Now the definition of Q implies xQw_G, thus also $xS_w w_G$. $\dashv$

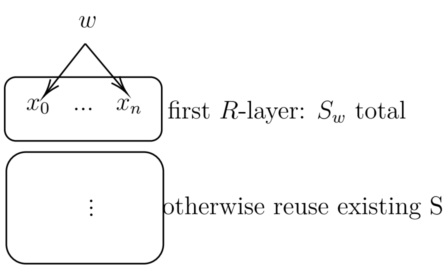
Theorem 6.11 The logic **IL**W is PSPACE-decidable.

Proof. The proof is almost exactly the same as the proof of Lemma 6.5, but this time referring to Lemmas 6.8–6.10. A small change is that we should convince ourselves that finding a good visibility graph in Algorithm (3) takes a polynomial amount of space. When we loop through the possible $(\Delta, C \rhd D)$-pairs (N, P), we should try all possible visibility graphs for P. A visibility graph for P is a binary relation on P, and thus requires at most $|P|^2$ bits for its representation. Verifying its reflexivity and transitivity is a simple PTIME operation. $\dashv$

6.4 The logic ILP

It is well known that **IL**P extends **IL**W. However, the logic itself is simpler. Models of this logic have a convenient property that $wRxRuS_wv$ implies uS_xv. For this reason, when building a world w and dealing with a formula $E \rhd G \in \Delta^+$ (we are trying to ensure $w \Vdash E \rhd G$), we only need to ensure a world satisfying G exists if there will be an immediate R-successor of w satisfying E (as far as the formula $E \rhd G$ is concerned; we might need a world satisfying G for other reasons, of course). If there will be non-immediate R-successors satisfying E, the construction of their S_w-successors satisfying G will be handled by their immediate R-predecessors (which is, by this case's assumption, not the world w). As is shown the picture below (where $x_0, \ldots, x_n$ are the worlds whose construction was triggered by the world w), we can make S_w total in the first R-layer. We could not have done this with **IL**W, since one formula had only one witness, and connecting too many worlds to it may have caused loops. Here, on the other hand, the worlds x_i (pictured) only serve to cover each other's needs: their children will take care of their S_w-needs themselves.

Thus, there will be some "repetition" and models will on average be larger than it was the case for **IL** and **IL**W, but the construction itself is arguably simpler.



In this section all models are **IL**P-models.

We do not need any auxiliary lemmas, so let us present the algorithm. As before, we have three algorithms (1), (2) and (3) with the same purpose as their **IL** and **IL**W-counterparts.

Let A be the formula whose satisfiability we are interested in, and $\Gamma := \mathrm{Sub}(A) \cup \{\bot\}$.

Algorithms (1) and (2) are defined just like before. Please see above for their descriptions (replacing each implicit or explicit occurrence of "**IL**" or "**IL**W" with "**IL**P").

Algorithm (3) takes as input a single formula $C \rhd D \in \Gamma$, a set $\Delta \subseteq \Gamma \setminus \{C \rhd D\}$ of formulas of the form $E \rhd G$; and a set of formulas $\mathcal{D} \subseteq \Gamma$. It returns a positive answer if for some $(\Delta, C \rhd D)$-pair (N, P), for every $G \in P$ there is a rooted Veltman model $(\mathfrak{M}_G, w_G)$ of G such that:

1. $\dot{R}[w_G] \Vdash \Delta, \neg\mathcal{D}, \neg N$;

2. $R[w_G] \Vdash \neg G$.

Note that these checks can be computed with Algorithm (1).

We will now proceed to prove that Algorithms (1), (2) and (3) are correct.

Lemma 6.12 Given $\Delta, \mathcal{B}, \mathcal{D} \subseteq \Gamma$, the following statements are equivalent:

1. there exists a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \Delta$, such that $\dot{R}[w] \Vdash \neg\mathcal{B}$ and $R[w] \Vdash \neg\mathcal{D}$;

2. there is a set $\Delta' \subseteq \Gamma$ and a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \text{Full}(\Delta')$, such that $R[w] \Vdash \neg\mathcal{B}, \neg\mathcal{D}$, the set Δ' is a Boolean satisfiable extension of Δ, and $\text{Full}(\Delta') \cup \neg\mathcal{B}$ is propositionally satisfiable.

Proof. See the proof of Lemma 6.2. Since we use the same model $(\mathfrak{M}, w)$ in both directions of the proof, we do not need to check if the characteristic property (P) is preserved. $\dashv$

Lemma 6.13 Let Δ be a Boolean satisfiable subset of Γ, and $\mathcal{B}, \mathcal{D} \subseteq \Gamma$. The following statements are equivalent:

1. there exists a rooted model $(\mathfrak{M}, w)$ such that $w \Vdash \text{Full}(\Delta)$ and $R[w] \Vdash \neg\mathcal{D}$;

2. for all $C \rhd D \in \Delta^-$, there is a rooted Veltman model $(\mathfrak{M}_{C\rhd D}, w_{C\rhd D})$ of $\{\neg(C \rhd D)\} \cup \Delta^+$, such that $R[w_{C\rhd D}] \Vdash \neg\mathcal{D}$.

Proof. See the proof of Lemma 6.3. Here we should make an additional check of whether we still have the characteristic property (P) once we merge (the roots of) the models $\mathfrak{M}_{C\rhd D}$ into one.

So, assume uS_wv and let w' be between w and u. Then for some $C \rhd D$ we have $uS_{w_{C\rhd D}}v$. Since $\mathfrak{M}_{C\rhd D}$ satisfies (P), we have $uS_{w'}v$ in $\mathfrak{M}_{C\rhd D}$, and thus $uS_{w'}v$ in $\mathfrak{M}$. $\dashv$

Lemma 6.14 Let $\Delta, \mathcal{D} \subseteq \Gamma$ where Δ is a set of formulas of the form $E \rhd G$, and assume $C \rhd D \in \Gamma \setminus \Delta$. The following are equivalent:

1. there exists a rooted model $(\mathfrak{M}, w)$ of $\{\neg(C \rhd D)\} \cup \Delta$, such that $R[w] \Vdash \neg\mathcal{D}$;

2. there is a $(\Delta, C \rhd D)$-pair (N, P) such that for every $G \in P$ there is a rooted Veltman model $(\mathfrak{M}_G, w_G)$ of G such that:

   (a) $\dot{R}[w_G] \Vdash \Delta, \neg\mathcal{D}, \neg N$;

   (b) $R[w_G] \Vdash \neg G$.

Proof. (1.) to (2.). This is exactly the same as in the proof of Lemma 6.4, but now we have an additional property to check: $\dot{R}[w_G] \Vdash \Delta$. Since wRw_G and $\mathfrak{M}$ satisfies (P), whenever $w \Vdash H \rhd J$, we have $\dot{R}[w_G] \subseteq R[w] \Vdash H \rhd J$.

(2.) to (1.). Let $(W_G, R_G, \{S^G_x : x \in W_G\}, \Vdash_G) = \mathfrak{M}_G$ for all $G \in P$. We construct $\mathfrak{M}$ as follows. First take the disjoint union U of the models $(\mathfrak{M}_G, w_G)$ for $G \in P$. Prepend a new world w as a predecessor of every world in U. Let $S_w = \{(w_G, x) : G \in P, x \in R[w]\} \cup \bigcup_{G \in P} S_{w_G}$. This concludes the construction of $\mathfrak{M}$. The following properties are easy to check: the transitivity and converse well-foundedness of R; the property that $xRyRz$ implies yS_xz for $x \in R[w]$, the reflexivity of S_x (on $R[x]^2$) for $x \in R[w]$, and the transitivity of S_x for $x \in R[w]$.

Reflexivity of S_w: assume wRx. If $x = w_G$ for some $G \in P$, we required that $(x, x) \in S_w$; otherwise we use the reflexivity of S_{w_G} for the formula $G \in P$ such that $x \in \mathfrak{M}_G$.

If $wRxRy$, and $x = w_G$ for some $G \in P$, then xS_wy follows from the definition of S_w. If $wRxRy$, and $x \neq w_G$ (for all $G \in P$), there must be some $G \in P$ such that w_GRxRy. Then $(x, y) \in S_{w_G} \subseteq S_w$.

Let us verify the transitivity of S_w. Assume aS_wbS_wc and not aS_wc. Thus, in particular, $a \neq w_G$ for all $G \in P$. This and aS_wb imply $aS_{w_G}b$ for some $G \in P$. Now w_GRb implies $b \neq w_H$ for all $H \in P$. This and bS_wc imply $bS_{w_H}c$ for some $H \in P$. Since $R[w_G] \cap R[w_H] = \emptyset$ if $G \neq H$, it follows that $bS_{w_G}c$. By the transitivity of S_{w_G}, we have $aS_{w_G}c$, and so aS_wc.

Let us verify that the characteristic property (P) is preserved. It is clearly preserved if only worlds from $R[w]$ are involved. So assume uS_wv and let w' be between w and u. Since $wRw'Ru$, we have $u \neq w_G$ for all $G \in P$, so $uS_{w_G}v$ for some $G \in P$. If w_GRw' and since (P) holds for $\mathfrak{M}_G$, we have $uS_{w'}v$. Otherwise, $w_G = w'$. In this case, we already obtained $uS_{w'}v$.

It remains to prove $w \Vdash \neg(C \rhd D), \Delta$ and $R[w] \Vdash \neg\mathcal{D}$. To see that $w \Vdash \neg(C \rhd D)$, note that $wRw_C \Vdash C$ and that, since $D \in N$, all the models $\mathfrak{M}_G$ satisfy $\neg D$. Similarly, $R[w] \Vdash \neg\mathcal{D}$. Let $E \rhd G \in \Delta$ be an arbitrary formula, and $x \in R[w]$ an arbitrary world such that $x \Vdash E$. If $x = w_H$ for some $H \in P$, we have $xS_ww_G \Vdash G$. Otherwise, there is some $H \in P$ such that wRw_HRx. By the assumption that $\dot{R}[w_H] \Vdash \Delta$, in particular we have $w_H \Vdash E \rhd G$. Since $w_HRx \Vdash E$, there is y with $xS_{w_H}y \Vdash G$. Since $S_{w_H} \subseteq S_w$, we have xS_wy. ⊣

Theorem 6.15 The logic **IL**P is PSPACE-decidable.

Proof. The proof is analogous to the proof of Theorem 6.11, except that now we use Lemmas 6.12, 6.13, and 6.14. ⊣

Corollary 6.16 The logic **IL**P is PSPACE-complete.

Proof. By Theorem 6.15 and the fact that it conservatively extends **GL**. ⊣

Note 6.17 Lemmas 6.2, 6.3, 6.4 (**IL**), 6.8, 6.9, 6.10 (**IL**W), 6.12, 6.12, and 6.12 (**IL**P) can be slightly strengthened by adding the condition of *finiteness* to one (or both) of the claims (1.) or (2.) that are present in all these lemmas. Depending on a lemma, this is either obvious, or a simple consequence of the finite model property (which all three logics, **IL**, **ILW** and **ILP**, have). This fact is not crucial for our proofs since we ensure finiteness by other means, namely by ensuring R-maximality with appropriate formulas.

6.5 Other logics

It is natural to ask whether the results of this chapter extend to other interpretability logics. Probably the best candidate logics to test next are interpretability logics that are already known to be decidable. These are (to the best of our knowledge) **ILM** ([18]), **ILM**$_0$ ([49]), **ILW*** ([49]), **ILR** ([50]) and **ILP**$_0$ ([50]).

Note also that in [49] and [50] the decidability of **ILM**$_0$, **ILW***, **ILR** and **ILP**$_0$ is proved using generalised Veltman semantics, in which S_w-successors are sets of worlds. Therefore, an adaptation of the technique of this chapter should take that into consideration. However, the remaining logic on this list, the logic **ILM**, does have the finite model property with respect to ordinary Veltman semantics. This is also the only interesting extension of **IL** that is known to be complete and to have the finite model property with respect to ordinary Veltman semantics, but for which a complexity result is not obtained yet (either in this chapter or elsewhere).

So let us briefly comment on what we tried to do and why we did not succeed in applying the approach that worked for **IL**, **ILW** and **ILP** to **ILM**.

In the piece-by-piece approach of building models that we've seen in this chapter, the set of (direct) successors of any world w is split into disjoint subsets (which we refer to as *pieces*). Every piece corresponds to a single formula of form $E \rhd G$ we wish to be true in the world w. Inside every piece, or in the case of **ILP** inside the first R-layer of every piece, we inserted as many S_w relation as required in order to have all formulas of form $E \rhd G$ we wish to be true in the world w, true. So, if we wanted to satisfy $E \rhd G$ in w and at the same time wanted to satisfy E in a world u in one of the pieces, we would connect u to the (single) distinguished witness for G in this piece. This does not work well for **ILM**. Recall that the frame condition (M) is that $wRuS_wvRz$ implies uRz. So, when selecting the target of an S_w-transition, one has to ensure that the set of boxed formulas of the source is a subset of the set of boxed formulas of the target. Otherwise, we would have uS_wv with $u \Vdash \Box B$ and $v \nVdash \Box B$, for some $\Box B$. Hence there would be z with $vRz \nVdash B$. The frame condition requires uRz, contradicting $u \Vdash \Box B$. Since for every formula we want a single distinguished witness per piece, we would have to select a particular set of boxed formulas for this distinguished witness. There is no obvious reason why this single set of boxed formulas would be compatible with all the worlds we would wish to connect to the witness. That is, we might want to connect different sources (satisfying different sets of boxed formulas) to this distinguished witness. Even if we drop the requirement for having a single distinguished witness, there is still no obvious reason why even a polynomial number of witnesses would suffice. The number of different

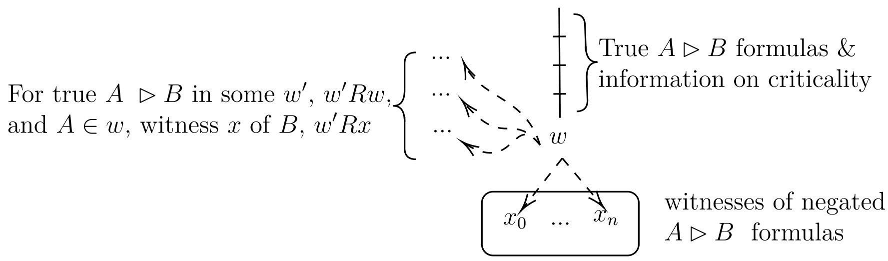


Figure 6.1: A "naive" approach to building models in a space-efficient manner.

possible selections of boxed formulas is, roughly, $2^{|\Gamma|}$, where Γ is the set of subformulas of the formula whose satisfiability our algorithm is currently verifying.

One alternative to the approach used in this chapter is to drop the general approach of building self-sufficient pieces altogether. A possible alternative is to try and implement something like the procedure used in Goris' and Joosten's construction method ([27, 28]). With this approach, every world would take care of all the problems (as in our approach) *and* all the deficiencies that the addition of this world caused in the model. The latter is unlike the approach we used in this chapter, since in this chapter we let the parent of a world take care of solving children's deficiencies (solving deficiencies can be done efficiently for the logics studied in this chapter, since we take care of all children's deficiencies in one go). It is easy to see that the alternative approach can quickly lead to an infinite branch if we're not careful. Suppose the ambient logic is **IL** and we have $w \Vdash E \rhd G, G \rhd E$ and $wRu \Vdash E$. We first need a witness v of G so that we can let uS_wv and solve the deficiency caused by $w \Vdash E \rhd G$ and $wRu \Vdash E$. The world v causes a new deficiency: $w \Vdash G \rhd E$ and $wRv \Vdash G$. So we need a witness for E, which will then require a witness for G, etc. For **IL** we can try to solve this by keeping track of all the formulas satisfied among all the S_w-transitions that occurred after the last R-transition. Whenever we make an S_w-transition uS_wv, we can add another S_w-transition vS_wu. Since the ambient logic is **IL**, no property of a Veltman model is invalidated by such additions. There will never be more than $|\Gamma|$ adjacent S_w-transitions since after $|\Gamma|$ adjacent S_w-transitions every possible deficiency can be solved by making a transition to one of the worlds in the preceding S_w-sequence. However, it's not clear if and how we can ensure (in polynomial space) that the total number of R-transitions is bound.

For **ILM**, the approach described in the last paragraph might be fruitful. The requirement that would work for **IL**, that whenever we make an S_w-transition uS_wv we add another S_w-transition vS_wu, is not sufficiently subtle for **ILM**, since it does not take into account the sets of boxed formulas. We should only allow S_w-transitions towards worlds whose sets of boxed formulas are supersets of the corresponding set for the current world. Every time we make a transition (either R or S_x for some x) the target world has at least as many boxed formulas. This property might[6] imply that the recursion depth does not

[6] We only sketch the approach in this section, so we cannot claim this categorically; there may be

exceed $|\Gamma|^3$:

- There are never more than $|\Gamma|$ R-transitions in a single branch since every world is R-maximal for some formula in Γ.
- Between two R-transitions we can make S_x-transitions, and x can be one of the at most $|\Gamma|$ worlds.
- For a fixed x, there can be at most $|\Gamma|$ adjacent S_x-transitions where the set of boxed formulas stays the same.

Similarly, we preserved the polynomial branching factor: every world causes at most $|\Gamma|$ problems, and at most $|\Gamma|^2$ deficiencies. Unfortunately, we did not yet succeed in turning this sketch into an actual algorithm. One of the issues we have to solve is how to resolve deficiencies that appear after performing the closure under the characteristic property (in this process new R-transitions are inserted in the model). If space was not a problem, we could keep track of the worlds that are going to become connected after the closure. However, it is not clear why would such a list of worlds be of polynomial length. A possible solution would be not to keep track of the worlds, but rather of their sets of true $\rhd$-formulas and their sets of boxed formulas. Since the sets of boxed formulas is monotonically increasing, the number of possible combinations would fit into polynomial space. We aim to continue working on this issue and resolve it in future work.

unforeseen issues.

Chapter 7

The approximating theory

In this chapter we prove the arithmetical validity of the principles contained in the two series of principles $\mathsf{R_n}$ and $\mathsf{R^n}$, which were recently introduced in [29]. Both series are already known to be arithmetically sound; this was established using reasoning based on definable cuts. We wanted to provide a different proof of the same result and aimed to use the method of approximating theories introduced in [40]. In [40] this approach is formalised as a system AtL. The benefit of AtL is that such proofs seem to highly resemble the purely modal proofs of these principles within the logic $\mathbf{IL}\mathsf{P}$.[1] It turned out that the original version of AtL, the version of this system presented in [40], did not suffice to prove the arithmetical soundness of the series $\mathsf{R^n}$ and $\mathsf{R_n}$. At least we were not able to find full proofs of the two series while staying inside the old version of AtL.

The aim of this chapter is to present an extended system (which we still call AtL), and prove that the two series are arithmetically valid by reasoning in the system. Together with the authors of [40] we extend and generalise the results of [40]. Unlike most other chapters, this content is not yet submitted to a journal, but we plan to do this soon ([39]).

This chapter is for the most part joint work with Joost J. Joosten and Albert Visser, and the modal soundness results w.r.t. generalised Veltman semantics are joint work with Jan Mas Rovira (see [44] for details).

7.1 Preliminaries

In this part of the thesis we will be using reasoning in and over weak arithmetics. To this end, let us start by describing the theory S^1_2, introduced by Buss in [9]. This is a finitely axiomatisable and weak first-order theory of arithmetic. The signature of S^1_2 is

$$(S, |\cdot|, \lfloor\frac{1}{2}\cdot\rfloor, +, \cdot, \#, =, \leq).$$

The intended interpretation of $|\cdot|$ is the length of its argument when expressed in the binary number system. In other words, $|n|$ is (in the intended interpretation) equal

[1] There is another extension of $\mathbf{IL}$, CuL, where arithmetical soundness proofs resemble $\mathbf{IL}\mathsf{M}$-proofs. See Section 8.4 for the definition of CuL.

to $\lceil \log_2(n+1) \rceil$. The intended interpretation of $\lfloor \frac{1}{2} \cdot \rfloor$ is precisely the one suggested by the notation. The symbol # is pronounced "smash" and has the following intended interpretation ("the smash function"):

$$n \# m = 2^{|n||m|}.$$

Other symbols are more standard and are intended to be interpreted in the expected way.

The motivation for the smash function is that it gives an upper bound to Gödel numbers of formulas obtained by substitution. Suppose A is a formula, x a variable and t a term. Given the Gödel numbers of A and t (denoted with $\lceil A \rceil$ and $\lceil t \rceil$, as usual), the Gödel number of $A(x \mapsto t)$ will not surpass $\lceil A \rceil \# \lceil t \rceil$. Here the assumption is that both the numeral representation and the Gödel numbers we work with are efficient. For example, we can take the Gödel number of a string of symbols to be its ordinal number in an arbitrary computationally very easy but otherwise fixed enumeration of all strings in the language of S^1_2.

As for the numerals, we can use *efficient numerals*, defined recursively as follows:

$$\underline{0} \mapsto 0;$$
$$\underline{2n} \mapsto (SS0) \cdot \underline{n};$$
$$\underline{2n+1} \mapsto S((SS0) \cdot \underline{n}).$$

Clearly, efficient numerals have about the same growth rate as the corresponding binary representations. We require the coding of our choice to have asymptotically the same growth order as the Gödel numbers of the efficient numerals. We also require that the code of a subterm is always smaller than the entire term, and similarly for formulas. We will consider such coding *natural*. For example, using powers of prime numbers to code sequences and terms of the form $SSS...0$ as numerals will not be considered natural in this context. See [10] for details.

Before introducing (some of) the axioms of S^1_2, we will first define a certain hierarchy of formulas in the language of S^1_2. We will say that a quantifier is *bounded* if it is of the form $(Qx \leq t)$ where t is a term that does not involve x.[2] A quantifier is *sharply bounded* if it is of the form $(Qx \leq |t|)$ where t is a term that does not involve x

Definition 7.1 ([10]) Let Δ^b_0, Σ^b_0, and Π^b_0 stand for the set of formulas all of whose quantifiers are sharply bounded. We define Δ^b_i, Σ^b_i, and Π^b_i for $i > 0$ as the minimal sets satisfying the following conditions:

1. If A and B are Σ^b_i-formulas, then $A \wedge B$ and $A \vee B$ are Σ^b_i-formulas.

2. If A is a Π^b_i-formula and B is a Σ^b_i-formula, then $\neg A$ and $A \to B$ are Σ^b_i-formulas.

[2]By "$(Qx \leq t)$" we mean "$(\exists x)(x \leq t \wedge \ldots$" if Q is $\exists$, and similarly if Q is $\forall$.

3. If A is a Π^b_{i-1}-formula, then A is a Σ^b_i-formula.

4. If A is a Σ^b_i-formula, x a variable and t is a term not involving x, then $(\forall x \leq |t|)A$ is a Σ^b_i-formula.

5. If A is a Σ^b_i-formula, x a variable and t is a term not involving x, then $(\exists x \leq t)A$ and $(\exists x \leq |t|)A$ are Σ^b_i-formulas.

6. The first five conditions are to be repeated in the dual form: with the roles of Σ and Π, and $\exists$ and $\forall$, swapped in all places.

7. A formula A is a Δ^b_i-formula if it is equivalent over predicate logic both to a Σ^b_i-formula and to a Π^b_i-formula.

Thus, this hierarchy is analogous to the standard arithmetical hierarchy, with bounded quantifiers in the role of unbounded quantifiers, and sharply bounded quantifiers in the role of bounded quantifiers.

Definition 7.2 (The polynomial induction schema [10]) Let Φ be a set of formulas which may contain zero or more free variables. We define Φ-PIND axioms to be the formulas

$$A(x \mapsto 0) \wedge (\forall x)(A(x \mapsto \lfloor \frac{1}{2} x \rfloor) \to A) \to \forall x A,$$

for all $A \in \Phi$ and all variables x.

Thus, when proving facts using the schema of polynomial induction, in the inductive step we are only allowed to refer to the property obtained for $\lfloor \frac{1}{2} n \rfloor$. This is, of course, less convenient than the standard schema of mathematical induction where we can use the property obtained for $n - 1$.

We obtain S^1_2 by extending a certain list of 32 quantifier-free formulas (dubbed *BASIC*, see e.g. [10]) with all Σ^b_1-PIND axioms.

This somewhat unusually axiomatised theory has a nice connection to computational complexity, as the next theorem shows.

Theorem 7.3 ([9]) We have the following.

- Suppose $\mathsf{S}^1_2 \vdash (\forall x)(\exists y)A(x, y)$ for some Σ^b_1-formula A. Then there is a PTIME-computable function f_A such that if $f_A(x) = y$ then $A(x, y)$ holds (f_A is a *witnessing function* for A), and $\mathsf{S}^1_2 \vdash (\forall x)A(x, f_A(x))$.

- Conversely, suppose f is a PTIME-computable function. Then there is a Σ^b_1-formula A_f such that $A_f(x, y)$ holds if and only if $f(x) = y$, and $\mathsf{S}^1_2 \vdash (\forall x)(\exists y)A_f(x, y)$.

Theories in this chapter will be Δ_1^b-axiomatised theories (i.e. having PTIME-decidable axiomatisations). Moreover, we will always assume that any theory we consider comes with a designated interpretation of S_2^1. That is, when we say "a theory", we mean a pair of an actual theory together with some singled-out and fixed interpretation of S_2^1.

As S_2^1 is finitely axiomatisable, and in this chapter we work with a principle similar to a principle of interpretability P which is valid in finitely axiomatisable theories, the theory S_2^1 is a natural choice. However, we could have used another weak theory, such as $I\Delta_0 + \Omega_1$. Here $I\Delta_0$ refers to Q extended with the schema of induction, restricted to Δ_0-formulas of the standard hierarchy. The axiom Ω_1 states the totality of ω_1, the function $n \mapsto 2^{|n|^2}$ (note the similarity with the smash function). This theory is a less natural choice as it is not yet known whether it is finitely axiomatisable. An even more powerful system is $\mathrm{I}\Delta_0 + \mathsf{exp}$. Here, exp states the totality of exponentiation. This can be expressed by a Π_2 statement since the graph of the exponentiation function can be given by a bounded formula.

A principle similar to induction is that of *collection*, in particular Σ_1-collection.

Definition 7.4 (Σ_1-collection) The schema

$$(\forall n)((\forall x < n)(\exists y)A(x, y) \to (\exists m)(\forall x < n)(\exists y < m)A(x, y))$$

where A is restricted to Σ_1-formulas possibly with parameters, is the Σ_1-collection schema. The principle is occasionally useful, however we will have to find ways to avoid it as it is not available in S_2^1.

7.1.1 Formalised interpretability

Before introducing formalised interpretability, let us say a few words on formalised provability. *The provability predicate* of a theory T, usually denoted as Pr_T, is the natural formalisation of the statement "there exists a T-proof of a given formula". Let us denote the efficient numeral of the (natural) Gödel number of A by $\ulcorner A \urcorner$. Sufficiently strong theories (such as S_2^1 and $\mathrm{I}\Delta_0 + \Omega_1$) prove the *Hilbert–Bernays–Löb derivability conditions* ([32]):

1. for all A, if $T \vdash A$, then $T \vdash \mathsf{Pr}_T(\ulcorner A \urcorner)$;
2. for all A, $T \vdash \mathsf{Pr}_T(\ulcorner A \to B \urcorner) \to (\mathsf{Pr}_T(\ulcorner A \urcorner) \to \mathsf{Pr}_T(\ulcorner B \urcorner))$;
3. for all A, $T \vdash \mathsf{Pr}_T(\ulcorner A \urcorner) \to \mathsf{Pr}_T(\ulcorner \mathsf{Pr}_T(\ulcorner A \urcorner) \urcorner)$.

These conditions suffice to show that $T \vdash \mathsf{Pr}_T(\ulcorner 0 = 1 \urcorner)$ and consequently $T \vdash 0 = 1$ follows from $T \vdash \neg\mathsf{Pr}_T(\ulcorner 0 = 1 \urcorner)$, i.e. the Gödel's second incompleteness theorem. These conditions also suffice to show that the following holds:

$$\text{if } T \vdash \mathsf{Pr}_T(\ulcorner A \urcorner) \to A, \text{ then } T \vdash A.$$

Thus T is only "aware" that $\mathsf{Pr}_T(\ulcorner A \urcorner)$ implies A in case the conditional is trivially satisfied by the provability of its consequent. This entailment is known as *Löb's rule*. In fact, T is "aware" of this limitation (*formalised Löb's rule*):

$$T \vdash \mathsf{Pr}_T(\ulcorner \mathsf{Pr}_T(\ulcorner A \urcorner) \to A \urcorner) \to \mathsf{Pr}_T(\ulcorner A \urcorner).$$

If we replace occurrences of $\mathsf{Pr}_T(\ulcorner A \urcorner)$ with $\Box A$, for all A, we can formulate the facts and rules above in the language of propositional modal logic. The provability logic **GL** is the extension of the basic modal logic **K** with an additional axiom schema representing Löb's formalised rule:

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

In his well-known result, Solovay [57] established arithmetical completeness for this logic.

The predicate Pr_T satisfies the following property:

$$T \vdash A \text{ if and only if } \mathbb{N} \models \mathsf{Pr}_T(\ulcorner A \urcorner). \tag{7.1}$$

Now, one might ask if there are other predicates, apart from Pr_T, that satisfy the same Property (7.1) (i.e. which extensionally coincide with the predicate Pr_T). Indeed, there are many other such predicates. One such predicate is given in [22] as the formalisation of "provable by the union of all consistent initial segments of T".[3] Let us call this notion *Feferman-provability*. As we're interested only in consistent theories, clearly this predicate has the same extension as the predicate Pr_T. However, it is provable within **PA** that $0 = 1$ is not Feferman-provable. This is of course not the case with $\mathsf{Pr}_{\mathbf{PA}}$, as that would contradict Gödel's second incompleteness theorem.

If we're dealing with some poly-time decidable theory T, by Theorem 7.3 there is a Σ_1^b-predicate verifying whether a number codes a T-proof of a formula. This implies that the provability predicate, claiming that a proof exists for some given formula, is a $\exists\Sigma_1^b$-predicate. This is convenient because for S_2^1 we have provable $\exists\Sigma_1^b$-completeness.

We now move on and consider interpretability. There are various notions of formalised interpretability (see Theorem 1.2.10. of [37] for a discussion on their relationships). Here we are interested in *theorems interpretability*, i.o.w. we say that k is an interpretation of V in U (we write $k : U \rhd V$) if and only if

$$\forall \phi \ (\Box_V \phi \to \Box_U \phi^k).$$

Here $\Box_V$ and $\Box_U$ are to be understood as, of course, the provability predicates of V and U, respectively. The k-translation of ϕ is denoted as ϕ^k. If V is a finitely axiomatisable theory, then $U \rhd V$ is in fact a $\exists\Sigma_1^b$ sentence. This is due to the fact that for finitely

[3] In [22] the theory T is assumed to be a consistent reflexive extension of **PA**.

axiomatisable theories V, their interpretability in U boils down to the statements stating the provability of the translation of the conjunction of these axioms. As the theories studied in this chapter are all Δ_1^b-axiomatisable, the aforementioned statement is $\exists\Delta_1^b$, in particular $\exists\Sigma_1^b$.

7.2 Tweaking the axiom set

For finitely axiomatised theories V, we have:

$$\mathsf{S}_2^1 \vdash U \rhd V \to \Box_{\mathsf{S}_2^1}(U \rhd V),$$

because $U \rhd V$ is a $\exists\Sigma_1^b$-sentence. Recall that in this chapter all theories are assumed to be Δ_1^b-axiomatised. If this were not the case, $U \rhd V$ need not, of course, be a Σ_1^b-sentence, even for finitely axiomatised theories V.

To mimic the P-style behaviour for an arbitrary theory V, we will modify V to a new theory V' that approximates V to obtain $\mathsf{S}_2^1 \vdash U \rhd V \to \Box_{\mathsf{S}_2^1}(U \rhd V')$. Of course, the new theory V' should be sufficiently like V to be useful. Thus, we define a theory V' that is extensionally the same as V, but for which $U \rhd V'$ is a statement that is so simple that under the assumption that $U \rhd V$, we can easily infer $\Box_{\mathsf{S}_2^1}(U \rhd V')$.

7.2.1 The approximating theory defined

The idea is simple and as follows. Given some translation k, let us *define* the set of axioms V' as consisting of just those axioms ϕ of V such that $U \vdash \phi^k$. Note that, if $k : U \rhd V$, then V and V' have the same axioms. However, when V is not finitely axiomatisable in general, we cannot take this insight with us when we proceed to reason inside a box. In formulas: we do have $k : U \rhd V \;\Rightarrow\; V \equiv V'$ but in general we do not have $k : U \rhd V \;\Rightarrow\; \Box(V \equiv V')$.

This idea works modulo some trifling details. Firstly, the definition of the new axiom set does not have the right complexity. Secondly, if the argument is not set up in a careful way, we may seem to need both Σ_1-collection and exp. We shall use a variation of Craig's trick so that the axiom sets that we consider will remain to be Δ_1^b-definable. The same trick makes the use of strong principles, like Σ_1-collection and exp, superfluous.

Definition 7.5 Let U and V be Δ_1^b-axiomatised theories. Moreover, let k be a translation of the language of V into the language of U that includes a domain specifier. We define $V^{[U,k]}$ as follows.

$$\begin{aligned}\mathsf{axioms}_{V^{[U,k]}}(x) \quad := \quad & \exists\, p, \varphi{<}x \ \Big(x = \ulcorner \varphi \wedge (\underline{p} = \underline{p})\urcorner \wedge \\ & \qquad\qquad \mathsf{axioms}_V(\varphi) \wedge \mathsf{proof}_U(p, \varphi^k)\Big).\end{aligned}$$

It appears at first sight that we have defined a Σ_1^b-formula. It is clear that $\mathsf{axioms}_{V^{[U,k]}}(x)$ is poly-time decidable if $\mathsf{axioms}_V(x)$ and $\mathsf{axioms}_U(x)$ are. Thus, this formula clearly describes a PTIME-computable procedure. By Theorem 7.3, we know it has to be Δ_1^b-definable.

The following lemma tells us that S^1_2 verifies that $k : U \rhd V$ implies that V and $V^{[U,k]}$ are *extensionally equal*. Actually, $V \rhd V^{[U,k]}$ always holds and does not depend on the assumption $k : U \rhd V$.

Lemma 7.6 Let U and V be Δ_1^b-axiomatised theories. We have

1. $\mathsf{S}^1_2 \vdash \forall k\ (\mathsf{id} : V \rhd V^{[U,k]})$.
2. $\mathsf{S}^1_2 \vdash \forall k\ (k : U \rhd V \to \mathsf{id} : V^{[U,k]} \rhd V)$.

Proof. Ad (1). Reason in S^1_2. We have to show: $\Box_{V^{[U,k]}}\varphi \to \Box_V\varphi$. This is easily seen to be true, since we can replace every axiom $\varphi \wedge (\underline{p} = \underline{p})$ of $V^{[U,k]}$ by a proof of $\varphi \wedge (\underline{p} = \underline{p})$ from the V-axiom φ. The resulting transformation is clearly p-time.

Ad (2). Reason in S^1_2. Suppose $k : U \rhd V$ and $\Box_V\varphi$. We set out to prove $\Box_{V^{[U,k]}}\varphi$. From our assumption $\Box_V\varphi$ we obtain a proof p of φ from V-axioms $\tau_0, \ldots, \tau_n$. We would be done obtaining a proof p' if we could replace every axiom occurrence of τ_i in p by

$$\frac{\tau_i \wedge (\underline{q_i} = \underline{q_i})}{\tau_i}\ \wedge E, l$$

where q_i were a proof in U of τ_i^k, so that we would obtain a $V^{[U,k]}$-proof r of φ. Clearly, for each τ_i we have that $\Box_V\tau_i$, so that by our assumption $k : U \rhd V$ we obtain a U proof q_i of τ_i^k. However, these proofs q_i may be cofinal and thus we would need a form of collection to exclude that possibility to keep the resulting syntactical object p' finite.

It turns out that we can perform a little trick to avoid the use of collection. To this end, let τ be the (possibly non-standard) conjunction of these axioms. Note that, by the naturality conditions on our coding, τ is bounded by p. Since clearly, we have $\Box_V\tau$, we may find, using $k : U \rhd V$, a U-proof q of τ^k (recall that we employ theorems interpretability in this chapter). We may use q to obtain U-proofs of q_i of ${\tau_i}^k$. Clearly, $|q_i|$ is bounded by a term of order $|q|^2$. We can now replace every axiom occurrence of τ_i in p by

$$\frac{\tau_i \wedge (\underline{q_i} = \underline{q_i})}{\tau_i}\ \wedge E, l$$

and obtain a $V^{[U,k]}$-proof r of φ. We find that $|r|$ is bounded by a term of order $|p| \cdot |q|^2$. So r can indeed be found in p-time from the given p and q. ⊣

For the previous lemma to hold it is essential that we work with efficient numerals $\underline{p}$. The reader may find it instructive to rephrase the lemma in terms of provability.

Corollary 7.7 For U and V, Δ_1^b-axiomatised theories we have

1. $\mathsf{S}^1_2 \vdash \forall k\, \forall \varphi\, (\Box_{V^{[U,k]}}\varphi \to \Box_V \varphi)$.

2. $\mathsf{S}^1_2 \vdash \forall k\, \Big(k : U \rhd V \to \forall \varphi\, (\Box_{V^{[U,k]}}\varphi \leftrightarrow \Box_V \varphi)\Big)$.

As mentioned before, even though we have extensional equivalence of V and $V^{[U,k]}$ under the assumption that $k : U \rhd V$, we do not necessarily have this under a provability predicate. That is, although we do have $\Box_{\mathsf{S}^1_2}(\Box_{V^{[U,k]}}\varphi \to \Box_V\varphi)$ we shall, in general, not have $k : U \rhd V \to \Box_{\mathsf{S}^1_2}(\Box_V\varphi \to \Box_{V^{[U,k]}}\varphi)$.

7.2.2 A P-like principle for the approximated theory

The theory $V^{[U,k]}$ is exactly defined so that it being interpretable in U is true almost by definition. This is even independent on k being or not an interpretation of V in U. The following lemma reflects this insight.

Lemma 7.8 For U and V, Δ^b_1-axiomatised theories we have $\mathsf{S}^1_2 \vdash \forall k\, (k : U \rhd V^{[U,k]})$.

Proof. Reason in S^1_2. Suppose p is a $V^{[U,k]}$-proof of ϕ. We want to construct a U-proof of ϕ^k. As a first step we transform p into a V-proof p' as we did in the proof of Lemma 7.6,(1). Next we transform p', using k, into a predicate logical proof q of ϕ^k from assumptions τ_i^k, where each τ_i is a V-axiom. It is well known that this transformation is p-time. Finally, each axiom τ_i extracted from p, comes from a $V^{[U,k]}$-axiom $\tau_i \wedge (\underline{r}_i = \underline{r}_i)$, where r_i is a U-proof of τ_i^k. So our final step is to extend q to a U-proof q' by prepending the U-proofs r_i above the corresponding τ_i^k. This extension will at most double the number of symbols of q, so $q' \approx q^2$. $\dashv$

As a direct consequence of this lemma, we see via necessitation that $\mathsf{S}^1_2 \vdash \Box_{\mathsf{S}^1_2}\forall k\, (k : U \rhd V^{[U,k]})$ so that in a trivial way we obtain something that comes quite close to the P-schema:

$$\mathsf{S}^1_2 \vdash U \rhd V \to \Box_{\mathsf{S}^1_2}\forall k\, (k : U \rhd V^{[U,k]}). \tag{7.2}$$

This is not yet what we are looking for. Therefore, we go to the setting of interpretability logics where all theories that we consider come as sentential extensions of some base theory. In this context we can prove the following lemma.

Lemma 7.9 Let T be a Δ^b_1-axiomatised theory, containing S^1_2 (as always), and let α and β be T-sentences.

$$\mathsf{S}^1_2 \vdash k : (T + \alpha) \rhd (T + \beta) \to \Box_{\mathsf{S}^1_2} k : (T + \alpha) \rhd (T^{[T+\alpha,k]} + \beta)$$

Proof. We reason in S^1_2 and suppose

$$k : (T + \alpha) \rhd (T + \beta). \tag{7.3}$$

In particular, $k : (T + \alpha) \rhd T$ whence by (7.2) we conclude

$$\Box_{\mathsf{S}^1_2} k : (T + \alpha) \rhd T^{[T+\alpha,k]}. \tag{7.4}$$

We need to conclude $\Box_{\mathsf{S}^1_2} k : (T + \alpha) \rhd (T^{[T+\alpha,k]} + \beta)$. In other words, we need to show

$$\Box_{\mathsf{S}^1_2}\Big(\forall\varphi\,(\Box_{T^{[T+\alpha,k]}+\beta}\varphi \;\to\; \Box_{T+\alpha}\varphi^k)\Big). \tag{7.5}$$

By (7.3) we know that $\Box_{T+\alpha}\beta^k$ so that by provable Σ_1-completeness we also have $\Box_{\mathsf{S}^1_2}\Box_{T+\alpha}\beta^k$. Thus, under the $\Box_{\mathsf{S}^1_2}$ we may use $\Box_{T+\alpha}\beta^k$. With this, we can now set out to prove (7.5). Thus we reason under the $\Box_{\mathsf{S}^1_2}$, fix some arbitrary φ and assume $\Box_{T^{[T+\alpha,k]}+\beta}\varphi$. By the formalised deduction theorem we get $\Box_{T^{[T+\alpha,k]}}(\beta \to \varphi)$. Thus, via (7.4) and by observing that $(\beta \to \varphi)^k = (\beta^k \to \varphi^k)$ we obtain $\Box_{T+\alpha}(\beta^k \to \varphi^k)$. This is combined with $\Box_{T+\alpha}\beta^k$ to obtain $\Box_{T+\alpha}\varphi^k$ which was needed for (7.5) so that the proof is finished. $\dashv$

This lemma can be simplified slightly which shall be the final version of our approximation of the principle P.

Theorem 7.10 Let T be a Δ^b_1-axiomatised theory, containing S^1_2, and let α and β be T-sentences.

$$\mathsf{S}^1_2 \vdash k : (T + \alpha) \rhd (T + \beta) \to \Box_{\mathsf{S}^1_2} k : (T + \alpha) \rhd (T^{[T,k]} + \beta)$$

Proof. Similar to the proof of Lemma 7.9. We now start by applying (7.2) to $U{=}V{=}T$ to conclude $\Box_{\mathsf{S}^1_2} k : T \rhd T^{[T,k]}$ so in particular $\Box_{\mathsf{S}^1_2} k : (T + \alpha) \rhd T^{[T,k]}$. The remainder of the proof is unaltered. $\dashv$

We observe that, *a priori*, Lemma 7.9 is slightly stronger than Theorem 7.10 since $T^{[T+\alpha,k]}$ can contain more axioms than $T^{[T,k]}$. However, for our purposes we did not see a need for the additional freedom generated by using $T^{[T+\alpha,k]}$ instead of $T^{[T,k]}$.

7.2.3 Iterated approximations

As it turns out, we will need to apply our technique of approximating theories to theories that themselves are already approximations[4]. To this end we generalise the definition of approximated theories to sequences of interpretations as follows.

Definition 7.11 Let $V^{[\langle U,k\rangle]} := V^{[U,k]}$. We recursively define

$$V^{[\langle U_0,k_0\rangle,\ldots,\langle U_n,k_n\rangle,\langle U_{n+1},k_{n+1}\rangle]}$$

[4] An example can be found in the proof of Lemma 7.30.

for $n \geq 0$ to stand for $\left(V^{[\langle U_0,k_0\rangle,\ldots,\langle U_n,k_n\rangle]}\right)^{[U_{n+1},k_{n+1}]}$, i.e.:

$$\begin{aligned}\mathsf{axioms}_{V^{[\langle U_0,k_0\rangle,\ldots,\langle U_n,k_n\rangle,\langle U_{n+1},k_{n+1}\rangle]}}(x) \;:=&\\ \exists\, p,\varphi{<}x\, \Big(x = \ulcorner \varphi \wedge (\underline{p} = \underline{p})\urcorner \wedge&\\ &\mathsf{axioms}_{V^{[\langle U_0,k_0\rangle,\ldots,\langle U_n,k_n\rangle]}}(\varphi) \wedge\\ &\mathsf{proof}_{U_{n+1}}(p, \varphi^{k_{n+1}})\Big).\end{aligned}$$

If x denotes a finite sequence $\langle U_0, k_0\rangle, \ldots, \langle U_n, k_n\rangle$, then we understand $V^{[x,\langle U_{n+1},k_{n+1}\rangle]}$ as $V^{[\langle U_0,k_0\rangle,\ldots,\langle U_n,k_n\rangle,\langle U_{n+1},k_{n+1}\rangle]}$.

Lemma 7.9 can be adapted to this new setting so that we get the following.

Lemma 7.12 Let T be a Δ_1^b-axiomatised theory, containing S_2^1, and let α and β be T-sentences. Let x be a sequence of pairs $\langle U_i, k_i\rangle$. We have:

$$\mathsf{S}_2^1 \vdash k : (T+\alpha) \rhd (T^{[x]} + \beta) \rightarrow \Box_{\mathsf{S}_2^1}(T+\alpha \rhd T^{[x,\langle T+\alpha,k\rangle]} + \beta).$$

Proof. Reason in S_2^1 and suppose $k : (T+\alpha) \rhd (T^{[x]}+\beta)$. Via a straightforward and minor modification of Lemma 7.9 we get

$$\Box_{\mathsf{S}_2^1} k : (T+\alpha) \rhd \left(\left(T^{[x]}\right)^{[T+\alpha,k]} + \beta\right),$$

i.e., $\Box_{\mathsf{S}_2^1} k : (T+\alpha) \rhd (T^{[x,\langle T+\alpha,k\rangle]} + \beta)$ as was to be shown. ⊣

Again, it seems that there is no need to keep track of the α formulas in the $T^{[T+\alpha,k]}$ definition. Therefore, we shall in the sequel work with simply sequences of interpretations of T in T rather than sequences of pairs of theory and interpretation. The corresponding definition is as follows where $\langle\rangle$ denotes the empty sequence and for a sequence x, we use $x \star k$ or sometimes simply x, k to denote the concatenation of x with $\langle k\rangle$.

Definition 7.13 For T a Δ_1^b-axiomatised theory we define $T^{[\langle\rangle]} := T$ and $T^{[x\star k]} := \left(T^{[x]}\right)^{[T,k]}$.

From now on, we shall write $T^{[k]}$ instead of $T^{[\langle k\rangle]}$. With the simplified notion of iteration we can formulate a friendlier P-flavoured principle.

Theorem 7.14 Let T be a Δ_1^b-axiomatised theory, containing S_2^1, and let α and β be T-sentences. Let x be a sequence of interpretations. We have:

$$\mathsf{S}_2^1 \vdash k : (T+\alpha) \rhd (T^{[x]} + \beta) \rightarrow \Box_{\mathsf{S}_2^1} k : (T+\alpha) \rhd (T^{[x,k]} + \beta).$$

7.3 A modal logic for approximation

In this section we will present a modal logical system to reason about interpretations and approximations based on them.

7.3.1 The logic AtL

We proceed to articulate modal principles reflecting facts about approximations. The main idea is to label our modalities with sequences x of interpretation variables. Of course, in the arithmetical part, these sequences x will indeed be interpreted via some map κ as a sequence $\kappa(x)$ of translations from the language of T to the language of T. In the next subsection we shall make the arithmetical reading precise but the idea is that $A \rhd^x B$ will stand for $T + \alpha \rhd T^{[\kappa(x)]} + \beta$ whenever A is interpreted by the arithmetical sentence α and B by β. Likewise, $\Box^x A$ will be interpreted as $\Box_{T^{[\kappa(x)]}}\alpha$.

As in [40], we will call our modal system AtL. We first specify the language. We have propositional variables $p_0, p_1, p_2 \ldots$ We will use $p, q, r, \ldots$ to range over them, and we have interpretation variables $k_0, k_1, k_2, \ldots$. We have one interpretation constant id. The meta-variables $k, \ell, m, \ldots$ will range over the interpretation terms (i.e. interpretation variables and id). The meta-variables $x, y, z, \ldots$ will range over finite sequences of interpretation variables. The modal language is the smallest language containing the propositional variables, closed under the propositional connectives, including $\top$ and $\bot$, and given an interpretation term k, the modal operators $\Box^k$ and $\rhd^k$ (we let $\Diamond^x A$ abbreviate $\neg\Box^x\neg A$), and closed under the following rule.

- If $A \rhd^x B$ is in the language and k is an interpretation term not contained in x, then $A \rhd^{x,k} B$ is in the language. Similarly for $\Box^x A$.

We will write $\rhd$ for $\rhd^{\mathsf{id}}$, and analogously for $\Box$ and $\Diamond$. Our logic AtLwill have axioms $\vdash A$ for any tautology A. Moreover, AtL has the obvious interchange rules to govern interaction between both sides of the turnstyle $\vdash$ based on the deduction theorem so that $\Delta, \Gamma \vdash C \ \Leftrightarrow \Delta \vdash \bigwedge \Gamma \to C$. Apart from modus ponens, AtL has the following axioms and rules.

$(\to \Box)^{x,k}$	$\vdash \Box^{x,k} A \to \Box^x A$
$(\to \rhd)^{x,k}$	$\vdash A \rhd^x B \to A \rhd^{x,k} B$
L_1^x	$\vdash \Box^x (A \to B) \to (\Box^x A \to \Box^x B)$
$\mathsf{L}_2^{x,y}$	$\vdash \Box^y A \to \Box^x \Box^y A$
L_3^x	$\vdash \Box^x (\Box^x A \to A) \to \Box^x A$
J_1^x	$\vdash \Box^x (A \to B) \to A \rhd^x B$
$\mathsf{J}_2^x\mathsf{a}$	$\vdash (A \rhd B) \wedge (B \rhd^x C) \to A \rhd^x C$
$\mathsf{J}_2^{x,y}\mathsf{b}$	$\vdash (A \rhd^x B) \wedge \Box^{x \star y} (B \to C) \to A \rhd^x C$
J_3^x	$\vdash (A \rhd^x C) \wedge (B \rhd^x C) \to A \vee B \rhd^x C$
J_4^x	$\vdash A \rhd^x B \to (\Diamond A \to \Diamond^x B)$
$\mathsf{J}_5^{x,y}$	$\vdash A \rhd^x \Diamond^y B \to A \rhd^y B$
$\mathsf{P}^{x,y,k}$	$\Gamma, \Delta, \Box^y (A \rhd^{x,k} B) \vdash C \Rightarrow \Gamma, A \rhd^x B \vdash C$
Nec^x	$\vdash A \Rightarrow \vdash \Box^x A$

In the above, the rule $\mathsf{P}^{x,y,k}$ is subject to the following conditions:

1. k is an interpretation variable;
2. k does not occur in x, Γ, A, B, C;
3. Δ consists of formulas of the form $E \rhd^{x,k} F \to E \rhd^x F$ and $\Box^x E \to \Box^{x,k} E$.

We will call the licence to use $\Box^x E \to \Box^{x,k} E$ provided by $\mathsf{P}^{x,y,k}$: $(\mathsf{E}\Box)^{\mathsf{k}}$, and we will call the licence to use $E \rhd^{x,k} F \to E \rhd^x F$: $(\mathsf{E}\rhd)^{\mathsf{k}}$.

We observe that by taking the empty sequence we get various special cases of our axioms. For example, a special case of $(\to \Box)^{x,k}$ would be $\vdash \Box^k A \to \Box A$. Furthermore, successive applications of $(\to \Box)^{x,k}$ yield $\vdash \Box^x A \to \Box A$

Likewise, a special case of $(\to \rhd)^{x,k}$ gives us $\vdash A \rhd B \to A \rhd^k B$. A special case of $\mathsf{J}_2^{x,y}\mathsf{b}$ is given by what we could call $\mathsf{J}_2^x\mathsf{b}$ given by $(A \rhd^x B) \wedge \Box^x (B \to C) \to A \rhd^x C$ by taking y to be the empty sequence in $\mathsf{J}_2^{x,y}\mathsf{b}$. In our applications so far, we only saw need for the special case of the axiom. It is unknown if AtL with $\mathsf{J}_2^y\mathsf{b}$ instead of $\mathsf{J}_2^{x,y}\mathsf{b}$ proves the same set of interpretability-variable-free formulas (and likewise for the other special cases).

Furthermore, we observe that J_1^x follows from the classical J_1 principle since

$$\begin{array}{rcl} \Box^x (A \to B) & \to & \Box (A \to B) \\ & \to & A \rhd B \\ & \to & A \rhd^x B. \end{array}$$

As a first and simple derivation in our system we have the following strengthening of the principle P_0 in AtL (recall that P_0 is the scheme $A \rhd \Diamond B \to \Box (A \rhd B)$).

Lemma 7.15 Let x be a finite non-empty sequence of interpretation variables. We may fix y so that $x = y\star\langle k\rangle$ with the understanding that $y = \mathsf{id}$ in case the length of x is just one. Now, let Δ consists of formulas of the form $E \rhd^{y,k} F \to E \rhd^{y} F$ and $\Box^{y}E \to \Box^{y,k}E$. We have the following rule to be derivable over AtL:

$$\Gamma, \Delta, \Box(A \rhd B) \vdash C \Rightarrow \Gamma, A \rhd^{x} \Diamond B \vdash C$$

Proof. This follows from the rule P^{x} combined with the schema J_5^{x}. ⊣

7.3.2 Quantifying over interpretations

In order to set up arithmetical semantics, we would like to quantify over sensible translations. However, how are we to separate the sensible from the non-sensical translations? In this subsection we shall provide a construction to guarantee that we only use sensible translations.

Being a sensible translation or not shall be gauged via some fixed interpretation of the natural numbers. Thus, let T be any theory with a designated interpretation, say N, of S^1_2. Let $\alpha^\star$ be a conjunction of T-axioms that implies $(\mathsf{S}^1_2)^N$. We fix T, N and $\alpha^\star$ for the remainder of this section.

We can now provide a construction that guarantees that we only deal with sensible translations. Define, for any translation k of the language of T to the language of T:

$$\mathsf{s}(k) := \begin{cases} k, & \text{if } (\alpha^\star)^k \\ \mathsf{id}, & \text{otherwise} \end{cases}$$

Thus, $\mathsf{s}(k)$ is an interpretation defined by cases. Via an easy induction on the complexity of φ we can prove that over predicate logic we have

$$\varphi^{\mathsf{s}(k)} \longleftrightarrow \Big((\alpha^\star)^k \wedge \varphi^k\Big) \vee \Big(\neg(\alpha^\star)^k \wedge \varphi\Big). \tag{7.6}$$

Since the needed induction to prove this is on the *length* of φ and since the proof can be uniformly constructed in p-time from φ, we have access to (7.6) when reasoning inside S^1_2.

We observe that $\mathsf{s}(k)$ should be evaluated where it occurs. For example in the formula $\exists k \Box_U \phi^{\mathsf{s}(k)}$, the choice of whether $\mathsf{s}(k)$ will be id or k will depend on whether $(\alpha^\star)^k$ holds under the $\Box_U$ even though the k comes from a quantifier outside the box. In contrast, in the expression $\exists k \Box_{U^{[\mathsf{s}(k)]}}\varphi$, the nature of $U^{[\mathsf{s}(k)]}$ depends on whether $(\alpha^\star)^k$ holds outside the box.

Let us proceed by making some easy observations on $\mathsf{s}(k)$. In the following lemma, we start by observing that regardless of the nature of k, the derived $\mathsf{s}(k)$ provides us an interpretation of $\alpha^\star$ in T. Next, we see that any other interpretation of $\alpha^\star$ in T will also occur as an image of s. Thus, modulo T-provable equivalence, $\mathsf{s}(k)$ ranges precisely over all interpretations of $\alpha^\star$.

Lemma 7.16 We have, verifiably in S^1_2, for any k,

1. $T \vdash (\alpha^\star)^{\mathsf{s}(k)}$,

2. for any formula ϕ we have $T \vdash (\alpha^\star)^k \to (\phi^{\mathsf{s}(k)} \leftrightarrow \phi^k)$,

Proof. Let us prove the first claim. Reason in S^1_2 and let k be arbitrary. Now reason in T or more formally, under the $\Box_T$. We distinguish cases. If $(\alpha^\star)^k$, then $\mathsf{s}(k) = k$, and $(\alpha^\star)^k$ holds by the case assumption. Otherwise, $\mathsf{s}(k) = \mathsf{id}$. The choice of T (see beginning of the subsection) implies $T \vdash \alpha^\star$, as required.

Similarly, to see the second claim we should show that either $(\alpha^\star)^k \to (\phi^k \leftrightarrow \phi^k)$ if $(\alpha^\star)^k$, or $(\alpha^\star)^k \to (\phi \leftrightarrow \phi^k)$ if $(\alpha^\star)^k$ does *not* hold. Both claims are obviously true under the respective assumptions. ⊣

In the light of Theorem 7.14 it is desirable that approximating theories contain S^1_2. The following lemma tells us that approximating theories indeed contain a sufficient amount of arithmetic. We recall that where the lemma mentions the theory of form $T^{[T,\mathsf{s}(k)]}$ we really mean the theory axiomatised by

$$\begin{aligned}\mathsf{axioms}_{T^{[T,\mathsf{s}(k)]}}(x) = \exists\, p, \varphi < x \Big(& x = \ulcorner \varphi \wedge (\underline{p} = \underline{p})\urcorner \wedge \\ & \mathsf{axioms}_T(\varphi) \wedge \mathsf{proof}_T(p, \varphi^{\mathsf{s}(k)})\Big).\end{aligned} \tag{7.7}$$

In this formula we can expand $\varphi^{\mathsf{s}(k)}$ as in (7.6).

Lemma 7.17 $\mathsf{S}^1_2 \vdash \forall k\, \Box_{T^{[T,\mathsf{s}(k)]}}(\mathsf{S}^1_2)^N$.

Proof. Reason in S^1_2. Consider any translation k from the language of T to the language of T. Lemma 7.16 tells us there is a proof in T of $(\alpha^\star)^{\mathsf{s}(k)}$. Hence, we have proofs p_i of $(\alpha_i)^{\mathsf{s}(k)}$, for (standardly) finitely many T-axioms $\alpha_1, \ldots, \alpha_n$. We would like to show that $T^{[T,\mathsf{s}(k)]}$ proves each of these α_i, since then $T^{[T,\mathsf{s}(k)]}$ proves $(\mathsf{S}^1_2)^N$. We take arbitrary α_i and put $x = \ulcorner \alpha_i \wedge (\underline{p_i} = \underline{p_i})\urcorner$. If we now substitute this x into (7.7), we immediately get the first two conjuncts of (7.7). Furthermore, T proves $(\alpha_i)^{\mathsf{s}(k)}$ because p_i is a proof of this formula in T. So, $\alpha_i \wedge (\underline{p_i} = \underline{p_i})$ is an axiom of $T^{[T,\mathsf{s}(k)]}$, whence $T^{[T,\mathsf{s}(k)]}$ proves α_i. ⊣

Recall that we work with the theories T that interpret S^1_2 and that we fix a designated interpretation $N : T \rhd \mathsf{S}^1_2$. We defined a variety of other theories of the form $T^{[x]}$, but we did not specify what interpretation of S^1_2 we are supposed to bundle them with. The preceding lemma tells us that we can reuse N. Thus, we will take N as the designated interpretation of S^1_2 in the $T^{[x]}$.

7.3.3 Arithmetical soundness

As usual, the modal logics are related to arithmetic via so-called *realisations*. Realisations map the propositional variables to sentences in the language of arithmetic. However, we now also have to deal with the interpretation sequences. Thus, our realisations for the arithmetical interpretation are pairs (σ, κ), where:

- σ maps the propositional variables to T-sentences, and
- κ maps the interpretation variables to translations from the language of T to the language of T.

We stipulate that the σ are $\top$ for all but finitely many arguments and likewise, we stipulate that the κ are id for all but finitely many arguments. The realisations are lifted to the arithmetical language in the obvious way by having them commute with the logical connectives and by taking:

$$(\Box^{k_1,\ldots,k_n} A)^{\sigma,\kappa} := \Box_{T[\,\langle \mathsf{s}(\kappa(k_1)),\ldots,\mathsf{s}(\kappa(k_n))\rangle\,]} A^{\sigma,\kappa}, \text{ and}$$

$$(A \rhd^{k_1,\ldots,k_n} B)^{\sigma,\kappa} := \\ (T + A^{\sigma,\kappa}) \rhd (T^{[\,\langle \mathsf{s}(\kappa(k_1)),\ldots,\mathsf{s}(\kappa(k_n))\rangle\,]} + B^{\sigma,\kappa}).$$

We observe that the nested modalities make sense because of Lemma 7.17. Note that the interpretation $\mathsf{s}(k)$ is applied only at locations where it can be expanded to a formula. Thus, we can arithmetise its use in T. For this reason, we can internally quantify over interpretations and the statement of the following theorem makes sense.

Theorem 7.18 Let T be a Δ^b_1-axiomatisable theory containing S^1_2. Furthermore, let $\forall^{T\text{-realisation}}$ flag that we quantify over realisations that are lifted to the entire modal language using formalised provability and interpretability over the base theory T. We then have

$$\Gamma \vdash_{\mathsf{AtL}} A \quad \Longrightarrow \quad \forall^{T\text{-realisation}}\sigma \;\; \mathsf{S}^1_2 \vdash \forall\kappa \, (\bigwedge \Gamma^{\sigma,\kappa} \to A^{\sigma,\kappa}).$$

Proof. Via an easy induction on AtLproofs. We refer the reader to [40] for details. An important ingredient is given in Theorem 7.14. ⊣

The use of a P-flavoured rule instead of an axiom is suggested since it better allocates flexibility in collecting all applications of Lemma 7.6 and Corollary 7.7 in our reasoning. To be on the safe side, we consider that $\mathsf{AtL}$is presented using multi-sets so that we can allocate of applications of Lemma 7.6 and Corollary 7.7 *after* a $\mathsf{P}^{x,y,k}$ rule is applied.

7.3.4 Room for generalisations

We already observed that our modal system does not directly allocate the extra flexibility that Lemma 7.12 has over Theorem 7.14. If we would like our logics to reflect the extra flexibility, we could work with sequences of pairs of formulas and translations instead of just sequences of translations. These formulas can then be added to the base theory. Similar, but even more general, is the following notion where assignments for the arithmetical interpretation are triples (σ, κ, τ), where:

- σ maps the propositional variables to T-sentences,
- κ maps the interpretation variables to translations from the language of T to the language of T, and
- τ maps the interpretation variables to theories in the language of T.

As before, we stipulate that the σ are $\top$ for all but finitely many arguments; the κ are id for all but finitely many arguments; and the τ are T for all but finitely many arguments. The assignments are lifted to the arithmetical language in the obvious way as before, but now taking:

$$
\begin{aligned}
&(\Box^{k_1,\ldots,k_n} A)^{\sigma,\kappa,\tau} := \Box_{T[\,\langle \tau(k_1),\mathsf{s}(\kappa(k_1))\rangle,\ldots,\langle \tau(k_n),\mathsf{s}(\kappa(k_n))\rangle\,]} A^{\sigma,\kappa,\tau},\\
&(A \rhd^{k_1,\ldots,k_n} B)^{\sigma,\kappa,\tau} :=\\
&\qquad (T + A^{\sigma,\kappa,\tau}) \rhd (T^{[\,\langle \tau(k_1),\mathsf{s}(\kappa(k_1))\rangle,\ldots,\langle \tau(k_n),\mathsf{s}(\kappa(k_n))\rangle\,]} + B^{\sigma,\kappa,\tau}).
\end{aligned}
$$

This notion gives rise to the following notion of consequence and soundness of $\mathsf{AtL}$with respect to this notion of consequence is readily proven.

$$\Gamma \models_T A :\Leftrightarrow \forall \sigma \ \ \mathsf{S}^1_2 \vdash \forall \kappa, \tau \left(\bigwedge \Gamma^{\sigma,\kappa,\tau} \to A^{\sigma,\kappa,\tau}\right).$$

Another way of possible generalisation is given by approximating both the interpreted *and the interpreting* theory. We observe that currently we only approximate the interpreted theory. Alternatively, we could label the binary modality $\rhd$ by a pair of sequences x, y of translations with the intended reading of $A \rhd^{x,y} B$ being $T^{[\mathsf{s}(x)]} + \alpha \rhd T^{[\mathsf{s}(y)]} + \beta$ when α and β are the intended readings of A and B respectively. Such a generalisation would allow for a more sophisticated transitivity axiom:

$$(A \rhd^{x,y} B) \wedge (B \rhd^{y,z} C) \to (A \rhd^{x,z} C).$$

We leave these observations for future investigations.

7.3.5 An alternative system

The idea of approximating finite axiomatisability can be realised in a different way. Let $A \rhd^0 B$ stand for $T + \alpha \rhd \mathsf{S}^1_2 + \beta$ when α and β are the arithmetical interpretation of A and B respectively. Likewise, $\Box^0 A$ will stand for $\Box_{\mathsf{S}^1_2}\alpha$. Using this notation, we can formulate a new sound principle.

Lemma 7.19 Given an interpretation k, we have

$$\vdash k : A \rhd B \to \Box^0(k : A \rhd^0 B).$$

Proof. Assume $k : A \rhd B$. Clearly $k : A \rhd^0 B$, by the assumption that all our base theories T extend S^1_2. Since $\mathsf{S}^1_2 + B$ is finitely axiomatisable, $k : A \rhd^0 B$ is a $\exists\Sigma^b_1$-statement. By $\exists\Sigma^b_1$-completeness we get the required $\Box^0(k : A \rhd^0 B)$. ⊣

Lemma 7.20

$$\vdash A \rhd \Diamond B \to \Box^0(A \rhd B).$$

Proof. Using Lemma 7.19 and noticing that from $k : A \rhd^0 \Diamond B$ we can obtain $A \rhd B$ just like with the rule $\mathsf{J}^{x,y}_5$ of AtL. ⊣

The relation between the logic AtL, i.e. reasoning with iterated approximations, and reasoning with $\rhd^0$ and non-iterated approximations, is unknown. In particular, we do not know if both systems prove the same theorems in their common language or in the language without any interpretability variable at all. However, we do observe that both systems are sufficiently strong for the principles appearing in this chapter.

7.4 On principles in IL(All)

In this section, we give arithmetical soundness proofs for some well-known interpretability principles that hold in all reasonable arithmetical theories. For this purpose we will employ the system AtL.

To avoid repeating too much content from [40], here we study only the following principles, but with proofs written in more detail compared to [40]. For other well-known principles please refer to [40].

$$\begin{array}{ll} \mathsf{W} & \vdash A \rhd B \to A \rhd (B \land \Box\neg A) \\ \mathsf{M}_0 & \vdash A \rhd B \to (\Diamond A \land \Box C) \rhd (B \land \Box C) \\ \mathsf{R} & \vdash A \rhd B \to \neg(A \rhd \neg C) \rhd B \land \Box C \end{array}$$

7.4.1 The principle W

We start with the **IL**P-proof of the principle W, which we will later convert to an AtL proof of W.

Fact 7.21 **IL**$\mathsf{P} \vdash \mathsf{W}$.

Proof. We reason in **IL**P. Suppose $A \rhd B$. Then, $\Box(A \rhd B)$. Hence, $(*)$ $\Box(\Diamond A \to \Diamond B)$, and, thus, $(**)$ $\Box(\Box\neg B \to \Box\neg A)$.

Moreover, from $A \rhd B$, we have $A \rhd (B \wedge \Box\neg A) \vee (B \wedge \Diamond A)$. So it is sufficient to show: $B \wedge \Diamond A \rhd B \wedge \Box\neg A$.

We have:

$$\begin{array}{llll} B \wedge \Diamond A & \rhd & \Diamond B & \text{by } (*) \\ & \rhd & \Diamond(B \wedge \Box\neg B) & \text{by } \mathsf{L}_3 \\ & \rhd & B \wedge \Box\neg B & \text{by } \mathsf{J}_5 \\ & \rhd & B \wedge \Box\neg A. & \text{by } (**) \end{array}$$

⊣

To prove arithmetical soundness of W we will essentially replicate the modal proof of W in **IL**P. We first give a more formal version of the proof that uses the rule $\mathsf{P}^{x,y,k}$ in the way we formally defined it. Afterwards we will give a more natural proof.

Lemma 7.22 The following holds:

$$\begin{array}{ll} & \Box(A \rhd^{[k]} B), \\ & (B \wedge \Diamond A \rhd^{[k]} B \wedge \Box^{[k]}\neg B) \to (B \wedge \Diamond A \rhd B \wedge \Box^{[k]}\neg B) \\ \vdash_{\mathsf{AtL}} & B \wedge \Diamond A \rhd B \wedge \Box\neg A. \end{array}$$

Proof. Reason in AtL. Some simple uses of rules and axiom schemas of AtL are left implicit.

$$\begin{array}{lrr} \Box(A \rhd^{[k]} B) & \text{assump.} & (7.8) \\ (B\wedge\Diamond A\rhd^{[k]}B\wedge\Box^{[k]}\neg B)\to(B\wedge\Diamond A\rhd B\wedge\Box^{[k]}\neg B) & \text{assump.} & (7.9) \\ \Box(\Diamond A \to \Diamond^{[k]} B) & \text{by } (7.8), \mathsf{J}_4^k & (7.10) \\ \Box(\Box^{[k]}\neg B \to \Box\neg A) & \text{by } (7.10) & (7.11) \\ \Diamond A \rhd \Diamond^{[k]} B & \text{by } (7.10), \mathsf{J}_1 & (7.12) \\ B \wedge \Diamond A \rhd \Diamond^{[k]} B & \text{by } (7.12), \mathsf{J}_1, \mathsf{J}_2 & (7.13) \\ B \wedge \Diamond A \rhd \Diamond^{[k]}(B \wedge \Box^k\neg B) & \text{by } (7.13), \mathsf{L}_3^k, \mathsf{J}_2^k & (7.14) \\ B \wedge \Diamond A \rhd^{[k]} B \wedge \Box^k\neg B & \text{by } (7.14), \mathsf{J}_5^k & (7.15) \\ B \wedge \Diamond A \rhd B \wedge \Box^{[k]}\neg B & \text{by } (7.9), (7.15) & (7.16) \\ B \wedge \Diamond A \rhd B \wedge \Box\neg A. & \text{by } (7.11), (7.16) & (7.17) \end{array}$$

⊣

Proposition 7.23 The principle W is arithmetically valid.

$$\mathsf{AtL} \vdash A \rhd B \to A \rhd B \wedge \Box\neg A.$$

Proof. Reason in AtL. By P^k and Lemma 7.22 we get

$$A \rhd B \vdash_{\mathsf{AtL}} B \wedge \Diamond A \rhd B \wedge \Box\neg A. \ (*)$$

Now assume $A \rhd B$. Combining $A \rhd B$ with $(*)$ we get

$$B \wedge \Diamond A \rhd B \wedge \Box\neg A. \ (**)$$

Clearly $A \rhd B$ implies

$$A \rhd (B \wedge \Box\neg A) \vee (B \wedge \Diamond A). \ (***)$$

From $(**)$ and $(***)$ by J_3 we obtain $A \rhd B \wedge \Box\neg A$. Thus

$$\mathsf{AtL} \vdash A \rhd B \to A \rhd B \wedge \Box\neg A,$$

as required. ⊣

The proof presented in Proposition 7.23 (and Lemma 7.22) resembles the proof we gave earlier demonstrating that $\mathbf{IL}\mathsf{P} \vdash \mathsf{W}$. However, the resemblance is not exactly obvious; we had to turn our proof "inside-out" in order to use the rule P^k (resulting in the contrived statement of Lemma 7.22). This can be avoided by applying the rule P^k in a different way. When we want to conclude something starting from $A \rhd^x B$, we introduce a fresh interpretation variable k and get $\Box^y(A \rhd^{x,k} B)$ (for whichever y we find suitable). Now we have to be a bit more careful; we can't end the proof before we eliminate this k. We also have to be careful in how we use the rules $(\mathsf{E}\Box)^{\mathsf{k}}$ and $(\mathsf{E}\rhd)^{\mathsf{k}}$. Essentially, any proof in the new form must be formalisable in the system AtL as it was defined earlier. Let us demonstrate this with the principle W.

Reason in AtL. Suppose that $A \rhd B$. By P^k we have that, for some k, $\Box(A \rhd^{[k]} B)$. Hence, by J_4^k, we have $(*)$ $\Box(\Diamond A \to \Diamond^{[k]} B)$ and, so, $(**)$ $\Box(\Box^{[k]}\neg B \to \Box\neg A)$.

Moreover, from $A \rhd B$, we have $A \rhd (B \wedge \Box\neg A) \vee (B \wedge \Diamond A)$. So it is sufficient to show $B \wedge \Diamond A \rhd B \wedge \Box\neg A$. We have:

$$\begin{array}{lcll} B \wedge \Diamond A & \rhd & \Diamond^{[k]} B & \text{by } (*) \\ & \rhd & \Diamond^{[k]}(B \wedge \Box^{[k]}\neg B) & \text{by } \mathsf{L}_3^k \\ & \rhd & B \wedge \Box^{[k]}\neg B & \text{by } \mathsf{J}_5^k \text{ and } (\mathsf{E}\rhd)^{\mathsf{k}} \\ & \rhd & B \wedge \Box\neg A. & \text{by } (**) \ . \end{array}$$

7.4.2 The principle $\mathsf{M_0}$

Another good test case is the principle $\mathsf{M_0}$, since both $\mathbf{IL}\mathsf{W} \nvdash \mathsf{M_0}$ and $\mathbf{IL}\mathsf{M_0} \nvdash \mathsf{W}$. Although we will later demonstrate the method for the principle R too and $\mathbf{IL}\mathsf{R} \vdash \mathsf{M_0}$, the proof for R is more complex. For this reason we include the principle $\mathsf{M_0}$.

We start with the $\mathbf{IL}\mathsf{P}$-proof of $\mathsf{M_0}$.

Fact 7.24 $\mathbf{IL}\mathsf{P} \vdash \mathsf{M_0}$.

Proof. Reason in $\mathbf{IL}\mathsf{P}$.

$$\begin{array}{rcll}
A \rhd B & \to & \Box(A \rhd B) & \text{by } \mathsf{P} \\
 & \to & \Box(\Diamond A \to \Diamond B) & \text{by } \mathsf{J_4} \\
 & \to & \Box(\Diamond A \wedge \Box C \to \Diamond B \wedge \Box C) & \\
 & \to & \Diamond A \wedge \Box C \rhd \Diamond B \wedge \Box C & \text{by } \mathsf{J_1} \\
 & \to & \Diamond A \wedge \Box C \rhd \Diamond(B \wedge \Box C) & \\
 & \to & \Diamond A \wedge \Box C \rhd B \wedge \Box C. & \text{by } \mathsf{J_5}
\end{array}$$

⊣

Now we adapt this proof to fit AtL. We will not write the more formal version of the proof (see the commentary in Subsection 7.4.1).

P-style soundness proof of $\mathsf{M_0}$ Reason in AtL.

$$\begin{array}{rcll}
A \rhd B & \to & \Box(A \rhd^{[k]} B) & \text{by } \mathsf{P}^k \\
 & \to & \Box(\Diamond A \to \Diamond^{[k]} B) & \text{by } \mathsf{J}_4^k \\
 & \to & \Box(\Diamond A \wedge \Box C \to \Diamond^{[k]} B \wedge \Box C) & \\
 & \to & \Diamond A \wedge \Box C \rhd \Diamond^{[k]} B \wedge \Box C & \text{by } \mathsf{J_1} \\
 & \to & \Diamond A \wedge \Box C \rhd \Diamond^{[k]} B \wedge \Box^{[k]} \Box C & \text{by } \mathsf{L}_2^k \\
 & \to & \Diamond A \wedge \Box C \rhd \Diamond^{[k]}(B \wedge \Box C) & \\
 & \to & \Diamond A \wedge \Box C \rhd^{[k]} B \wedge \Box C. & \text{by } \mathsf{J}_5^k \\
 & \to & \Diamond A \wedge \Box C \rhd B \wedge \Box C. & \text{by } (\mathsf{E}\rhd)^{\mathsf{k}}
\end{array}$$

7.4.3 The principle R

As the final example, we will prove that the principle R is arithmetically valid.

Before we see that $\mathbf{IL}\mathsf{P} \vdash \mathsf{R}$, we first prove an auxiliary lemma.

Lemma 7.25 $\mathbf{IL} \vdash \neg(A \rhd \neg C) \wedge (A \rhd B) \to \Diamond(B \wedge \Box C)$.

Proof. We prove the $\mathbf{IL}$-equivalent formula $(A \rhd B) \wedge \Box(B \to \Diamond\neg C) \to A \rhd \neg C$. But this is clear, as $\mathbf{IL} \vdash (A \rhd B) \wedge \Box(B \to \Diamond\neg C) \to A \rhd \Diamond\neg C$ and $\mathbf{IL} \vdash \Diamond\neg C \rhd \neg C$. ⊣

Fact 7.26 $\mathbf{IL}\mathsf{P} \vdash \mathsf{R}$.

Proof. We reason in $\mathbf{IL}\mathsf{P}$. Suppose $A \rhd B$. It follows that $\Box(A \rhd B)$. Using this together with Lemma 7.25 we get:

$$\begin{array}{rcll} \neg(A \rhd \neg C) & \rhd & \neg(A \rhd \neg C) \wedge (A \rhd B) & \\ & \rhd & \Diamond(B \wedge \Box C) & \text{by Lemma 7.25} \\ & \rhd & B \wedge \Box C. & \text{by } \mathsf{J}_5 \end{array}$$

⊣

P-style soundness proof of R Reason in AtL. We first show that $(A \rhd^{[k]} B) \wedge \neg(A \rhd \neg C) \to \Diamond^{[k]}(B \wedge \Box C)$. We show an equivalent claim $(A \rhd^{[k]} B) \wedge \Box^{[k]}(B \to \Diamond\neg C) \to A \rhd \neg C$.

Suppose that $A \rhd^{[k]} B$ and $\Box^{[k]}(B \to \Diamond\neg C)$. Thus, $A \rhd^{[k]} \Diamond\neg C$ by $\mathsf{J}_2^k\mathsf{b}$. By J_5^k we get $A \rhd \neg C$, as required. By necessitation,

$$\Box((A \rhd^{[k]} B) \wedge \neg(A \rhd \neg C) \to \Diamond^{[k]}(B \wedge \Box C)). \tag{7.18}$$

We now turn to the main proof. Suppose $A \rhd B$. Then, for some k, we have $\Box(A \rhd^{[k]} B)$ and, thus,

$$\begin{array}{rcll} \neg(A \rhd \neg C) & \rhd & \neg(A \rhd \neg C) \wedge (A \rhd^{[k]} B) & \\ & \rhd & \Diamond^{[k]}(B \wedge \Box C) & \text{by (7.18)} \\ & \rhd & B \wedge \Box C. & \text{by } \mathsf{J}_5^k \text{ and } (\mathsf{E}\rhd)^{\mathsf{k}} \end{array}$$

7.5 Two series of principles

In [29] two series of interpretability principles are presented. One series is called the *broad series*, denoted R^n for $n \in \omega$. The other series is called the *slim hierarchy*, denoted by R_n. The latter is actually a hierarchy of principles of increasing logical strength.

Both series of principles are proven to be arithmetically sound in any reasonable arithmetical theory. The methods used to prove this soundness in [29] involve definable cuts and in essence can be carried out in the system called CuL (we use this system in this thesis too, please see Chapter 8). In the next two sections we will see how both series admit a soundness proof based on the method of finite approximations of target theories as embodied in our logic AtL. We will also use this opportunity to state the results concerning modal semantics we obtained in collaboration with Jan Mas Rovira, which concern the two series. The proofs of these results can be found in his Master's thesis [44].

7.6 Arithmetical soundness of the slim hierarchy

As already mentioned, the slim series R_n defined in [29] is actually a hierarchy. Thus, to prove arithmetical soundness it suffices to study a cofinal sub-series. In our case we will study the certain sub-series $\widetilde{\mathsf{R}}_n$. Let us define the original sequence first; even though we will use the sub-series for the most part. Let a_i, b_i, c_i and e_i denote different propositional variables, for all $i \in \omega$.[5] We define a series of principles as follows.

$$
\begin{array}{lcl}
\mathsf{R}_0 & := & a_0 \rhd b_0 \to \neg(a_0 \rhd \neg c_0) \rhd b_0 \wedge \Box c_0 \\
\\
\mathsf{R}_{2n+1} & := & R_{2n}[\neg(a_n \rhd \neg c_n)/\neg(a_n \rhd \neg c_n) \wedge (e_{n+1} \rhd \Diamond a_{n+1}); \\
 & & b_n \wedge \Box c_n / b_n \wedge \Box c_n \wedge (e_{n+1} \rhd a_{n+1})] \\
\\
\mathsf{R}_{2n+2} & := & R_{2n+1}[b_n / b_n \wedge (a_{n+1} \rhd b_{n+1}); \\
 & & \Diamond a_{n+1} / \neg(a_{n+1} \rhd \neg c_{n+1}); \\
 & & (e_{n+1} \rhd a_{n+1})/(e_{n+1} \rhd a_{n+1}) \wedge (e_{n+1} \rhd b_{n+1} \wedge \Box c_{n+1})]
\end{array}
$$

We proceed with defining the sub-series $\widetilde{\mathsf{R}}_n$ (see [29], below Lemma 3.1):

$$
\begin{aligned}
\mathsf{X}_0 &:= A_0 \rhd B_0 \\
\mathsf{X}_{n+1} &:= A_{n+1} \rhd B_{n+1} \wedge (\mathsf{X}_n) \\
\mathsf{Y}_0 &:= \neg(A_0 \rhd \neg C_0) \\
\mathsf{Y}_{n+1} &:= \neg(A_{n+1} \rhd \neg C_{n+1}) \wedge (E_{n+1} \rhd \mathsf{Y}_n) \\
\mathsf{Z}_0 &:= B_0 \wedge \Box C_0 \\
\mathsf{Z}_{n+1} &:= B_{n+1} \wedge (\mathsf{X}_n) \wedge \Box C_{n+1} \wedge (E_{n+1} \rhd A_n) \wedge (E_{n+1} \rhd \mathsf{Z}_n) \\
\widetilde{\mathsf{R}}_n &:= \mathsf{X}_n \to \mathsf{Y}_n \rhd \mathsf{Z}_n.
\end{aligned}
$$

For convenience, define $\mathsf{X}_{-1} = \top$. With this we have $\mathsf{X}_n \equiv_{\mathbf{IL}} A_n \rhd B_n \wedge (\mathsf{X}_{n-1})$ for all $n \in \omega$. The first two schemas:

$$
\begin{array}{lcl}
\widetilde{\mathsf{R}}_0 & := & A_0 \rhd B_0 \to \neg(A_0 \rhd \neg C_0) \rhd B_0 \wedge \Box C_0; \\
\widetilde{\mathsf{R}}_1 & := & A_1 \rhd B_1 \wedge (A_0 \rhd B_0) \to \neg(A_1 \rhd \neg C_1) \wedge (E_1 \rhd \neg(A_0 \rhd \neg C_0)) \rhd \\
 & & \qquad B_1 \wedge (A_0 \rhd B_0) \wedge \Box C_1 \wedge (E_1 \rhd A_0) \wedge (E_1 \rhd B_0 \wedge \Box C_0).
\end{array}
$$

In the proof that $\mathsf{AtL} \vdash \widetilde{\mathsf{R}}_n$ (see the proof of Theorem 7.28) we use the following lemma.

[5] The series is originally defined as a series of modal formulas, not a series of *schemas* of modal formulas. In the remainder of the chapter we will work with schemas, as that is the way we treat other principles in this thesis.

Lemma 7.27 For all $n \in \omega$, and all interpretation variables k:

$$\mathsf{AtL} \vdash (A_n \rhd^k B_n \wedge (\mathsf{X}_{n-1})) \wedge \mathsf{Y}_n \to \Diamond^k(\mathsf{Z}_n).$$

Proof. Let $n = 0$ and fix k. We are to prove

$$\mathsf{AtL} \vdash (A_0 \rhd^k B_0 \wedge \top) \wedge \neg(A_0 \rhd \neg C_0) \to \Diamond^k(B_0 \wedge \Box C_0).$$

Equivalently,

$$\mathsf{AtL} \vdash (A_0 \rhd^k B_0) \wedge \Box^k(B_0 \to \Diamond\neg C_0) \to A_0 \rhd \neg C_0.$$

Assume $(A_0 \rhd^k B_0) \wedge \Box^k(B_0 \to \Diamond\neg C_0)$. By J_2^k, this yields $A_0 \rhd^k \Diamond\neg C_0$, whence by J_5^k, $A_0 \rhd \neg C_0$.

Let us now prove the claim for $n+1$. Fix k. Unpacking, we are to show that:

$$\begin{aligned}\mathsf{AtL} \vdash &(A_{n+1} \rhd^k B_{n+1} \wedge (\mathsf{X}_n)) \wedge \neg(A_{n+1} \rhd \neg C_{n+1}) \wedge (E_{n+1} \rhd \mathsf{Y}_n)\\ &\to \Diamond^k\Big(B_{n+1} \wedge (\mathsf{X}_n) \wedge \Box C_{n+1} \wedge (E_{n+1} \rhd A_n) \wedge (E_{n+1} \rhd \mathsf{Z}_n)\Big).\end{aligned}$$

Equivalently, we are to show that:

$$\begin{aligned}\mathsf{AtL} \vdash &(A_{n+1} \rhd^k B_{n+1} \wedge (\mathsf{X}_n)) \wedge (E_{n+1} \rhd \mathsf{Y}_n)\\ &\wedge \Box^k\Big((B_{n+1} \wedge (\mathsf{X}_n)) \to \Diamond\neg C_{n+1} \vee \neg(E_{n+1} \rhd A_n) \vee \neg(E_{n+1} \rhd \mathsf{Z}_n)\Big) \\ &\to A_{n+1} \rhd \neg C_{n+1}.\end{aligned} \tag{7.19}$$

Assume the conjunction on the left-hand side of (7.19). The first and the third conjunct imply

$$A_{n+1} \rhd^k B_{n+1} \wedge (\mathsf{X}_n) \wedge \Big(\Diamond\neg C_{n+1} \vee \neg(E_{n+1} \rhd A_n) \vee \neg(E_{n+1} \rhd \mathsf{Z}_n)\Big),$$

whence by weakening,

$$A_{n+1} \rhd^k (\mathsf{X}_n) \wedge \Big(\Diamond\neg C_{n+1} \vee \neg(E_{n+1} \rhd A_n) \vee \neg(E_{n+1} \rhd \mathsf{Z}_n)\Big). \tag{7.20}$$

We now aim to get $A_{n+1} \rhd^k \Diamond\neg C_{n+1}$. To this end, we set out to eliminate the last two disjuncts within (7.20).

From $E_{n+1} \rhd \mathsf{Y}_n$ (the second conjunct on the left-hand side of (7.19)) we have $E_{n+1} \rhd \neg(A_n \rhd \neg C_n)$, thus $E_{n+1} \rhd \Diamond A_n$, whence $\Box^k(E_{n+1} \rhd A_n)$ by the generalised P_0 (Lemma 7.15). We now combine $\Box^k(E_{n+1} \rhd A_n)$ with (7.20), simplify and weaken to obtain

$$A_{n+1} \rhd^k (\mathsf{X}_n) \wedge (\Diamond\neg C_{n+1} \vee \neg(E_{n+1} \rhd \mathsf{Z}_n)). \tag{7.21}$$

Thus, we have eliminated the second disjunct within (7.20), and it remains to eliminate $\neg(E_{n+1} \rhd \mathsf{Z}_n)$. We will now use the second conjunct on the left-hand side of (7.19), $E_{n+1} \rhd \mathsf{Y}_n$, again. We wish to apply the rule P^j_k, so assume $\Box^k(E_{n+1} \rhd^j \mathsf{Y}_n)$. Combining $\Box^k(E_{n+1} \rhd^j \mathsf{Y}_n)$ with (7.21) and unpacking X_n,

$$A_{n+1} \rhd^k (A_n \rhd B_n \wedge (\mathsf{X}_{n-1})) \wedge (E_{n+1} \rhd^j \mathsf{Y}_n) \wedge (\Diamond\neg C_{n+1} \vee \neg(E_{n+1} \rhd \mathsf{Z}_n)). \tag{7.22}$$

Reason under $\Box^k$. We wish to apply the rule P^ℓ_j with $A_n \rhd B_n \wedge (\mathsf{X}_{n-1})$, so assume $\Box^j(A_n \rhd^\ell B_n \wedge (\mathsf{X}_{n-1}))$. Combining $\Box^j(A_n \rhd^\ell B_n \wedge (\mathsf{X}_{n-1}))$ with $E_{n+1} \rhd^j \mathsf{Y}_n$ we obtain (still under the $\Box^k$) that $E_{n+1} \rhd^j (A_n \rhd^\ell B_n \wedge (\mathsf{X}_{n-1})) \wedge \mathsf{Y}_n$. Applying this to (7.22) we may conclude

$$A_{n+1} \rhd^k \Big(E_{n+1} \rhd^j (A_n \rhd^\ell B_n \wedge (\mathsf{X}_{n-1})) \wedge \mathsf{Y}_n\Big) \wedge (\Diamond\neg C_{n+1} \vee \neg(E_{n+1} \rhd \mathsf{Z}_n)).$$

The induction hypothesis allows us to replace $A_n \rhd^\ell B_n \wedge (\mathsf{X}_{n-1}) \wedge \mathsf{Y}_n$ with $\Diamond^\ell(\mathsf{Z}_n)$.

$$A_{n+1} \rhd^k (E_{n+1} \rhd^j \Diamond^\ell(\mathsf{Z}_n)) \wedge (\Diamond\neg C_{n+1} \vee \neg(E_{n+1} \rhd \mathsf{Z}_n)).$$

By $\mathsf{J}_5^{k,\ell}$,

$$A_{n+1} \rhd^k (E_{n+1} \rhd^\ell \mathsf{Z}_n) \wedge (\Diamond\neg C_{n+1} \vee \neg(E_{n+1} \rhd \mathsf{Z}_n)).$$

By our last application of P^ℓ_j and $(\mathsf{E}\rhd)^\ell$, we can substitute $\rhd$ for $\rhd^\ell$:

$$A_{n+1} \rhd^k (E_{n+1} \rhd \mathsf{Z}_n) \wedge (\Diamond\neg C_{n+1} \vee \neg(E_{n+1} \rhd \mathsf{Z}_n)).$$

Finally, we can simplify, weaken and apply J5 to obtain $A_{n+1} \rhd \neg C_{n+1}$. ⊣

We can now prove soundness for the slim hierarchy. It suffices to do this for the cofinal sub-hierarchy $\widetilde{\mathsf{R}}_n$.

Theorem 7.28 For all $n \in \omega$, $\mathsf{AtL} \vdash \widetilde{\mathsf{R}}_n$.

Proof. Let $n \in \omega$ be arbitrary. Assume $\Box^k(A_n \rhd B_n \wedge (\mathsf{X}_{n-1}))$. Clearly

$$\mathsf{Y}_n \rhd (A_n \rhd^k B_n \wedge (\mathsf{X}_{n-1})) \wedge \mathsf{Y}_n.$$

Now Lemma 7.27 implies

$$\mathsf{Y}_n \rhd \Diamond^k(\mathsf{Z}_n),$$

whence by the generalised J5,

$$\mathsf{Y}_n \rhd^k \mathsf{Z}_n.$$

By the rule P^k, we can replace our assumption $\Box^k(A_n \rhd B_n \wedge (\mathsf{X}_{n-1}))$ with X_n. Furthermore, by the same application of P^k, and by $(\mathsf{E}\rhd)^k$, we have $\mathsf{Y}_n \rhd \mathsf{Z}_n$. Thus, $X_n \to \mathsf{Y}_n \rhd \mathsf{Z}_n$, i.e.

$\widetilde{\mathsf{R}}_n$. ⊣

Finally, as we announced earlier, we quote the result obtained in collaboration with Jan Mas Rovira. To state the generalised frame condition for the principle R_1 (which lies strictly between $\widetilde{\mathsf{R}}_0$ and $\widetilde{\mathsf{R}}_1$) we let $R^{-1}[E] := \{x : (\exists y \in E)xRy\}$, and $R_x^{-1}[E] := R^{-1}[E] \cap R[x]$.

Theorem 7.29 The frame condition for the principle R_1 with respect to generalised Veltman semantics is the following condition:

$$\begin{aligned}\forall w, x, u, \mathbb{B}, \mathbb{C}, \mathbb{E}\Big(& wRxRuS_w\mathbb{B}, \mathbb{C} \in \mathcal{C}(x,u) \\ &\Rightarrow\ (\exists \mathbb{B}' \subseteq \mathbb{B})\Big(xS_w\mathbb{B}', R[\mathbb{B}'] \subseteq \mathbb{C}, (\forall v \in \mathbb{B}')(\forall c \in \mathbb{C}) \\ &\qquad (vRcS_xR_x^{-1}[\mathbb{E}] \Rightarrow (\exists \mathbb{E}' \subseteq \mathbb{E})cS_v\mathbb{E}')\Big)\Big).\end{aligned}$$

Proof. Please see [44] for the proof (including a formalisation in Agda). ⊣

7.7 Arithmetical soundness of the broad series

In order to define the second series we first define a series of auxiliary formulas. For any $n \geq 1$ we define the schemata U_n as follows.

$$\begin{aligned}\mathsf{U}_1 &:= \Diamond\neg(D_1 \rhd \neg C),\\ \mathsf{U}_{n+2} &:= \Diamond((D_{n+1} \rhd D_{n+2}) \wedge \mathsf{U}_{n+1}).\end{aligned}$$

Now, for $n \geq 0$ we define the schemata R^n as follows.

$$\begin{aligned}\mathsf{R}^0 &:= A \rhd B \to \neg(A \rhd \neg C) \rhd B \wedge \Box C,\\ \mathsf{R}^{n+1} &:= A \rhd B \to \Big(\mathsf{U}_{n+1} \wedge (D_{n+1} \rhd A)\Big) \rhd B \wedge \Box C.\end{aligned}$$

As an illustration we present the first three principles.

$$\begin{aligned}\mathsf{R}^0 &:= A \rhd B \to \neg(A \rhd \neg C) \rhd B \wedge \Box C;\\ \mathsf{R}^1 &:= A \rhd B \to \Diamond\neg(D_1 \rhd \neg C) \wedge (D_1 \rhd A) \rhd B \wedge \Box C;\\ \mathsf{R}^2 &:= A \rhd B \to \Diamond\Big[(D_1 \rhd D_2) \wedge \Diamond\neg(D_1 \rhd \neg C)\Big] \wedge (D_2 \rhd A) \rhd B \wedge \Box C.\end{aligned}$$

7.7.1 Proof using iterated approximations

When working with this series it is convenient to also have the following schemas:

$$\mathsf{V}_1 := \Box(D_1 \rhd \neg C),$$
$$\mathsf{V}_{n+1} := \Box(D_n \rhd D_{n+1} \to V_n) \text{ for } n \geq 1.$$

Alternatively, we could have defined $\mathsf{V}_n := \neg \mathsf{U}_n$ for $n \geq 1$.

Lemma 7.30 For all $n \in \omega \setminus \{0\}$, and all finite sequences x consisting of interpretation variables:

$$\mathsf{AtL} \vdash D_n \rhd^x \Diamond\neg C \to \mathsf{V}_n.$$

Proof. Let $n = 1$ and x be arbitrary. We are to prove $\mathsf{AtL} \vdash D_1 \rhd^x \Diamond\neg C \to \Box(D_1 \rhd \neg C)$. This is an instance of the generalised $\mathsf{P_0}$ schema as we stated in Lemma 7.15.

Let us now prove the claim for $n + 1$. Thus, we fix an arbitrary sequence of interpretations x. We are to show that

$$\mathsf{AtL} \vdash D_{n+1} \rhd^x \Diamond\neg C \to \Box(D_n \rhd D_{n+1} \to \mathsf{V}_n).$$

Thus, reasoning in AtL, we assume $D_{n+1} \rhd^x \Diamond\neg C$. We now wish to apply the rule P^k with this formula, where k is an arbitrary variable not used in x. So, assume $\Box(D_{n+1} \rhd^{x,k} \Diamond\neg C)$. Reason under a box. Assume $D_n \rhd D_{n+1}$. Now $D_n \rhd D_{n+1}$ and $D_{n+1} \rhd^{x,k} \Diamond\neg C$ imply $D_n \rhd^{x,k} \Diamond\neg C$. By the necessitated induction hypothesis, this implies V_n. Thus, $\Box(D_n \rhd D_{n+1} \to \mathsf{V}_n)$, as required. ⊣

Lemma 7.31 For all interpretation variables k we have the following:

$$\mathsf{AtL} \vdash \mathsf{U}_n \wedge (D_n \rhd A) \wedge (A \rhd^k B) \rhd^k B \wedge \Box C.$$

Proof. It is clear that the claim to be proved follows by necessitation, J_1, and $\mathsf{J}_5^{\mathsf{k}}$ from the following:

$$\mathsf{AtL} \vdash \mathsf{U}_n \wedge (D_n \rhd A) \wedge (A \rhd^k B) \to \Diamond^k(B \wedge \Box C).$$

This formula is equivalent to

$$(D_n \rhd A) \wedge (A \rhd^k B) \wedge \Box^k(B \to \Diamond\neg C) \to \mathsf{V}_n.$$

On the left-hand side we get $D_n \rhd^k \Diamond\neg C$. Now V_n follows from Lemma 7.30. ⊣

Theorem 7.32 For all $n \in \omega$, $\mathsf{AtL} \vdash \mathsf{R}^n$.

Proof. Case $n = 0$ is clear. Let $n > 0$ be arbitrary and let us prove R^n. Reason in AtL.

Assume $A \rhd B$. We wish to apply the rule P^k here. So, assume $\Box(A \rhd^k B)$. We have:

$$\mathsf{U}_n \wedge (D_n \rhd A) \rhd \mathsf{U}_n \wedge (D_n \rhd A) \wedge (A \rhd^k B).$$

Lemma 7.31 and the rule $\mathsf{J2}$ imply

$$\mathsf{U}_n \wedge (D_n \rhd A) \rhd^k B \wedge \Box C,$$

and by $(\mathsf{E}\rhd)^{\mathsf{k}}$,

$$\mathsf{U}_n \wedge (D_n \rhd A) \rhd B \wedge \Box C.$$

⊣

7.7.2 Proof using $\rhd^0$ (S^1_2)

Here we present an alternative proof which avoids iterated approximations, and instead uses the idea exploited in Lemma 7.19 and Lemma 7.20. The proof is essentially the same, but slightly shorter. We note here that we also wrote an alternative proof for the series $\mathsf{R_n}$ but we omit it in this thesis it as the proofs are very similar in that case too.

Lemma 7.33 For all $n \in \omega \setminus \{0\}$:

$$\mathsf{AtL} \vdash D_n \rhd^0 \Diamond\neg C \to \mathsf{V}_n.$$

Proof. Let $n = 1$. We are to prove $\mathsf{AtL} \vdash D_1 \rhd^0 \Diamond\neg C \to \Box(D_1 \rhd \neg C)$. This is an instance of the generalised $\mathsf{P_0}$ schema (Lemma 7.20).

Let us now prove the claim for $n+1$. We are to show that

$$\mathsf{AtL} \vdash D_{n+1} \rhd^0 \Diamond\neg C \to \Box(D_n \rhd D_{n+1} \to \mathsf{V}_n).$$

Assume $D_{n+1} \rhd^0 \Diamond\neg C$. By Lemma 7.19, we have $\Box^0(D_{n+1} \rhd^0 \Diamond\neg C)$. Reason under a box. Assume $D_n \rhd D_{n+1}$. Now $D_n \rhd D_{n+1}$ and $D_{n+1} \rhd^0 \Diamond\neg C$ imply $D_n \rhd^0 \Diamond\neg C$. By the induction hypothesis, this implies V_n, as required. ⊣

Lemma 7.34 Given an interpretation variable k,

$$\mathsf{AtL} \vdash \mathsf{U}_n \wedge (D_n \rhd A) \wedge (A \rhd^k B) \rhd^k B \wedge \Box C.$$

Proof. It is clear that the claim to be proved follows by necessitation, $\mathsf{J_1}$, and $\mathsf{J^k_5}$ from the following:

$$\mathsf{AtL} \vdash \mathsf{U}_n \wedge (D_n \rhd A) \wedge (A \rhd^k B) \to \Diamond^k(B \wedge \Box C).$$

This formula is equivalent to

$$(D_n \rhd A) \wedge (A \rhd^k B) \wedge \Box^k(B \to \Diamond\neg C) \to \mathsf{V}_n.$$

On the left-hand side we get $D_n \rhd^k \Diamond\neg C$. In particular, $D_n \rhd^0 \Diamond\neg C$. Now V_n follows from Lemma 7.33. ⊣

Theorem 7.35 For all $n \in \omega$, $\mathsf{AtL} \vdash \mathsf{R}^n$.

Proof. The proof is exactly the same as the proof of Theorem 7.32. ⊣

Finally, we state the generalised frame condition for the series $\mathsf{R^n}$, obtained in joint work with Jan Mas Rovira.

Theorem 7.36 Let $n \in \omega$ be arbitrary. We have $\mathfrak{F} \Vdash \mathsf{R}^n$ if and only if for all w, x_0, …, x_{n-1}, y, z, $\mathbb{A}$, $\mathbb{B}$, $\mathbb{C}$, $\mathbb{D}_0$, …, $\mathbb{D}_{n-1}$ we have the following:

$$\begin{aligned}
& wRx_{n-1}R\ldots Rx_0RyRz,\\
& (\forall u \in R[w] \cap \mathbb{A})(\exists V)uS_wV \subseteq \mathbb{B},\\
& (\forall u \in R[x_{n-1}] \cap \mathbb{D}_{n-1})(\exists V)uS_{x_{n-1}}V \subseteq \mathbb{A},\\
& (\forall i \in \{1,\ldots,n-2\})(\forall u \in R[x_i] \cap \mathbb{D}_i)(\exists V)uS_{x_i}V \subseteq \mathbb{D}_{i+1},\\
& (\forall V \in S_y[z])V \cap \mathbb{C} \neq 0,\\
& z \in \mathbb{D}_0\\
\Rightarrow\ & (\exists V \subseteq \mathbb{B})(x_{n-1}S_wV \ \&\ R[V] \subseteq \mathbb{C}).
\end{aligned}$$

Proof. Please see [44] for the proof (including a formalisation in Agda). ⊣

Chapter 8

An R^*-flavoured series of principles

In this chapter we further explore the question of the completeness of the logic $\mathbf{IL}\mathsf{WR}$ which we touched upon in Section 4.3 and in particular Subsection 4.3.1. We discussed the *label iteration problem* and described *labelling systems* which can be used to resolve the label iteration problem at least in the simpler case of $\mathbf{IL}\mathsf{WP}$-frames. We postponed the remainder of the discussion on $\mathbf{IL}\mathsf{WR}$ for this chapter.

The contents of this chapter is unpublished, with a preliminary report appearing as a short paper [46].

Introduction

In Subsection 4.3.1 we commented on the situation that may appear in a model we are building (which is meant to be an $\mathbf{IL}\mathsf{WR}$-model):

$$B \rhd C \in w \prec_S x \prec_T u \ni B.$$

In such situations we need to find a world v with

$$w \prec_{S \cup x^{\Box}_{T \cup \{\Box\neg B\}} \cup \{\Box\neg B\}} v \ni B.$$

However, using the principles W and R directly only gives us $w \prec_{S \cup x^{\Box}_{T} \cup \{\Box\neg B\}} v$ (we only get T where we'd like to have $T \cup \{\Box\neg B\}$). Similarly for longer sequences: suppose $B \rhd C \in w \prec_S x \prec_T y \prec_U u \ni B$. In this case we'd like to have

$$w \prec_{S \cup x^{\Box}_{T \cup \{\Box\neg B\} \cup y^{\Box}_{U \cup \{\Box\neg B\}}} \cup \{\Box\neg B\}} v \ni C$$

In order to prove the characteristic property we will actually require slightly more complex labels. We make the recursive labelling pattern for $\mathbf{IL}\mathsf{WR}$ explicit in Definition 8.1.

Definition 8.1 For $n \in \omega \setminus \{0\}$, let $\{w_0, \dots, w_n\}$ be a finite sequence of $\mathbf{IL}$WR-MCS's, let $\{S_1, \dots, S_n\}$ be a finite sequence of sets of formulas, and B a formula. We recursively define n sets of formulas, one for every $j \in \{0, \ldots, n-1\}$:

$$Q(\{w_0, \ldots, w_n\}, \{S_1, \ldots, S_n\}, B, j).$$

Usually the MCS's $\{w_0, \ldots, w_n\}$, and the sets of formulas $\{S_1, \ldots, S_n\}$ will be clear from the context, so we will write $Q_j(B)$ for $Q(\{w_0, \ldots, w_n\}, \{S_1, \ldots, S_n\}, B, j)$.

$$\begin{aligned} & Q_0(B) := \emptyset; \\ \text{for } j \in \{1, \ldots, n-1\}: \quad & Q_j(B) := {w_j}^{\Box}_{S_j \cup \{\Box\neg B\} \cup Q_{j-1}(B)} \cup Q_{j-1}(B). \end{aligned}$$

Given a sufficiently large sequence of worlds w_i and the corresponding labels S_i, we have:

$$\begin{aligned} Q_1(B) &= {w_1}^{\Box}_{S_1 \cup \{\Box\neg B\}}; \\ Q_2(B) &= {w_2}^{\Box}_{S_2 \cup \{\Box\neg B\} \cup {w_1}^{\Box}_{S_1 \cup \{\Box\neg B\}}} \cup {w_1}^{\Box}_{S_1 \cup \{\Box\neg B\}}; \\ Q_3(B) &= {w_3}^{\Box}_{S_3 \cup \{\Box\neg B\} \cup \left({w_2}^{\Box}_{S_2 \cup \{\Box\neg B\} \cup {w_1}^{\Box}_{S_1 \cup \{\Box\neg B\}}}\right) \cup {w_1}^{\Box}_{S_1 \cup \{\Box\neg B\}}} \cup {w_2}^{\Box}_{S_2 \cup \{\Box\neg B\} \cup {w_1}^{\Box}_{S_1 \cup \{\Box\neg B\}}} \cup {w_1}^{\Box}_{S_1 \cup \{\Box\neg B\}}. \end{aligned}$$

Given a set S and a formula B, we will often need to refer to the set $S \cup \{\Box\neg B\}$; for this reason we will sometimes denote $S \cup \{\Box\neg B\}$ as S^* (the formula B will usually be obvious from the context). Using this notation we have:

$$\begin{aligned} Q_1 &= {w_1}^{\Box}_{S_1^*}; \\ Q_2 &= {w_2}^{\Box}_{S_2^* \cup {w_1}^{\Box}_{S_1^*}} \cup {w_1}^{\Box}_{S_1^*}; \\ Q_3 &= {w_3}^{\Box}_{S_3^* \cup \left({w_2}^{\Box}_{S_2^* \cup {w_1}^{\Box}_{S_1^*}}\right) \cup {w_1}^{\Box}_{S_1^*}} \cup {w_2}^{\Box}_{S_2^* \cup {w_1}^{\Box}_{S_1^*}} \cup {w_1}^{\Box}_{S_1^*}. \end{aligned}$$

Considering what it takes to be able to claim the existence of worlds (i.e. maximal consistent sets w.r.t. **IL**WR) v with $w \prec_{S_n^* \cup Q_{n-1}(B)}$, which is a labelling that can be used to prove the characteristic property, we arrive at a new series of principles. If the principles in this series are provable in **IL**WR, this approach leads to a completeness proof of **IL**WR.

Let us lay down the plan for this chapter. In the next section we first define the new series of principles W_n for $n \in \omega$ and then we prove the "deficiency-solving" lemma, i.e. that the new series is sufficiently strong for obtaining the labels $Q_j(B)$ defined above.

One we defined the principles, we introduce **IL**W_ω-structures (where **IL**W_ω stands for the extension of **IL** with the set of principles $\{\mathsf{W}_n : n \in \omega\}$) and prove that if **IL**WR $\vdash \mathsf{W}_n$ (for all $n \in \omega$), then **IL**WR is complete w.r.t. generalised semantics. Also, if the principles W_n are valid in **IL**$_{\text{set}}$WR-frames, then **IL**W_ω is complete w.r.t. generalised semantics. Both claims we prove conditionally; we do not know if their antecedents are true.

Having discussed generalised semantics, we then turn to discuss ordinary Veltman semantics for $\mathbf{IL}\mathsf{W}_\omega$ and prove that the principles W_n are valid in ordinary $\mathbf{IL}\mathsf{WR}$-frames. If $\mathbf{IL}\mathsf{WR} \vdash \mathsf{W}_n$ (for all $n \in \omega$), then the principles W_n are valid in generalised $\mathbf{IL}\mathsf{WR}$-frames too. We believe to have determined that the principles W_1–W_4 are valid in $\mathbf{IL}_{\mathrm{set}}\mathsf{WR}$-frames.[1] The proof that the principle W_5 is valid in ordinary $\mathbf{IL}\mathsf{WR}$-frames amounts to showing that a certain model with about 500 worlds cannot exist. The principle W_5 is the point where our earlier strategies do not seem to be sufficiently strong to prove the validity w.r.t. the appropriate generalised frames. It is hard to analyse such large models (the generalised model whose existence we are to show to be impossible has at least as many worlds as the corresponding model in ordinary semantics, i.e. more than 500 worlds). We leave resolving this question for future work.

In the penultimate section we show that the new series is arithmetically valid. This is a prerequisite for even considering this series in the context of $\mathbf{IL}$(All); the section comes next-to-last for purely technical reasons: no other content depends on the results of this section, while that section depends on Section 8.1.

In the final section we give a recap of what we know regarding $\mathbf{IL}\mathsf{W}_\omega$, and what remains to be answered in future work.

8.1 The logic ILW_ω

Definition 8.2 We define a series of principles $(\mathsf{W}_n)_{n\in\omega}$.

$$
\begin{array}{rll}
 & \mathsf{U}_1 & := \bot; \\
\text{for } n > 1: & \mathsf{U}_n & := \Diamond C_{n-1} \vee \cdots \vee \Diamond C_1; \\
 & \mathsf{V}_1 & := A; \\
\text{for } n > 1: & \mathsf{V}_n & := \neg((C_{n-1} \rhd \Diamond A \vee B_{n-1} \vee \mathsf{U}_{n-1}) \to (\mathsf{V}_{n-1} \rhd B_{n-1}));
\end{array}
$$

$$
\text{for } n > 0: \quad \mathsf{W}_n := (A \rhd \Diamond A \vee B_n \vee \mathsf{U}_n) \to (\mathsf{V}_n \rhd B_n).
$$

Note that most parentheses in the definition are not required by our reading convention; they are added here solely for additional clarity.

We can let $\mathsf{W}_0 := \mathsf{W}$ so that W_n is defined for all $n \in \omega$. We could also shift the indices down, so that W_1 becomes W_0 etc., which we do not do in order to have the indices align better in some situations. We are really only interested in W_n for $n > 0$, and only treat that case in proofs.

[1]More precisely, a script employing a certain amount of semantic reasoning determined that an $\mathbf{IL}\mathsf{WR}$-counterexample cannot exist for W_1–W_4. The reasoning principles that this script employs haven't been formally verified and it is possible, though we believe it to be unlikely, that W_3 or W_4 are not valid in $\mathbf{IL}_{\mathrm{set}}\mathsf{WR}$-frames.

By $\mathbf{IL}\mathsf{W}_\omega$ we denote the extension of $\mathbf{IL}$ with the set of principles $\{\mathsf{W}_n : n \in \omega\}$. Similarly, by $\mathbf{IL}\mathsf{W}_\omega\mathsf{X}$ we understand the extension of $\mathbf{IL}\mathsf{W}_\omega$ by X.

Let us unravel the definition of W_n for some small values of n:

$$
\begin{aligned}
\mathsf{W}_1 :& A \rhd \Diamond A \vee B_1 \to A \rhd B_1;\\
\mathsf{W}_2 :& A \rhd \Diamond A \vee B_2 \vee \Diamond C_1 \to \neg(C_1 \rhd \Diamond A \vee B_1 \to A \rhd B_1) \rhd B_2;\\
\mathsf{W}_3 :& A \rhd \Diamond A \vee B_3 \vee \Diamond C_2 \vee \Diamond C_1 \to\\
& \neg(C_2 \rhd \Diamond A \vee B_2 \vee \Diamond C_1 \to \neg(C_1 \rhd \Diamond A \vee B_1 \to A \rhd B_1) \rhd B_2) \rhd B_3.
\end{aligned}
$$

The first two principles are well-known, although not in the form presented above. We show that $\mathbf{IL}\mathsf{W}_1 = \mathbf{IL}W$ and $\mathbf{IL}\mathsf{W}_1\mathsf{W}_2 = \mathbf{IL}\mathsf{R}^*$:

- $\mathbf{IL}\mathsf{W}_1 = \mathbf{IL}W$. Clearly $\mathbf{IL}W \vdash A \rhd \Diamond A \vee B_1 \to A \rhd (\Diamond A \vee B_1) \wedge \Box\neg A, \to A \rhd B_1$. On the other hand, $\mathbf{IL}\mathsf{W}_1 \vdash A \rhd B \to A \rhd B \wedge (\Box\neg A \vee \Diamond A), \to A \rhd (B \wedge \Box\neg A) \vee \Diamond A, \to A \rhd B \wedge \Box\neg A$.

- $\mathbf{IL}\mathsf{W}_1\mathsf{W}_2 = \mathbf{IL}\mathsf{R}^*$.

$$
\begin{aligned}
\mathbf{IL}\mathsf{R}^* &\vdash A \rhd \Diamond A \vee B_2 \vee \Diamond C_1 \to \neg(A \rhd C_1) \rhd B_2 \wedge \Box\neg C_1\\
\mathbf{IL}\mathsf{R}^* &\vdash A \rhd \Diamond A \vee B_2 \vee \Diamond C_1 \to \neg(A \rhd C_1) \rhd B_2\\
\mathbf{IL}\mathsf{R}^* &\vdash A \rhd \Diamond A \vee B_2 \vee \Diamond C_1 \to \neg(C_1 \rhd \Diamond A \vee B_1 \to A \rhd B_1) \rhd B_2
\end{aligned}
$$

To see the last inference, suppose $C_1 \rhd \Diamond A \vee B_1$, $\neg(A \rhd B_1)$ and (for a contradiction) $A \rhd C_1$. Then $A \rhd \Diamond A \vee B_1$, and thus by W ($\mathbf{IL}\mathsf{R}^* \vdash \mathsf{W}$), $A \rhd B_1$, a contradiction.

For the other direction,

$$
\begin{aligned}
\mathbf{IL} &\vdash A \rhd B \to A \rhd (B \wedge \Box\neg A \wedge \Box\neg C) \vee \Diamond A \vee \Diamond C\\
\mathbf{IL}\mathsf{W}_2 &\vdash A \rhd B \to \neg(C \rhd \Diamond A \vee C \to A \rhd C) \rhd B \wedge \Box\neg A \wedge \Box\neg C\\
\mathbf{IL}\mathsf{W}_2 &\vdash A \rhd B \to \neg(A \rhd C) \rhd B \wedge \Box\neg A \wedge \Box\neg C\\
\mathbf{IL}\mathsf{W}_2 &\vdash A \rhd B \to \neg(A \rhd C) \rhd B \wedge \Box\neg C
\end{aligned}
$$

For the second line take $B_2 := B$ and $C_1, B_1 := C$.

We note here that there is an easy way of forming W_{n+1} from W_n. Let OLD be the result of replacing the first occurrence of A in W_n (the occurrence of A at the very beginning of W_n) with C_n. We then have:

$$
\mathsf{W}_{n+1} = A \rhd \Diamond A \vee B_{n+1} \vee \Diamond C_n \vee \cdots \vee \Diamond C_1 \to \neg(\mathsf{OLD}) \rhd B_{n+1}.
$$

8.2 Generalised semantics

In this section we prove two conditional claims:

1. If $\mathbf{IL}\mathsf{WR} \vdash \mathsf{W}_n$ for all $n \in \omega$, then $\mathbf{IL}\mathsf{WR}$ is complete with respect to its generalised semantics.

2. If the generalised frame condition of $\mathbf{IL}\mathsf{W}_\omega$ is equivalent to $(\mathsf{WR})_{\text{gen}}$, then $\mathbf{IL}\mathsf{W}_\omega$ is complete w.r.t. the class of $\mathbf{IL}_{\text{set}}\mathsf{W}_\omega$-frames.

The structure of this proof resembles Section 4.3. Our first goal is to prove a deficiency-solving lemma for $\mathbf{IL}\mathsf{W}_\omega$. This is not as trivial as it was for other logics considered in this thesis. We first introduce a technical notion, CB-sequences, and prove they can be constructed. We will need CB-sequences in the deficiency-solving lemma; essentially, they will populate the labels used in that lemma with the correct choice of formulas.

Definition 8.3 Let X be an arbitrary collection of modal formulas. Let $n \in \omega \setminus \{0\}$. Let $\{w_0, \ldots, w_n\}$ be a finite sequence of $\mathbf{IL}\mathsf{W}_\omega\mathsf{X}$-MCS's, let $\{S_1, \ldots, S_n\}$ be a finite sequence of sets of formulas and let $E \rhd G$ a formula such that:

$$E \rhd G \in w_n \prec_{S_n} w_{n-1} \prec_{S_{n-1}} \cdots \prec_{S_1} w_0 \ni E.$$

We define a CB-sequence of length $m \in \{1, \ldots, n-1\}$ to be a pair of any finite sequences $\{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ and $\{\mathcal{B}_1, \ldots, \mathcal{B}_m\}$ of sets of formulas, and furthermore for all $i \in \{1, \ldots, m\}$:

$$\mathcal{B}_i \subseteq \neg S_i;$$
$$\bigvee \mathcal{C}_i \rhd \Diamond E \vee \bigvee \mathcal{B}_i \vee \bigvee_{1 \leq j \leq i-1} \Diamond \bigvee \mathcal{C}_j \in w_i.$$

Lemma 8.4 Let X be an arbitrary collection of modal formulas. Let $n \in \omega \setminus \{0, 1\}$. Let $\{w_0, \ldots, w_n\}$ be a finite sequence of $\mathbf{IL}\mathsf{W}_\omega\mathsf{X}$-MCS's, let $\{S_1, \ldots, S_n\}$ be a finite sequence of sets of formulas and $E \rhd G$ a formula such that:

$$E \rhd G \in w_n \prec_{S_n} w_{n-1} \prec_{S_{n-1}} \cdots \prec_{S_1} w_0 \ni E.$$

Let $m \in \{1, \ldots, n-1\}$ be arbitrary. If $\Box\neg D \in Q_m(E)$, then[2] there exist a CB-sequence $(\{\mathcal{C}_1, \ldots, \mathcal{C}_m\}, \{\mathcal{B}_1, \ldots, \mathcal{B}_m\})$ of length m with $D \in \bigcup_{i \in \{1, \ldots, m\}} \mathcal{C}_i$.

Proof. Let $n \in \omega \setminus \{0, 1\}$ be arbitrary. We prove the claim for this value of n and all values of $m \in \{1, \ldots, n-1\}$ by induction on m. Base case is $m = 1$. Since $\Box\neg D \in Q_1(E)$, for some finite J we have $D \rhd \bigvee_{j \in J} \neg S^1_j \vee \Diamond E \in w_1$ where $S^1_j \in S_1$ for all $j \in J$. We let $\mathcal{C}_1 = \{D\}$ and $\mathcal{B}_1 = \{\neg S^1_j : j \in J\}$. Clearly all required properties are satisfied.

[2]Note that all formulas in $Q_m(E)$ are of the form $\Box\neg D$.

Suppose the claim holds for all values less than some value m, with $2 \leq m \leq n-1$. We now prove the claim for m. Since $\Box\neg D \in Q_m(E)$ and $2 \leq m \leq n-1$, there are two cases to distinguish: (1) $\Box\neg D \in Q_{m-1}(E)$ and (2) $\Box\neg D \in w_m{}^{\Box}_{S_m \cup \{\Box\neg E\} \cup Q_{m-1}(E)}$.

In Case (1) we use the induction hypothesis for $m-1$ and obtain a CB-sequence $(\{\mathcal{C}_1, \ldots, \mathcal{C}_{m-1}\}, \{\mathcal{B}_1, \ldots, \mathcal{B}_{m-1}\})$ of length $m-1$ with $D \in \bigcup_{i \in \{1,\ldots,m-1\}} \mathcal{C}_i$. We extend this sequence by defining $\mathcal{B}_m = \emptyset$ and $\mathcal{C}_m = \emptyset$. Clearly the required properties of a CB-sequence are satisfied, and $D \in \bigcup_{i \in \{1,\ldots,m\}} \mathcal{C}_i$.

In Case (2), for some finite J and K we have $D \rhd \bigvee_{j\in J} \neg S^m_j \vee \Diamond E \vee \bigvee_{k \in K} \neg\Box\neg H_k \in w_m$ where $S^m_j \in S_m$ for all $j \in J$ and $\Box\neg H_k \in Q_{m-1}(E)$ for all $k \in K$. We define a CB-sequence of length m as follows. For every $k \in K$, since $\Box\neg H_k \in Q_{m-1}(E)$, the induction hypothesis implies there is a CB-sequence $(\{\mathcal{C}^k_1, \ldots, \mathcal{C}^k_{m-1}\}, \{\mathcal{B}^k_1, \ldots, \mathcal{B}^k_{m-1}\})$ of length $m-1$. Put $\mathcal{C}_m = \{D\}$ and $\mathcal{B}_m = \{\neg S^m_j : j \in J\}$. For i with $1 \leq i \leq m-1$, put $\mathcal{C}_i = \bigcup_{k\in K} \mathcal{C}^k_i$ and $\mathcal{B}_i = \bigcup_{k \in K} \mathcal{B}^k_i$. It is easy to see we have all the required properties. In particular, for the property that

$$\bigvee \mathcal{C}_m \rhd \Diamond E \vee \bigvee \mathcal{B}_m \vee \bigvee_{1 \leq j \leq m-1} \Diamond \bigvee \mathcal{C}_j \in w_m$$

note that for all $k \in K$, by the induction hypothesis, we have

$$H_k \in \bigcup_{1 \leq j \leq m-1} \mathcal{C}_j.$$

Thus, this property follows from

$$D \rhd \bigvee_{j \in J} \neg S^m_j \vee \Diamond E \vee \bigvee_{k\in K} \neg\Box\neg H_k \in w_m.$$

$\dashv$

We need the following lemma for the “deficiency-solving” business.

Lemma 8.5 Let X be an arbitrary collection of modal formulas. Let $n \in \omega \setminus \{0\}$. Let $\{w_0, \ldots, w_n\}$ be a finite sequence of $\mathbf{IL}\mathsf{W}_\omega\mathsf{X}$-MCS's, let $\{S_1, \ldots, S_n\}$ be a finite sequence of sets of formulas and let $E \rhd G$ be a formula such that:

$$E \rhd G \in w_n \prec_{S_n} w_{n-1} \prec_{S_{n-1}} \cdots \prec_{S_1} w_0 \ni E.$$

Then there is an $\mathbf{IL}\mathsf{W}_\omega\mathsf{X}$-MCS v such that $w_n \prec_{S^*_n \cup Q_{n-1}(E)} v$ and $G, \Box\neg G \in v$.

Proof. For $n = 1$ this is the labelling lemma for $\mathbf{IL}\mathsf{W}$ (see Lemma 4.26). Let $n > 1$ be arbitrary. Suppose there is no such consistent set v, i.e. the following set is inconsistent:

$$\{G, \Box\neg G, \neg A, \Box\neg A\}$$

where A is a formula such that for some finite sets J and K we have:

$$A \rhd \bigvee_{j\in J} \neg S^n_j \vee \Diamond E \vee \bigvee_{k\in K} \neg\Box\neg H_k \in w_n,$$

where $S^n_j \in S_n$ for all $j \in J$ and $\Box\neg H_k \in Q_{n-1}(E)$ for all $k \in K$. The inconsistency of the set $\{G, \Box\neg G, \neg A, \Box\neg A\}$ implies $\mathbf{IL} \vdash G \rhd A$, whence by $\mathsf{J2}$:

$$E \rhd \bigvee_{j\in J} \neg S^n_j \vee \Diamond E \vee \bigvee_{k\in K} \neg\Box\neg H_k \in w_n.$$

We define a CB-sequence of length $n-1$ as follows. For every $k \in K$, since $\Box\neg H_k \in Q_{n-1}(E)$, Lemma 8.4 implies there is a CB sequence $(\{\mathcal{C}^k_1, \dots, \mathcal{C}^k_{n-1}\}, \{\mathcal{B}^k_1, \dots, \mathcal{B}^k_{n-1}\})$ of length $n-1$. For i with $1 \leq i \leq n-1$, put $\mathcal{C}_i = \bigcup_{k\in K} \mathcal{C}^k_i$ and $\mathcal{B}_i = \bigcup_{k\in K} \mathcal{B}^k_i$.

For convenience, we also define $\mathcal{C}_n = \{E\}$ and $\mathcal{B}_n = \{\neg S^n_j : j \in J\}$.

Finally, we define for all $i \in \{1, \dots, n-1\}$:

$$B_i := \bigvee \mathcal{B}_i;$$
$$C_i := \bigvee \mathcal{C}_i.$$

We prove the following claim by induction: for all $i \in \{n, \dots, 2, 1\}$,

$$\mathsf{V}_i \rhd B_i \in w_i.$$

As the base case we take $i = n$. Unpacking the definitions of $\mathcal{C}_n$ and $\mathcal{B}_n$ and since for every $k \in K$ we have $H_k \in \bigcup_{1\leq j\leq n-1} \mathcal{C}_j$, we have: $E \rhd \Diamond E \vee B_n \vee \Diamond C_{n-1} \vee \dots \vee \Diamond C_1 \in w_n$. The principle W_n implies $\mathsf{V}_n \rhd B_n \in w_n$. Suppose the claim holds for $i+1$ and let us prove the claim for i. By the induction hypothesis and after unpacking $\mathsf{V}_{i+1} \rhd B_{i+1}$ we have

$$\neg(C_i \rhd \Diamond E \vee B_i \vee \mathsf{U}_i \to \mathsf{V}_i \rhd B_i) \rhd B_{i+1} \in w_{i+1}.$$

Since $w_{i+1} \prec_{S_{i+1}} w_i$ and using the property of a CB-sequence that $\mathcal{B}_{i+1} \subseteq \neg S_{i+1}$, we get

$$C_i \rhd \Diamond E \vee B_i \vee \mathsf{U}_i \to \mathsf{V}_i \rhd B_i \in w_i.$$

Again, since $(\{\mathcal{C}^k_1, \dots, \mathcal{C}^k_{n-1}\}, \{\mathcal{B}^k_1, \dots, \mathcal{B}^k_{n-1}\})$ is a CB-sequence, we have $\bigvee C_i \rhd \Diamond E \vee \bigvee B_i \vee \bigvee_{1\leq j\leq i-1} \Diamond C_j \in w_i$. Thus,

$$\mathsf{V}_i \rhd B_i \in w_i,$$

as required. This concludes the proof by induction.

In particular this implies that $\mathsf{V}_1 \rhd B_1 \in w_1$, i.e. $E \rhd B_1 \in w_1$. This is impossible since $B_1 = \bigvee \mathcal{B}_1$, $\mathcal{B}_1 \subseteq \neg S_1$ and $w_1 \prec_{S_1} w_0 \ni E$. $\dashv$

Now that we have our deficiency-solving lemma, we can proceed to the definition of an $\mathbf{IL}\mathsf{W}_\omega$-structure for a set of formulas $\mathcal{D}$. Here we assume that $\mathcal{D}$ has the same properties that were assumed in Chapter 4, i.e. it is a finite set closed under subformulas and single negations, and contains $\top$ (see Definition 4.5).

The $\mathbf{IL}\mathsf{W}_\omega$-structures will resemble $\mathbf{IL}\mathsf{WP}$-structures defined earlier (see Definition 4.35). As was the case with $\mathbf{IL}\mathsf{W}$-structures (Definition 4.27), we distinguish two cases in the definition of S_w. And since $\mathbf{IL}\mathsf{W}_\omega$ is affected by the problem of label iteration (see Section 4.3), we need to use a labelling system. The labelling system for $\mathbf{IL}\mathsf{W}_\omega$ is defined in Definition 8.1.

Definition 8.6 We say that $\mathfrak{M} = (W, R, \{S_w : w \in W\}, \Vdash)$ is the $\mathbf{IL}\mathsf{W}_\omega$-structure for a set of formulas $\mathcal{D}$ if:

- $W = \{w : w$ is an $\mathbf{IL}\mathsf{W}_\omega$-MCS and for some $B \in \mathcal{D},\ B \wedge \Box\neg B \in w\}$;
- $wRu \Leftrightarrow w \prec u$;
- $uS_wV \Leftrightarrow wRu$ and $V \subseteq R[w]$ and, moreover, one of the following holds:
  (a) $V \cap \dot{R}[u] \neq \emptyset$;
  (b) we have for all $n \in \omega \backslash \{0\},$ all $\{w_0, \ldots, w_n\},$ and all $\{S_1, \ldots, S_n\}$:
  $w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u \Rightarrow (\exists v \in V)(\exists B \in \mathcal{D} \cap \bigcup \dot{R}[u])\ w \prec_{S_n^* \cup Q_{n-1}(B)} v$;
- $w \Vdash p \Leftrightarrow p \in w$.

The following lemma is very similar to Lemma 4.36, and both share the same general format as Lemma 4.28. Essentially the only difference compared to Lemma 4.36 is a different labelling system, which affects the proof only slightly.

Lemma 8.7 The $\mathbf{IL}\mathsf{W}_\omega$-structure $\mathfrak{M}$ for $\mathcal{D}$ is a generalised Veltman model. Furthermore, the following holds for each $w \in W$ and $G \in \mathcal{D}$:

$$\mathfrak{M}, w \Vdash G \text{ if and only if } G \in w,$$

Proof. All the properties, except for quasi-transitivity, have easy proofs (see the proof of Lemma 4.36). Let us prove quasi-transitivity. Assume uS_wV, and vS_wU_v for all $v \in V$. Put $U = \bigcup_{v \in V} U_v$. We claim that uS_wU. Clearly $U \subseteq R[w]$. To prove uS_wU we will distinguish the cases (a) and (b) from the definition of the relation S_w for uS_wV.

In Case (a), there exists $v_0 \in V$ for some $v_0 \in \dot{R}[u]$. We will next distinguish two cases from the definition of $v_0 S_w U_{v_0}$.

In Case (aa) we have $x \in U_{v_0}$ for some $x \in \dot{R}[v_0]$. Since $v_0 \in \dot{R}[u]$, we then have $x \in \dot{R}[u]$. Since $x \in U_{v_0} \subseteq U$, then $U \cap \dot{R}[u] \neq \emptyset$. So, we have uS_wU, as required.

In Case (ab) we have:

$$\text{For all } n \in \omega\backslash\{0\}, \text{ all } \{w_0,\dots,w_n\}, \text{ and all } \{S_1,\dots,S_n\} \text{ we have:} \tag{8.1}$$
$$w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = v_0 \Rightarrow (\exists x \in U_{v_0})(\exists B \in \mathcal{D} \cap \bigcup \dot{R}[v_0])\ w \prec_{S_n^* \cup Q_{n-1}(B)} x.$$

To prove uS_wU in this case, we will use Case (b) from the definition of the relation S_w. Let $n \in \omega\backslash\{0\}$ be arbitrary and let $\{w_0,\dots,w_n\}$ and $\{S_1,\dots,S_n\}$ be arbitrary such that $w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u$. If $u = v_0$, applying (8.1) with n, the world $\{w_0,\dots,w_n\}$ and the labels $\{S_1,\dots,S_n\}$, produces the required $x \in U_{v_0} \subseteq U$ and $B \in \mathcal{D} \cap \bigcup \dot{R}[v_0]$. Otherwise, i.e. if uRv_0, let $w'_0 = v_0$, $w'_{i+1} = w_i$, $S'_1 = \emptyset$, $S'_{i+1} = S_i$ and apply (8.1) with $n+1$, $\{w'_0,\dots,w'_{n+1}\}$, and $\{S'_1,\dots,S'_{n+1}\}$. This gives us a world $x \in U_{v_0}$ and $B \in \mathcal{D} \cap \bigcup \dot{R}[v_0]$ with $w \prec_{S_n^* \cup Q'_n(B)} x$, where, for every j, the notation $Q'_j(B)$ is short for $Q(\{w'_0,\dots,w'_{n+1}\},\{S'_1,\dots,S'_{n+1}\},B,j)$. It is easy to prove by induction that after replacing occurrences of $u^{\Box}_{\{\Box\neg B\}}$ (i.e. $w'^{\Box}_{1\,S'_1\cup\{\Box\neg B\}}$) with $\emptyset$ in the recursive definition of $Q'_n(B)$ we obtain $Q_{n-1}(B)$ and furthermore that $Q_{n-1}(B) \subseteq Q'_n(B)$. Thus, $w \prec_{S_n^* \cup Q_{n-1}(B)} x$. Since uRv_0, we have $\dot{R}[v_0] \subseteq \dot{R}[u]$. Thus, we can reuse the formula B for this S_w transition.

In Case (b), we have:

$$\text{For all } n \in \omega\backslash\{0\}, \text{ all } \{w_0,\dots,w_n\}, \text{ and all } \{S_1,\dots,S_n\} \text{ we have:} \tag{8.2}$$
$$w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u \Rightarrow (\exists v \in V)(\exists B \in \mathcal{D} \cap \bigcup \dot{R}[u])\ w \prec_{S_n^* \cup Q_{n-1}(B)} v.$$

To prove uS_wU we will use Case (b) from the definition of the relation S_w.

Let $n \in \omega\backslash\{0\}$ be arbitrary and let $\{w_0,\dots,w_n\}$ and $\{S_1,\dots,S_n\}$ be arbitrary such that $w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u$. By (8.2), there are $v_0 \in V$ and $B \in \mathcal{D} \cap \bigcup \dot{R}[u]$ such that $w \prec_{S_n^* \cup Q_{n-1}(B)} v_0$. From $v_0 \in V$ it follows that $v_0 S_w U_{v_0}$. We will next distinguish the possible cases in the definition of $v_0 S_w U_{v_0}$.

In Case (ba) we have $U_{v_0} \cap \dot{R}[v_0] \neq \emptyset$, i.e. there is $x \in U_{v_0}$ such that either $v_0 = x$ or $v_0 R x$. In both cases we have $w \prec_{S_n^* \cup Q_{n-1}(B)} x$.

In Case (bb), we have:

$$\text{For all } n' \in \omega\backslash\{0\}, \text{ all } \{w'_0,\dots,w'_{n'}\}, \text{ and all } \{S'_1,\dots,S'_{n'}\} \text{ we have:} \tag{8.3}$$
$$w = w'_{n'} \prec_{S'_{n'}} \cdots \prec_{S'_1} w'_0 = v_0 \Rightarrow (\exists x \in U_{v_0})(\exists B' \in \mathcal{D} \cap \bigcup \dot{R}[v_0])\ w \prec_{S^*_{n'} \cup Q'_{n'-1}(B)} x,$$

where $Q'_j(B)$ is short for $Q(\{w'_0,\dots,w'_{n'}\},\{S'_1,\dots,S'_{n'}\},B,j)$.

At the moment we only need to use (8.3) with $n' = 1$, which is the following statement: for every T, if $w \prec_T v_0$, there is $x \in U_{v_0}$ and $B' \in \mathcal{D} \cap \bigcup \dot{R}[v_0]$ such that $w \prec_{T \cup \{\Box\neg B\}} x$. Using this and $w \prec_{S_n^* \cup Q_{n-1}(B)} v_0$ we get that there is some $x \in U_{v_0} \subseteq U$ and $B' \in \mathcal{D} \cap \bigcup \dot{R}[v_0]$ such that $w \prec_{S_n^* \cup Q_{n-1}(B) \cup \{\Box\neg B'\}} x$. By weakening, $w \prec_{S_n^* \cup Q_{n-1}(B)} x$, as required.

We claim that for each formula $G \in \mathcal{D}$ and each world $w \in W$ the following holds:

$$\mathfrak{M}, w \Vdash G \text{ if and only if } G \in w.$$

The proof is by induction on the complexity of G. The only non-trivial case is when $G = B \rhd C$.

Assume $B \rhd C \in w$, wRu and $u \Vdash B$. Induction hypothesis implies $B \in u$. We claim that $uS_w[C]_w$ by Case (b) from the definition of S_w. Clearly wRu and $[C]_w \subseteq R[w]$.

Fix $n \in \omega \setminus \{0\}$, $\{w_0, \ldots, w_n\}$ and $\{S_1, \ldots, S_n\}$. Assume $w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = u$. Since $B \rhd C \in w_n$ and $B \in w_0$, Lemma 8.7 implies that there is an $\mathbf{IL}\mathsf{W}_\omega$-MCS v with $w_n \prec_{S_n^* \cup Q_{n-1}(B)} v$ and $C, \Box\neg C \in v$ (thus $v \in W$). Since $C \in v$, the induction hypothesis implies $v \Vdash C$. Since $w \prec v$, i.e. wRv, then $v \in [C]_w$. Finally, $B \in \mathcal{D}$ and $B \in u$ imply $B \in \mathcal{D} \cap \bigcup \dot{R}[u]$.

To prove the converse, assume $B \rhd C \notin w$. Since w is an $\mathbf{IL}\mathsf{W}_\omega$-MCS, we have $\neg(B \rhd C) \in w$. Lemma 4.25 implies there is u with $w \prec_{\{\Box\neg B, \neg C\}} u$ and $B \in u$. Since $w \prec_{\{\Box\neg B\}} u$, we have in particular that $\Box\neg B \in u$. So, $u \in W$. The induction hypothesis implies $u \Vdash B$. Let $V \subseteq R[w]$ be such that uS_wV. We will find a world $v \in V$ such that $w \prec_{\{\neg C\}} v$. We will distinguish Cases (a) and (b) from the definition of the relation S_w. Consider Case (a). Let v be an arbitrary world in $V \cap \dot{R}[u]$. If $v = u$, clearly $w \prec_{\{\Box\neg B, \neg C\}} v$. If uRv, then we have $w \prec_{\{\Box\neg B, \neg C\}} u \prec v$. This implies $w \prec_{\{\Box\neg B, \neg C\}} v$. Consider Case (b). From $w \prec_{\{\Box\neg B, \neg C\}} u$ and the definition of S_w it follows that there is $v \in V$ such that (for some formula D) $w \prec_{\{\Box\neg B, \neg C, \Box\neg D\}} v$. In both cases we have $w \prec_{\{\neg C\}} v$; thus $C \notin v$. Induction hypothesis implies $v \nVdash C$; whence $V \nVdash C$, as required. $\dashv$

Theorem 8.8 The logic $\mathbf{IL}\mathsf{WR}$ is complete w.r.t. $\mathbf{IL}_{\text{set}}\mathsf{WR}$-frames if $\mathbf{IL}\mathsf{WR} \vdash \mathsf{W}_n$ for all $n \in \omega$.

Proof. In the light of Lemma 8.7 it suffices to show that the $\mathbf{IL}\mathsf{W}_\omega$-structure $\mathfrak{M}$ for $\mathcal{D}$ possesses the properties $(\mathsf{W})_{\text{gen}}$ and $(\mathsf{R})_{\text{gen}}$. The proof that $(\mathsf{W})_{\text{gen}}$ holds is very similar to the proof that $\mathbf{IL}$W-structures satisfy $(\mathsf{W})_{\text{gen}}$ (Theorem 4.29). Let us now prove that $\mathfrak{M}$ possesses the property $(\mathsf{R})_{\text{gen}}$.

Assume $wRxRuS_wV$ and $C \in \mathcal{C}(x, u)$. We are to show that there is $U \subseteq V$ with xS_wU and $R[U] \subseteq C$. To do this, we distinguish two possible cases for uS_wV. If uS_wV holds by Case (a) from the definition of S_w, there is $v \in V$ such that either $u = v$ or uRv. In both cases xRv. Let $U = \{v\}$. Clearly $U \subseteq V$. Since xRv, $xS_w\{v\}$, i.e. xS_wU. For any $z \in R[v]$ we have $xRuRz$ and thus $uS_x\{z\}$. This implies $z \in C$. Thus $R[U](= R[v]) \subseteq C$. The remainder of the proof deals with the case when uS_wV holds by Case (b) from the definition of S_w.

We will first prove an auxiliary claim:

for all $n \in \omega\backslash\{0\}$, all $\{w_0, \dots, w_n\}$, and all $\{S_1, \dots, S_n\}$ we have:
$w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = x \Rightarrow (\exists v \in V)(\exists B \in \mathcal{D} \cap \bigcup \dot{R}[u])\ w \prec_{S_n^* \cup Q_{n-1}(B) \cup x_\emptyset^\Box} v \ \&\ R[v] \subseteq C$.

Fix $n \in \omega \setminus \{0\}$, and sets $\{w_0, \dots, w_n\}$ and $\{S_1, \dots, S_n\}$ with $w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = x$. Suppose (for a contradiction) that for every $v \in V$ and $B \in \mathcal{D} \cap \bigcup \dot{R}[u]$ with $w \prec_{S_n^* \cup Q_{n-1}(B) \cup x_\emptyset^\Box} v$, we have $R[v] \nsubseteq C$, that is, there is some $z_{v,B} \in R[v] \setminus C$. Let

$$Z = \{z_{v,B} : v \in V, B \in \mathcal{D} \cap \bigcup \dot{R}[u], w \prec_{S_n^* \cup Q_{n-1}(B) \cup x_\emptyset^\Box} v\}.$$

We claim that uS_xZ. We have xRu by assumption. To see that $Z \subseteq R[x]$, take any $z_{v,B} \in Z$ and apply Lemma 4.2 and Lemma 4.21 to the fact that $w \prec_{x_\emptyset^\Box} v \prec z$. To complete the proof that uS_xZ, we will use Case (b) from the definition of S_x. This part of the proof will also imply $Z \neq \emptyset$.

Fix $n' \in \omega \setminus \{0\}$, $\{w'_0, \dots, w'_{n'}\}$ and $\{S'_1, \dots, S'_{n'}\}$. Assume $x = w'_{n'} \prec_{S'_{n'}} \cdots \prec_{S'_1} w'_0 = u$. We have:

$$w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = x = w'_{n'} \prec_{S'_{n'}} \cdots \prec_{S'_1} w'_0 = u.$$

For $i \in \{1, \dots, n\}$, let $w'_{n'+i} = w_i$ and $S'_{n'+i} = S_i$. Let $n'' := n' + n$.

Recall that uS_wV (i.e. $w'_0 S_{w'_{n''}} V$) holds by Case (b). We can apply this clause with n'', $\{w'_0, \dots, w'_{n''}\}$, and $\{S'_1, \dots, S'_{n''}\}$. So there is some $v \in V$ and $B \in \mathcal{D} \cap \bigcup \dot{R}[u]$ such that

$$w \prec_{(S'_{n''})^* \cup Q'_{n''-1}(B)} v,$$

where, for every j, the notation $Q'_j(B)$ is short for $Q(\{w'_0, \dots, w'_{n''}\}, \{S'_1, \dots, S'_{n''}\}, B, j)$.

Note that in particular, $w \prec_{S_n^* \cup Q_{n-1}(B) \cup x_\emptyset^\Box} v$. To see this, note that $(S'_{n''})^* = S_n^*$, $Q_{n-1}(B) \subseteq Q'_{n''-1}(B)$ (this can be proved inductively using the recursive definition of $Q'_{n''-1}(B)$), and

$$x_\emptyset^\Box = {w'_{n'}}_\emptyset^\Box \subseteq {w'_{n'}}^\Box_{S'_{n'} \cup \{\Box\neg B\} \cup Q'_{n'-1}(B)} \subseteq Q'_{n'}(B) \subseteq Q'_{n''-1}(B).$$

Thus, $z_{v,B}$ is well-defined. Note that a part of the proof just given is ${w'_{n'}}^\Box_{S'_{n'} \cup \{\Box\neg B\} \cup Q'_{n'-1}(B)} \subseteq Q'_{n''-1}(B)$. Thus, we have

$$w \prec_{{w'_{n'}}^\Box_{S'_{n'} \cup \{\Box\neg B\} \cup Q'_{n'-1}(B)}} v \prec z_{v,B}.$$

This and Lemma 4.21 imply (recall once more that $x = w'_{n'}$)

$$x \prec_{S'_{n'} \cup \{\Box\neg B\} \cup Q'_{n'-1}(B)} z_{v,B}.$$

This concludes the proof of uS_xZ.

Now, uS_xZ and $C \in \mathcal{C}(x, u)$ imply $C \cap Z \neq \emptyset$, contradicting the definition and the non-emptiness of Z. This concludes the proof of the auxiliary claim.

Let $U = \{v \in V : R[v] \subseteq C\}$. Auxiliary claim implies $U \neq \emptyset$. To prove xS_wU we will use Case (b) from the definition of S_w. Fix $n \in \omega \setminus \{0\}$, $\{w_0, \ldots, w_n\}$, and $\{S_1, \ldots, S_n\}$. Assume $w = w_n \prec_{S_n} \cdots \prec_{S_1} w_0 = x$. The auxiliary claim implies there is $B \in \mathcal{D} \cap \bigcup \dot{R}[u]$ and $v \in U$ such that $w \prec_{S_n^* \cup Q_{n-1}(B) \cup x_\emptyset^\Box} v$. Thus, $B \in \mathcal{D} \cap \bigcup \dot{R}[x]$ and $w \prec_{S_n^* \cup Q_{n-1}(B)} v$. This concludes the proof that xS_wU. It is clear that $R[U] \subseteq C$. $\dashv$

8.3 Ordinary semantics

Due to difficulties in determining whether the principles W_n are valid in $\mathbf{IL}_{\text{set}}\mathsf{WR}$-frames, we tried to obtain validity in regular $\mathbf{IL}\mathsf{WR}$-frames. In this section we prove that, indeed, the logic $\mathbf{IL}\mathsf{W}_\omega$ is valid in regular semantics. What this entails is that either $\mathbf{IL}\mathsf{WR}$ is incomplete w.r.t. $\mathbf{IL}_{\text{set}}\mathsf{WR}$-frames, or $\mathbf{IL}\mathsf{W}_\omega \vdash \mathbf{IL}\mathsf{WR}$.

Theorem 8.9 For $n \in \omega$, the principle W_n is valid in ordinary $\mathbf{IL}$WR-frames.

Proof. For $n = 1$ and $n = 2$ the claim follows from the fact that $\mathbf{IL}\mathsf{W}_1 = \mathbf{IL}W$ and $\mathbf{IL}\mathsf{W}_1\mathsf{W}_2 = \mathbf{IL}\mathsf{R}^*$.

Suppose the claim does not hold for some $n > 2$. Then there is an $\mathbf{IL}$WR-model $\mathfrak{M}$, and formulas $A, B_1, \ldots, B_n, C_1, \ldots, C_{n-1}$ such that there is a world in $\mathfrak{M}$, we will denote this world as v_n, such that the instance of W_n with these formulas is not satisfied in v_n. Note that the schema W_n does not contain C_n. For this reason we can use the symbol C_n to refer to the formula denoted by A. This will eliminate the need for some case distinctions.

Before proceeding with the proof, let us sketch the proof. First we see there is a certain R-sequence in our model. Then we see that there is a structure, which we will call an "L-sequence", appended to the end of the aforementioned R-sequence. Next, we will construct two functions, $i \mapsto \bar{i}$ and $i \mapsto \tilde{i}$, both mapping the set $\{1, \ldots, n\}$ to some finite set of natural numbers. Finally, using these mappings we will see that our L-sequence must be finite, yielding a contradiction.

In the first part of this proof we establish the existence of certain worlds $v_n, v_{n-1}, \ldots, v_0$ such that $v_n\ R\ v_{n-1}\ R \ldots R\ v_0$ and prove the property (8.7) below.

We will write S_i for S_{v_i}.

By unravelling the definition of W_n, we see that there exist worlds $v_{n-1}, \ldots, v_0$ such that $v_n \, R \, v_{n-1} \, R \ldots R \, v_0$ and:

$$v_n \Vdash C_n \rhd \Diamond C_n \vee B_n \vee \mathsf{U}_n; \tag{8.4}$$

$$v_i \Vdash \mathsf{C}_i \rhd \Diamond C_n \vee B_i \vee \mathsf{U}_i,$$

$$\text{and for all } S_{i+1}\text{-successors } x \text{ of } v_i,\ x \Vdash \neg B_{i+1} \text{ (for } 1 \leq i \leq n-1); \tag{8.5}$$

$$v_0 \Vdash C_n, \text{ and for all } S_1\text{-successors } x \text{ of } v_0,\ x \Vdash \neg B_1. \tag{8.6}$$

Thus (8.5) expands to:

- $v_{n-1} \Vdash \mathsf{C}_{n-1} \rhd \Diamond C_n \vee B_{n-1} \vee \mathsf{U}_{n-1}$, and for all S_n-successors x of v_{n-1}, $x \Vdash \neg B_n$;
- $v_{n-2} \Vdash \mathsf{C}_{n-2} \rhd \Diamond C_n \vee B_{n-2} \vee \mathsf{U}_{n-2}$, and for all S_{n-1}-successors x of v_{n-2}, $x \Vdash \neg B_{n-1}$;
- $\ldots$
- $v_2 \Vdash \mathsf{C}_2 \rhd \Diamond C_n \vee B_2 \vee \mathsf{U}_2$, and for all S_3-successors x of v_2, $x \Vdash \neg B_3$;
- $v_1 \Vdash \mathsf{C}_1 \rhd \Diamond C_n \vee B_1 \vee \mathsf{U}_1$, and for all S_2-successors x of v_1, $x \Vdash \neg B_2$.

Let

$$Y = \{y \in R[v_1] : y \Vdash C_n \text{ and } (\forall x \in S_1[y]) x \Vdash \neg B_1\}.$$

Thus, any $y \in Y$ satisfies the properties we require from the world v_0. W.l.o.g. we can assume that the world v_0 is an $(S_1 \circ R \circ S_1)$-maximal world in Y. Otherwise, if no world in Y is $(S_1 \circ R \circ S_1)$-maximal, the relation $S_1 \circ R \circ S_1$ is not converse well-founded, contradicting the property (W). Furthermore, note that

$$\text{if } v_0 \ (S_1 \circ R \circ S_1) \ z, \text{ then } z \nVdash C_n. \tag{8.7}$$

If this were not the case, the existence of such a world z would contradict the $(S_1 \circ R \circ S_1)$-maximality of the world v_0 within Y (we would have $v_1 R z$ and $z \Vdash C_n$, and, since $S_1[z] \subseteq S_1[v_0]$, for all S_1-successors x of z, $x \Vdash \neg B_1$, so $z \in Y$).

In the next portion of the proof we define a certain type of sequence, an "L-sequence", and prove there is an infinite such sequence.

Let us define the L-sequences. These are a specific kind of either a finite sequence $\{x_0, \ldots, x_m\}$ for some $m \in \omega$, or of an infinite sequence $(x_i)_{i \in \omega}$. We denote the set of indices $\{0, \ldots, m\}$ or ω as I. Furthermore, we require $x_0 = v_0$ and that the following properties are satisfied too:

1. for all $i \in I$ such that $i + 1 \in I$ we have $x_i (R \circ (S_1 \cup \cdots \cup S_n)) x_{i+1}$;
2. for all $i \in I$ we have $x_i \nVdash C_1$ and $x_i \Vdash C_j$ for some $j \in \{2, \ldots, n\}$;
3. if $i, i{+}1 \in I$ and $x_i (R \circ S_k) x_{i+1}$, then $x_i S_j x_{i+1}$ for all $j \in \{1, \ldots, k\}$;

4. if $i, i{+}1 \in I$ and $x_i(R \circ S_k)x_{i+1}$, and j is the minimal ℓ such that $x_i \Vdash C_\ell$, then $x_i(R \circ S_j)x_{i+1}$.

As a consequence of (3.), L-sequences will also have these properties:

5. for all $i \in I$ we have $v_0 S_1 x_i$;

6. for every world $w \in \{v_1, \dots, v_n\}$ and all $i \in I$ we have wRx_i;

Since $v_n\ R\ v_{n-1}\ R\ \dots\ R\ v_1$, the property (6) follows from the property (5). Furthermore, since the property (3.) implies that whenever we have $x_i(R{\circ}S_k)x_{i+1}$, we also have $x_i S_1 x_{i+1}$, the property (3.) implies the property (5).

First we check that the sequence which consists only of the world v_0 is a finite L-sequence. All properties have simple proofs, except for the property $v_0 \nVdash C_1$. Suppose for a contradiction that $v_0 \Vdash C_1$. Since $v_1\ R\ v_0 \Vdash C_1$ and $v_1 \Vdash C_1 \rhd \Diamond C_n \vee B_1$, we have $v_0 S_1 u \Vdash \Diamond C_n \vee B_1$ for some u. Since $v_0 S_1 u$, by (8.6) we have $u \Vdash \neg B_1$, thus $u \Vdash \Diamond C_n$. Then there is u' with $uRu' \Vdash C_n$. However, we noted earlier (8.7) that if $v_0(S_1 \circ R \circ S_1)z$, then $z \nVdash C_n$. So, it cannot be the case that $v_0 S_1 u R u' S_1 u' \Vdash C_n$, i.e. $v_0\ (S_1{\circ}R{\circ}S_1)\ u' \Vdash C_n$. Thus, $v_0 \nVdash C_1$.

We will prove that an infinite L-sequence exists by showing that any finite L-sequence can be extended to a longer finite L-sequence. Once we do that, the required infinite L-sequence can be obtained as the appropriate union of finite L-sequences.

So suppose we have an L-sequence $L_1 = \{x_0, \dots, x_k\}$ and we wish to construct an appropriate extension L_2 of L_1, i.e. we wish to define x_{k+1}.

By the property (2.) we know that $x_k \Vdash C_i$ for some $i \in \{2, \dots, n\}$. Pick the minimal such i. By the property (6) we have $v_i\ R\ x_k$. This, together with the fact that $v_i \Vdash \mathsf{C}_i \rhd \Diamond C_n \vee B_i \vee \mathsf{U}_i$, implies there must exist a world y with $x_k S_i y \Vdash \Diamond C_n \vee B_i \vee \mathsf{U}_i$. Using the property (6) once more, we have $v_i\ R\ v_{i-1}\ R\ x_k$. So, $v_{i-1}\ S_i\ x_k$, and thus $v_{i-1}\ S_i\ y$. Since any S_i-successor of v_{i-1} satisfies $\neg B_i$ (see (8.5)), we have $y \Vdash \neg B_i$. Thus, $y \Vdash \Diamond C_n \vee \mathsf{U}_i$, i.e. $y \Vdash \Diamond C_n \vee \Diamond C_{i-1} \vee \dots \vee \Diamond C_1$. So, there must exist z with $yRz \Vdash C_j$ for some $j \in \{1, 2, \dots, i-2, i-1, n\}$. So, $x_k(R \circ (S_1 \cup \dots \cup S_n))z$.

Let L_2 denote the extension of L_1 with z, i.e. $x_{k+1} := z$. We have to check if L_2 is an L-sequence too. As we noted above, we only need to check properties (1.) (which is immediate), (2.), (3.), and (4.). Furthermore, Property (2.) only needs to be verified for the new world z, while Properties (3.) and (4.) only need to be checked for the newly added part of the L-sequence, i.e. for $i = k$.

Let us check (4.) first. This is immediate since we defined i to be the minimal ℓ such that $x_k \Vdash C_\ell$.

Next we check (3.). Let $j \in \{1, \dots, i\}$ be arbitrary. Since $R \circ S_i \subseteq S_i$, the case $j = i$ is easy. The other case is $j < i$, i.e. $v_i\ R\ v_j$. We have $v_i\ R\ v_j\ R\, x_k\, S_i\, y\, R\, z$. The characteristic property (R) implies $x_k S_j z$.

It remains to check (2.). The proof for this property is very similar to the proof that $v_0 \nVdash C_1$ which we gave earlier. We already know that $z \Vdash C_j$ for some $j \in \{1, 2, \ldots, i-2, i-1, n\}$. Thus, it suffices to show $z \nVdash C_1$. Suppose not, i.e. $z \Vdash C_1$. Since $v_1 \; R \; z \Vdash C_1$ and $v_1 \Vdash C_1 \rhd \Diamond C_n \vee B_1$, we have $zS_1u \Vdash \Diamond C_n \vee B_1$ for some u. The property (5) (which follows from (3.), so we can use it) implies $v_0 S_1 z$. Since $v_0 S_1 z S_1 u$, we must have $u \Vdash \neg B_1$, thus $u \Vdash \Diamond C_n$. Then there is u' with $uRu' \Vdash C_n$. However, we noted earlier (8.7) that if $v_0(S_1 \circ R \circ S_1)u$, then $u \nVdash C_n$. So, it cannot be the case that $v_0 S_1 u R u' S_1 u' \Vdash C_n$, i.e. $v_0 \; (S_1 \circ R \circ S_1) \; u' \Vdash C_n$. Thus, $z \nVdash C_1$.

This concludes our proof that an infinite L-sequence exists.

In the final part of the proof we define certain indices $\bar{i}$ and $\tilde{i}$. These indices will be associated with world whose existence will enable us to point to the contradiction.

Fix an arbitrary infinite L-sequence $L_\infty = (x_i)_{i \in \omega}$. We will recursively define mappings $i \mapsto \bar{i}$ and $i \mapsto \tilde{i}$ with the domain $\{1, \ldots, n\}$ and the codomain ω.

Let $\bar{1} = 1$ and $\tilde{1} = 0$. Now suppose the indices $\bar{1}, \ldots, \overline{i-1}$ and the indices $\tilde{1}, \ldots, \widetilde{i-1} \in \omega$ have been selected. We are to define $\bar{i}$ and $\tilde{i}$. For at least one $j \in \{2, \ldots, n\}$ and some $k > \widetilde{i-1}$, the formula C_j is true in x_k (this follows from applying (2.) to indices in $I = \omega$ that are larger than $\widetilde{i-1}$). Fix the minimal j such that there exists $k > \widetilde{i-1}$ and the formula C_j is true in x_k; and define $\bar{i} = j$. The definition of $\bar{i}$ implies that the following set is non-empty:

$$T = \{x_k \in \omega : k > \widetilde{i-1} \text{ and } x_k \Vdash C_{\bar{i}}\}.$$

By the characteristic property (W) there is at least one $(S_{\bar{i}} \circ R \circ S_{\bar{i}})$-maximal world in T. Let $\tilde{i}$ equal the index ℓ such that the world x_ℓ is an $(S_{\bar{i}} \circ R \circ S_{\bar{i}})$-maximal world in T.

We will show the following properties hold:

a. $\bar{i} \in \{\overline{i-1} + 1, \ldots, n\}$ for all $i \in \{2, \ldots, n\}$;

b. if $\tilde{i} < k$, then $x_k \nVdash C_j$ for all $i \in \{1, \ldots, n\}$ and $j \in \{1, 2, \ldots, \bar{i} - 1, \bar{i}\}$;

c. $\widetilde{i-1} < \tilde{i}$ for all $i \in \{2, \ldots, n\}$

d. if $\tilde{i} < k \leq \ell$, then $x_k S_j x_\ell$ for all $i \in \{1, \ldots, n\}$ and all $j \in \{1, 2, \ldots, \bar{i} - 1, \bar{i}\}$.

We prove (a.)–(d.) simultaneously by induction on i, with base cases being $i = 1$ for all four properties (thus there is nothing to prove in the base case for (a.) and (c.)). To make the proof more readable, we will group the induction's base and step by properties (a.)–(d.).

Let us check the property (a.). There is nothing to check for $i = 1$. Assume $i > 1$. By the property (b.) applied to $\overline{i-1}$ we know that for any k such that $\widetilde{i-1} < k$, for all $j \in \{1, \ldots, \overline{i-1}\}$ we have $x_k \nVdash C_j$. Since by the definition of $\bar{i}$ there is k with $\widetilde{i-1} < k$ and $x_k \Vdash C_{\bar{i}}$, we must have $\bar{i} > \overline{i-1}$. The definition of $\bar{i}$ clearly implies $\bar{i} \leq n$, so (a.) holds.

The property (c.) is obvious from the definition of $\tilde{i}$.

Next we check the property (d.). We will prove the induction base and step at the same time. Fix k and ℓ such that $\tilde{i} < k \leq \ell$, and $j \in \{1, \ldots, \bar{i}\}$. If $k = \ell$, the claim follows from the reflexivity of S_j (jRx_k follows from (6)). So assume $k \neq \ell$ and, in addition, that $\ell = k + 1$ (if the claim held for all pairs of adjacent indices, the required result would be implied by the transitivity of S_j). Let m be minimal such that $x_k \Vdash C_m$. Now the property (4.) implies $x_k(R \circ S_m)x_\ell$. By the definition of $\bar{i}$ we have $\bar{i} \leq m$. Since $j \leq \bar{i}$ (hence $j \leq m$), Property (3.) implies $x_k S_j x_\ell$.

Finally, the property (b.). The base case follows from Property (2.). We first consider the case $j < \bar{i}$. The definition of $\bar{i}$ implies that $\bar{i} \leq \ell$ for every ℓ such that C_ℓ is true in x_m for some $m > \tilde{i}$. So, it cannot be the case that $j < \bar{i}$ does not satisfy the property (b.).

It remains to consider the case $j = \bar{i}$. Let m be the minimal index such that $x_{\tilde{i}} \Vdash C_m$. Certainly $m \leq \bar{i}$ since $x_{\tilde{i}} \Vdash C_{\bar{i}}$ by the definition of $\tilde{i}$. We can also show that $m \geq \bar{i}$. Suppose otherwise, i.e. $x_{\tilde{i}} \Vdash C_m$ and $m < \bar{i}$. Then $\widetilde{i-1} < \tilde{i}$ and $x_{\tilde{i}} \Vdash C_m$, contradicting the definition of $\bar{i}$. Thus, we can conclude that $\bar{i} = m$. By Property (4.) we have $x_{\tilde{i}}(R \circ S_{\bar{i}})x_{\tilde{i}+1}$. By the same property we have $x_{\tilde{i}+1}(R \circ S_j)x_{\tilde{i}+2}$ for some $j \geq \bar{i}$, and applying the property (d.) to this, $x_{\tilde{i}+1}(R \circ S_{\bar{i}})x_{\tilde{i}+2}$. Thus, by the transitivity of $S_{\bar{i}}$, we get $x_{\tilde{i}}(R \circ S_{\bar{i}})x_{\tilde{i}+2}$. We can continue this process, and after finitely many steps ($\widetilde{i+1} - \tilde{i}$ steps) we conclude $x_{\tilde{i}}(R \circ S_{\bar{i}})x_{\widetilde{i+1}}$.

By Property (d.), we have $x_{\widetilde{i+1}} S_{\bar{i}} x_k$. Thus, $x_{\bar{i}}(S_{\bar{i}} \circ R \circ S_{\bar{i}})x_k$. Since $\tilde{i} < k$ and $x_{\tilde{i}}$ is an $(S_{\bar{i}} \circ R \circ S_{\bar{i}})$-maximal world in the set we denoted by T earlier, we must have $x \nVdash C_{\bar{i}}$.

We have checked all properties (a.)–(d.).

The property (b.) implies $x_{\widetilde{n+1}} \nVdash C_j$ for all j (the property (a.) implies $\overline{n} = n$). This contradicts the property (2.) satisfied by L_∞. $\dashv$

8.4 Arithmetical soundness

In this section we prove that the logic $\mathbf{IL}\mathsf{W}_\omega$ is arithmetically valid, i.e. a subset of $\mathbf{IL}(\text{All})$.

This semi-formal system CuL is defined in [40]. Essentially, this is an extension of the logic $\mathbf{IL}$ both in terms of the language and theoremhood. It is an alternative to the system AtL that we explored in Chapter 7. Both systems enable us to give modal-like proofs of arithmetical validity for many known principles (see [40]), and at the moment it is not clear whether these systems differ in power. In any case, will use CuL in this chapter. These are the axiom schemas and the rules of the system:

$(\rightarrow)^{\mathsf{J}}$	$\vdash \Box^I A \rightarrow \Box A$
$\mathsf{L}_1^{\mathsf{J}}$	$\vdash \Box^I(A \rightarrow B) \rightarrow (\Box^I A \rightarrow \Box^I B)$
$\mathsf{L}_2^{\mathsf{J}}$	$\vdash \Box^I A \rightarrow \Box^I \Box^J A$
$\mathsf{L}_3^{\mathsf{J}}$	$\vdash \Box^I(\Box^J A \rightarrow A) \rightarrow \Box^I A$
$\mathsf{J}_1^{\mathsf{J}}$	$\vdash \Box(A \rightarrow B) \rightarrow A \rhd B$
$\mathsf{J}_5^{\mathsf{J}}$	$\vdash \Diamond^J A \rhd A$
$\mathsf{Nec}^{\mathsf{J}}$	$\vdash A \Rightarrow \vdash \Box^I A$
M^{J}	$\Gamma, (A \wedge \Box^J C \rhd B \wedge \Box^{J'} C) \vdash D \Rightarrow \Gamma, A \rhd B \vdash D$
	Here J is a variable not occurring in Γ, A, B, D and $J \neq J'$

Reasoning in the system allows using all the regular principles like $\mathsf{J}_2 : (A \rhd B) \wedge (B \rhd C) \rightarrow A \rhd C$, $\mathsf{J}_3 : (A \rhd C) \wedge (B \rhd C) \rightarrow A \vee B \rhd C$ and $\mathsf{J}_4 : A \rhd B \rightarrow (\Diamond A \rightarrow \Diamond B)$.

The intended interpretations of superscripts (such as I and J) are definable cuts. See [40] for more details. Reasoning with definable cuts is the standard way of proving arithmetical validity; see [29] for a rather large application.

Before proving the main theorem of the section (the arithmetical validity of $\mathbf{IL}\mathsf{W}_\omega$), we first prove two auxiliary lemmas.

Lemma 8.10 Let $n \in \omega \setminus \{0\}$. Suppose $\Box(A \rightarrow \bigvee_{1 \leq i \leq n-1} \Diamond^K \neg C_i)$ and $C_{n-1} \rhd \Diamond A \vee B_{n-1} \vee \mathsf{U}_{n-1}$. Then for some cut J the following holds:

$$C_{n-1} \wedge \bigwedge_{1 \leq i \leq n-2} \Box^J \neg C_i \rhd B_{n-1}$$

Proof. Combining the two assumed formulas, we get

$$C_{n-1} \rhd \Diamond \left(\bigvee_{1 \leq i \leq n-1} \Diamond^K \neg C_i \right) \vee B_{n-1} \vee \mathsf{U}_{n-1}.$$

Applying (the contraposition of) $\mathsf{L}_2^{\mathsf{K}}$,

$$C_{n-1} \rhd \left(\bigvee_{1 \leq i \leq n-1} \Diamond \neg C_i \right) \vee B_{n-1} \vee \mathsf{U}_{n-1}.$$

We can unpack U_{n-1} next:

$$C_{n-1} \rhd \left(\bigvee_{1 \leq i \leq n-1} \Diamond \neg C_i \right) \vee B_{n-1} \vee \left(\bigvee_{1 \leq i \leq n-2} \Diamond \neg C_i \right).$$

Equivalently,

$$C_{n-1} \rhd \left(\bigvee_{1 \leq i \leq n-1} \Diamond \neg C_i \right) \vee B_{n-1}.$$

Applying the principle W,

$$C_{n-1} \rhd \left(\bigvee_{1 \leq i \leq n-2} \Diamond \neg C_i \right) \vee B_{n-1}.$$

Finally, there is a cut J such that

$$C_{n-1} \wedge \bigwedge_{1 \leq i \leq n-2} \Box^J \neg C_i \rhd \left(B_{n-1} \vee \bigvee_{1 \leq i \leq n-2} \Diamond \neg C_i \right) \wedge \bigwedge_{1 \leq i \leq n-2} \Box \neg C_i.$$

$\dashv$

Lemma 8.11 For all cut variables K and all $n \in \omega \setminus \{0\}$,

$$\vdash \mathsf{V}_n \rhd A \wedge \bigwedge_{1 \leq i \leq n-1} \Box^K \neg C_i.$$

Proof. We prove the claim by induction on n. If $n = 1$, $\mathsf{V}_n = A$, and clearly $\vdash A \rhd A \wedge \top$.

Suppose the claim holds for all $k \in \{1, \ldots, n-1\}$. Fix K. We are to show that

$$\vdash \mathsf{V}_n \rhd A \wedge \bigwedge_{1 \leq i \leq n-1} \Box^K \neg C_i.$$

We will do this by proving that

$$\vdash \mathsf{V}_n \to \Diamond (A \wedge \bigwedge_{1 \leq i \leq n-1} \Box^K \neg C_i),$$

equivalently (by unpacking V_n and rearranging),

$$\Box (A \to \bigvee_{1 \leq i \leq n-1} \Diamond^K C_i) \wedge (C_{n-1} \rhd \Diamond A \vee B_{n-1} \vee \mathsf{U}_{n-1}) \to \mathsf{V}_{n-1} \rhd B_{n-1}.$$

So, suppose $\Box(A \to \bigvee_{1 \leq i \leq n-1} \Diamond^K C_i)$ and $C_{n-1} \rhd \Diamond A \vee B_{n-1} \vee \mathsf{U}_{n-1}$. By the induction hypothesis,

$$\vdash \mathsf{V}_{n-1} \rhd A \wedge \bigwedge_{1 \leq i \leq n-2} \Box^K \neg C_i.$$

Applying $\Box(A \to \bigvee_{1 \leq i \leq n-1} \Diamond^K C_i)$,

$$\mathsf{V}_{n-1} \rhd \left(\bigvee_{1 \leq i \leq n-1} \Diamond^K C_i \right) \wedge \bigwedge_{1 \leq i \leq n-2} \Box^K \neg C_i.$$

Thus,

$$\mathsf{V}_{n-1} \rhd \Diamond^K C_{n-1} \wedge \bigwedge_{1\leq i\leq n-2} \Box^K \neg C_i.$$

Let J be the cut given by Lemma 8.10. Applying $\mathsf{L}_2^{\mathsf{J}}$,

$$\mathsf{V}_{n-1} \rhd \Diamond^K C_{n-1} \wedge \bigwedge_{1\leq i\leq n-2} \Box^K \Box^J \neg C_i.$$

Thus,

$$\mathsf{V}_{n-1} \rhd \Diamond^K (C_{n-1} \wedge \bigwedge_{1\leq i\leq n-2} \Box^J \neg C_i).$$

Applying $\mathsf{J}_5^{\mathsf{K}}$,

$$\mathsf{V}_{n-1} \rhd C_{n-1} \wedge \bigwedge_{1\leq i\leq n-2} \Box^J \neg C_i.$$

Finally, applying Lemma 8.10 and $\mathsf{J2}$:

$$\mathsf{V}_{n-1} \rhd B_{n-1}.$$

⊣

Theorem 8.12 For all $n \in \omega \setminus \{0\}$, $\vdash \mathsf{W}_n$, i.e.

$$\vdash A \rhd \Diamond A \vee B_n \vee \mathsf{U}_n \to \mathsf{V}_n \rhd B_n$$

Proof. Suppose $A \rhd \Diamond A \vee B_n \vee \mathsf{U}_n$. Applying the principle W, $A \rhd B_n \vee \mathsf{U}_n$. Then there is a cut K such that

$$A \wedge \bigwedge_{1\leq i\leq n-1} \Box^K \neg C_i \rhd (B_n \vee \mathsf{U}_n) \wedge \bigwedge_{1\leq i\leq n-1} \Box \neg C_i.$$

By unpacking U_n we see that

$$A \wedge \bigwedge_{1\leq i\leq n-1} \Box^K \neg C_i \rhd B_n \wedge \bigwedge_{1\leq i\leq n-1} \Box \neg C_i.$$

In particular,

$$A \wedge \bigwedge_{1\leq i\leq n-1} \Box^K \neg C_i \rhd B_n.$$

Lemma 8.11 implies

$$\mathsf{V}_n \rhd A \wedge \bigwedge_{1\leq i\leq n-1} \Box^K \neg C_i.$$

Applying $\mathsf{J2}$ gives $\mathsf{V}_n \rhd B_n$, as required. ⊣

Thus $\mathbf{IL}\mathsf{W}_\omega \subseteq \mathbf{IL}(\text{All})$. Note that this result does not mean that we found a better lower bound for $\mathbf{IL}(\text{All})$. The new principles might depend on some already introduced principles.

8.5 Status

At the moment we don't have answers to the following three questions:

1. Are the principles $(\mathsf{W}_n)_{n\in\omega}$ valid on generalised $\mathbf{IL}\mathsf{WR}$-frames?
2. Do we have $\mathbf{IL}\mathsf{WR} \Vdash W_n$ for all $n \in \omega$?
3. Do we have $\mathbf{IL}\mathsf{WR}_\omega\mathsf{R}^\omega \Vdash W_n$ for all $n \in \omega$?[3]

Of course, if (2) has a positive answer, then (1) and (3) have a positive answer too.

If (1), we have modal completeness of $\mathbf{IL}\mathsf{W}_\omega$ w.r.t. generalised semantics. This is a strictly stronger system than $\mathbf{IL}\mathsf{W}$ and $\mathbf{IL}\mathsf{R}$, and so would be the strongest system yet for which we have modal completeness.
If (1) and (2), then $\mathbf{IL}\mathsf{WR} = \mathbf{IL}\mathsf{W}_\omega$, and so we also have completeness of $\mathbf{IL}\mathsf{WR}$ w.r.t. generalised semantics.
If (1) and not (2) are the case, in addition we have incompleteness of $\mathbf{IL}\mathsf{WR}$ w.r.t. generalised semantics. This would be a first example of an interpretability logic that is incomplete w.r.t. generalised semantics. If (2) is not the case (regardless of (1)), we have incompleteness of $\mathbf{IL}\mathsf{WR}$ w.r.t. ordinary semantics.

If (1) is not the case, then $(\mathsf{W}_n)_{n\in\omega}$ is strictly stronger than $\mathbf{IL}\mathsf{WR}$; in this case, (2) is not the case either. With additional work we might still be able to prove completeness of $\mathbf{IL}\mathsf{W}_\omega$ w.r.t. generalised semantics. This would require us to first formulate the characteristic properties w.r.t. generalised semantics (since in this case, this is not merely a conjunction of $(\mathsf{W})_{\text{gen}}$ and $(\mathsf{R})_{\text{gen}}$).

If (3) is not the case, we have a (strictly) better lower bound of $\mathbf{IL}(\mathsf{All})$: the logic $\mathbf{IL}\mathsf{W}_\omega\mathsf{R}_\omega\mathsf{R}^\omega$.

To sum up, these are the possible outcomes we might hope for:

- Modal completeness of $\mathbf{IL}\mathsf{W}_\omega$ w.r.t. generalised semantics.
- Modal completeness of $\mathbf{IL}\mathsf{WR}$ w.r.t. generalised semantics.
- Modal incompleteness of $\mathbf{IL}\mathsf{W}_\omega$ w.r.t. ordinary semantics.
- Modal incompleteness of $\mathbf{IL}\mathsf{WR}$ w.r.t. ordinary semantics.
- A (strictly) better lower bound of $\mathbf{IL}(\mathsf{All})$: the logic $\mathbf{IL}\mathsf{W}_\omega\mathsf{R}_\omega\mathsf{R}^\omega$.

Currently, we are aiming to show that all the principles W_n are valid on $\mathbf{IL}_{\text{set}}\mathsf{W}_\omega$-frames, i.e. to answer the first question posed above positively.

[3]By $\mathbf{IL}\mathsf{WR}_\omega\mathsf{R}^\omega$ we denote the extension of $\mathbf{IL}\mathsf{W}_\omega$ with the principles $\mathsf{R_n}$ and $\mathsf{R^n}$ for all $n \in \omega$. Similar notation will be used elsewhere too.

Bibliography

[1] E. Beltrami. *Saggio di interpretazione della geometria non-euclidea.* 1868.

[2] A. Berarducci. The Interpretability Logic of Peano Arithmetic. *The Journal of Symbolic Logic*, 55(3):1059–1089, 1990.

[3] A. Berarducci and R. Verbrugge. On the provability logic of bounded arithmetic. *Annals of Pure and Applied Logic*, 61(1):75 – 93, 1993.

[4] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic.* Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2001.

[5] G. S. Boolos. *The Logic of Provability.* Cambridge University Press, Cambridge, 1993.

[6] F. Bou and J.J. Joosten. The closed fragment of IL is PSPACE hard. *Electronical Notes Theoretical Computer Science*, 278:47–54, 2011.

[7] J. P. Burgess. Basic tense logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, pages 89–133. Springer, Dordrecht, 1984.

[8] J. P. Burgess. Basic tense logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Second Edition: Volume VII: Extensions of Classical Logic*, pages 1–42. Kluwer Academic Publishers, Dordrecht, 2002.

[9] S. Buss. *Bounded arithmetic.* Bibliopolis, Napoli, 1986.

[10] S. Buss. Chapter 1: An introduction to proof theory & chapter 2: Firstorder proof theory of arithmetic. In S. R. Buss, editor, *Handbook of Proof Theory.* Elsevier, 1998.

[11] M. Bílková, E. Goris, and J.J. Joosten. Smart labels. In L. Afanasiev and M. Marx, editors, *Liber Amicorum for Dick de Jongh.* Intitute for Logic, Language and Computation, 2004. Electronically published, ISBN: 90 5776 1289.

[12] M. Bílková, D. de Jongh, and J.J. Joosten. Interpretability in PRA. *Annals of Pure and Applied Logic*, 161(2):128–138, 2009.

[13] V. Čačić and V. Kovač. On the share of closed **IL** formulas which are also in **GL**. *Archive for Mathematical Logic*, 54:741–767, 2015.

[14] V. Čačić and D. Vrgoč. A Note on Bisimulation and Modal Equivalence in Provability Logic and Interpretability Logic. *Studia Logica*, 101:31–44, 2011.

[15] V. Čačić and M. Vuković. A note on normal forms for closed fragment of system **IL**. *Mathematical Communications*, 17:195–204, 2012.

[16] A.V. Chagrov and M.N. Rybakov. How many variables does one need to prove PSPACE-hardness of modal logics. In P. Balbiani, N.-Y. Suzuki, F. Wolter, and M. Zakharyaschev, editors, *Advances in Modal Logic*, volume 4, pages 71–82. 2003.

[17] D.H.J. de Jongh, M. Jumelet, and F. Montagna. On the proof of Solovay's theorem. *Studia Logica*, 50:51–69, 1991.

[18] D.H.J. de Jongh and F. Veltman. Provability logics for relative interpretability. In P.P. Petkov, editor, *Mathematical Logic, Proceedings of the Heyting 1988 Summer School in Varna, Bulgaria*, pages 31–42. Plenum Press, Boston, New York, 1990.

[19] D.H.J. de Jongh and F. Veltman. Modal completeness of ILW. In J. Gerbrandy, M. Marx, M. Rijke, and Y. Venema, editors, *Essays dedicated to Johan van Benthem on the occasion of his 50th birthday*. Amsterdam University Press, Amsterdam, 1999.

[20] D.H.J. de Jongh and F.J.M.M. Veltman. Intensional logic. University of Amsterdam, unpublished course text, 1983.

[21] D.H.J. de Jongh, F.J.M.M. Veltman, and R. Verbrugge. Completeness by construction for tense logics of linear time. In J. van Benthem, A. Troelstra, F.J.M.M. Veltman, and A. Visser, editors, *Liber Amicorum for Dick de Jongh*. Institute of Logic, Language and Computation, Amsterdam, 2004.

[22] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta mathematicae*, 49(1):35–92, 1960.

[23] G. Gentzen. Die widerspruchsfreiheit der reinen zahlentheorie. *Mathematische annalen*, 112(1):493–565, 1936.

[24] K. Gödel. Über formal unentscheidbare sätze der principia mathematica und verwandter systeme i. *Monatshefte für mathematik und physik*, 38(1):173–198, 1931.

[25] K. Gödel. Eine interpretation des intuitionistischen aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums*, 4:39–40, 1933.

[26] E. Goris, M. Bílková, J.J. Joosten, and L. Mikec. Assuring and critical labels for relations between maximal consistent sets for interpretability logics, `https://arxiv.org/2003.04623`, 2020, submitted.

[27] E. Goris and J.J. Joosten. Modal Matters for Interpretability Logics. *Logic Journal of the IGPL*, 16(4):371–412, 2008.

[28] E. Goris and J.J. Joosten. A new principle in the interpretability logic of all reasonable arithmetical theories. *Logic Journal of the IGPL*, 19(1):14–17, 2011.

[29] E. Goris and J.J. Joosten. Two new series of principles in the interpretability logic of all reasonable arithmetical theories. *The Journal of Symbolic Logic*, 85(1):1–25, 2020.

[30] K. Gödel and G. W. Brown. *The consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of set theory.* Princeton University Press Princeton, NJ, 1940.

[31] D. Hilbert. Neubegründung der mathematik. erste mitteilung. In *Dritter Band: Analysis · Grundlagen der Mathematik · Physik Verschiedenes*, pages 157–177. Springer, 1935.

[32] D. Hilbert and P. Bernays. *Grundlagen der Mathematik II.* Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2013.

[33] P. Hájek. On interpretability in theories containing arithmetic. II. *Commentationes Mathematicae Universitatis Carolinae*, 022(4):667–688, 1981.

[34] T. Icard and J.J. Joosten. Provabilty and interpretability logics with restricted substitutions. *Notre Dame Journal of Formal Logic*, 53(2):133–154, 2012.

[35] G. Japaridze and D. de Jongh. The logic of provability. In S. Buss, editor, *Handbook of proof theory*, pages 475–546. North-Holland Publishing Co., Amsterdam, 1998.

[36] J.J. Joosten. Towards the interpretability logic of all reasonable arithmetical theories. Master's thesis, University of Amsterdam, 1998.

[37] J.J. Joosten. *Interpretability Formalized.* PhD thesis, Utrecht University, 2004.

[38] J.J. Joosten, J. Mas Rovira, L. Mikec, and M. Vuković. An overview of Generalised Veltman Semantics, 2020. Preprint.

[39] J.J. Joosten, L. Mikec, and A. Visser. Feferman axiomatisations, definable cuts and principles of interpretability. *forthcoming*, 2020.

[40] J.J. Joosten and A. Visser. How to derive principles of interpretability logic, A toolkit. In J. van Benthem, F. Troelstra, A. Veltman, and A. Visser, editors, *Liber Amicorum for Dick de Jongh.* Intitute for Logic, Language and Computation, 2004. Electronically published, ISBN: 90 5776 1289.

[41] R. Kaye and T. L. Wong. On interpretations of arithmetic and set theory. *Notre Dame Journal of Formal Logic*, 48(4):497–510, 2007.

[42] T. Kurahashi and Y. Okawa. Modal completeness of sublogics of the interpretability logic **IL**, `https://arxiv.org/2004.03813`, 2020.

[43] R.E. Ladner. The computational complexity of provability in systems of modal propositional logic. *SIAM Journal of Computing*, 6(3):467–480, 1977.

[44] J. Mas Rovira. Interpretability Logics and Generalised Veltman Semantics in Agda. Master's thesis, Master in Pure and Applied Logic, University of Barcelona, 2020. http://diposit.ub.edu/dspace/handle/2445/133559.

[45] J. Mas Rovira, L. Mikec, and J.J. Joosten. Generalised Veltman semantics in Agda. In R. Verbrugge and N. Olivetti, editors, *Short Papers, Advances in Modal Logic, AiML 2020*, pages 86–90, 2020.

[46] L. Mikec, J.J. Joosten, and M. Vuković. A w-flavoured series of interpretability principles. In R. Verbrugge and N. Olivetti, editors, *Short Papers, Advances in Modal Logic, AiML 2020*, pages 60–64, 2020.

[47] L. Mikec, F. Pakhomov, and M. Vuković. Complexity of the interpretability logic IL. *Logic Journal of the IGPL*, 27(1):1–7, 2018.

[48] L. Mikec and T. Perkov. Existentially valid formulas corresponding to some normal modal logics. In G. Metcalfe G. Bezhanishvili, G. D'Agostino and T. Studer, editors, *Short Papers, Advances in Modal Logic, AiML 2018*, pages 76–80, 2018.

[49] L. Mikec, T. Perkov, and M. Vuković. Decidability of interpretability logics ILM_0 and ILW^*. *Logic Journal of the IGPL*, 25(5):758–772, 2017.

[50] L. Mikec and M. Vuković. Interpretability logics and generalised Veltman semantics. *The Journal of Symbolic Logic*, to appear, 2020.

[51] F. Pakhomov. On the Complexity of the Closed Fragment of Japaridze's Provability Logic. *Archive for Mathematical Logic*, 53(7–8):949–967, 2014.

[52] T. Perkov and L. Mikec. Existential definability of modal frame classes. *Mathematical Logic Quarterly*, 66(3):316–325, 2020.

[53] T. Perkov and M. Vuković. Filtrations of generalized Veltman models. *Mathematical Logic Quarterly*, 62(4-5):412–419, 2016.

[54] K. Segerberg. An essay in classical modal logic. *Filosofiska Föreningen och Filosofiska Institutionen vid Uppsala Universitet*, 1971.

[55] I. Shapirovsky. PSPACE-decidability of Japaridze's polymodal logic. In C. Areces and R. Goldblatt, editors, *Advances in Modal Logic*, volume 8, pages 289–304. 2008.

[56] V.Y. Shavrukov. The logic of relative interpretability over Peano arithmetic. Preprint, Steklov Mathematical Institute, Moscow, 1988. In Russian.

[57] R. M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 28:33–71, 1976.

[58] E. Spaan. *Complexity of Modal Logics*. Haveka, B.V., 1993.

[59] V. Švejdar. Modal analysis of generalized Rosser sentences. *The Journal of Symbolic Logic*, 48:986–999, 1983.

[60] V. Švejdar. Some independence results in interpretability logic. *Studia Logica*, 50:29–38, 1991.

[61] A. Tarski, A. Mostowski, and R. Robinson. *Undecidable Theories*. North-Holland, Amsterdam, 1953.

[62] L.C. Verbrugge. Verzamelingen-Veltman frames en modellen (Set Veltman frames and models). Unpublished manuscript, Amsterdam, 1992.

[63] L.C. Verbrugge. *Efficient Metamathematics*. ILLC dissertation series. Fac. Wiskunde en Informatica, 1993.

[64] L.C. Verbrugge. Provability Logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2017 edition, 2017.

[65] A. Visser. Preliminary notes on interpretability logic. Technical Report LGPS 29, Department of Philosophy, Utrecht University, 1988.

[66] A. Visser. Interpretability logic. In P.P. Petkov, editor, *Mathematical Logic, Proceedings of the Heyting 1988 summer school in Varna, Bulgaria*, pages 175–209. Plenum Press, Boston, New York, 1990.

[67] A. Visser. The formalization of interpretability. *Studia Logica*, 50(1):81–106, 1991.

[68] A. Visser. An overview of interpretability logic. In M. Kracht, M. de Rijke, and H. Wansing, editors, *Advances in modal logic*, volume 1, pages 307–359. 1997.

[69] D. Vrgoč and M. Vuković. Bisimulations and bisimulation quotients of generalized Veltman models. *Logic Journal of the IGPL*, 18(6):870–880, 2009.

[70] M. Vuković. The interpretability logic ILF. *Mathematical Communications*, 2:205–210, 1997.

[71] M. Vuković. Some correspondence of principles in interpretability logic. *Glasnik Matematički*, 31(51):193–200, 1996.

[72] M. Vuković. The principles of interpretability. *Notre Dame Journal of Formal Logic*, 40(2):227–235, 1999.

[73] M. Vuković. Bisimulations between generalized Veltman models and Veltman models. *Mathematical Logic Quarterly*, 54(4):368–373, 2008.

Index of symbols and terms

List of Figures

Biography

Born on 19. March 1993 in Varaždin, Croatia, I attended primary school in Varaždin and Zadar, Croatia and secondary school in Zadar. During this time I competed in various school competitions; most notably, I won the national competition in logic (2010), and received an honourable mention in the International Philosophy Olympiad (2011).

I obtained a BA degree in Computer Science and Philosophy (summa cum laude) in 2014 from the University of Rijeka, Croatia, together with the Dean's Award for the most successful student of the Faculty. During this time I worked as a student teaching assistant for various computer science and maths courses.

I obtained an MSc degree in Computer Science and Mathematics (summa cum laude) in 2016 from the University of Zagreb, Croatia. After that I first enrolled in the joint Croatian PhD program in Mathematics, and then also in the PhD program in Mathematics and Computer Science, University of Barcelona.

I worked as a teaching and research assistant at the University of Rijeka and the University of Zagreb. The list of courses I taught over the last four years:

- Application of Computers in Mathematics (16/17);
- Computer Networks (18/19);
- Database Systems (17/18, 18/19, 19/20, 20/21; two sessions each year);
- Elementary Mathematics (16/17);
- Computer Lab 1 (19/20; two sessions in 20/21);
- Computer Lab 3 (20/21);
- Computer Lab for Mathematics Education 1 (18/19);
- Computer Lab for Mathematics Education 2 (17/18);
- Programming 1 (17/18, 18/19);
- Programming 2 (17/18, 18/19, 19/20);
- Programming for Contemporary Processors (18/19);
- Numerical Mathematics (17/18);
- Numerical Mathematics for Computer Science (17/18);
- Set Theory (two sessions in 16/17).

Published full papers: [49], [47], [50], [52]. Published short papers: [48], [46] and [45].