Categories for the working modal logician

Giovanni Cinà

Categories for the working modal logician

ILLC Dissertation Series DS-2017-04

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

Copyright © 2017 by Giovanni Cinà

Cover: Simonetta Chierici, *Al di là delle colonne di Ercole 4*.
Back cover: Simonetta Chierici, *Le colonne di Ercole 3*.

Printed and bound by Ipskamp Printing.

ISBN: 978-94-028-0601-4

Categories for the working modal logician

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 16 juni 2017, te 11.00 uur

door

Giovanni Cinà

geboren te Milano, Italië

Promotiecommisie

Promotor:	Prof. dr. J.F.A.K. van Benthem	Universiteit van Amsterdam
Co-promotor:	Dr. A. Baltag	Universiteit van Amsterdam
Overige leden:	Prof. dr. Y. Venema	Universiteit van Amsterdam
	Dr. N. Bezhanishvili	Universiteit van Amsterdam
	Prof. dr. D.J.N. van Eijck	Universiteit van Amsterdam
	Prof. dr. S. Abramsky	University of Oxford
	Prof. dr. S. Ghilardi	University of Milan
	Prof. dr. B.P. Kooi	Rijksuniversiteit Groningen

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

a Simonetta e Giuseppe,
a Sara, Giovanni, Nelda e Franco,
ai greci e agli arabi,
agli ulivi,
alla ddisa che cova il fuoco

Contents

Acknowledgments

First and foremost, I want to thank my supervisors Alexandru Baltag and Johan van Benthem. Alexandru's enthusiasm for research and the clarity of his vision have been a continuous source of inspiration for me. With hindsight, his example was what convinced me to pursue a PhD. It has been a privilege to work alongside him in these years. Johan joined the team only in my third year, but it is hard for me to imagine how this thesis would have been possible without him. He encouraged me constantly, with patience and altruism, even from the opposite side of the world; he brought out the best in me and for this I am grateful.

I also would like to extend my gratitude to the members of my thesis committee: Samson Abramsky, Nick Bezhanishvili, Jan van Eijck, Silvio Ghilardi, Barteld Kooi, and Yde Venema. I really appreciate their commitment to make sense of this long document, despite the University rules that prevent me from including their feedback in the final version.

Next in line I want to mention some of the academics I had the honor to collaborate with, for I learned a great deal from them. Ulle Endriss, who helped me move my first steps into the article-writing business; Benno van den Berg, who taught me how to teach; Sebastian Enqvist, for sharing his kindness, modesty and friendship. I furthermore want to thank the members of the Amsterdam Dynamics Group and participants to the LIRa seminar; our meetings were a fantastic opportunity to broaden and enrich my knowledge.

The ILLC is an extraordinary environment for a young researcher and most of the credit for this goes to the staff. Tanja, Karine, Peter, Marco and Jenny deserve a special mention for being always exceptionally kind and helpful. My thanks also to the ILLC directors, first Yde Venema and now Sonja Smets, for building and maintaining such a unique place.

A heartfelt thanks to my office mates Malvin, Dieuwke and Jouke for making our room welcoming and relaxed. There is a future in which I will sit on a giant rubber ball and install Linux, I promise. As per the rest of the ILLC members, I am grateful for sharing these years with the extended family of PhD colleagues.

From serious political conversations to light-hearted mockery, I feel lucky for all the time spent together. In my mind all these moments are condensed in one still image: we are all laughing around the coffee machine, setting aside worries and deadlines, briefly liberated by each other's company.

I tried to give back to this community as much as possible, serving in the ILLC PhD Council and the FNWI PhD Council. These activities gave me purpose when I felt I had none, so I am thankful and honored to have worked as representative. To the friends I met in these councils, as well as those met during the protests in 2015, I say: keep up the fight. It is a worthy cause.

In these years I basked in the energy of Amsterdam like a lizard under the sun. Here I re-discovered one of my strongest passions, dancing, which complemented perfectly the purely abstract and intellectual endeavour of a PhD in Logic. I owe much of my mental sanity to the dancers in the Rock'n Roll and Lindy Hop scene. One group I hold especially dear, the Clandestien gang, for they took me in as a family member and showed me the never-ending joy of the dance floor. To my dancing partner Tanne, thank you for sharing your dedication with me and for trusting me acrobatically.

Ultimately, friends made this city into my home: Kleitia, my long-time pal and Seinfeld buddy, my sister in arms Willy, Ana Lucia, the sweetest person on this planet, Iris and Ronald, who never cease to inspire me, l'amico ritrovato Basim, the unstoppable Camilo, Melle please tell another quirky story, Elia e le ragguardevoli serate, Shanti e il nostro dialetto, and the boulder climbers, the swing dancers, the artists gravitating around Rongwrong, the storytellers from Mezrab, the buzuki players, the barbershop singers and the triathlon runners.

Amsterdam is also a harbour, bustling with newcomers full of hopes as well as people hugging their loved ones before embarking for the next adventure. I met many friends and saw many leave. Claudio, Tanvi, Corver, Pavel, Johannes, Stephen, Stefania, Hugo, just to name a few, joined the diaspora of the people I care about. Thanks for sharing this leg of your journey with me. On this note, I would like to hug all the friends that sent me their love across countries and continents. In particular Leone, Giorgio, Cigo, Mambo, Simona, Emanuele, Elena e Mattia, you guys have been on my side since forever, walked with me, hiked with me, helped me climb real and metaphorical mountains. I miss you.

Finally, my warmest thoughts go to the two people who made all this possible, my parents Giuseppe and Simonetta. They are my foundations, my truest supporters: I owe them everything.

Amsterdam
May, 2017.

Giovanni Cinà

Chapter 1

Introduction

This thesis revolves around the connection between Category Theory and Modal Logic, focusing on the bearing of the former discipline on the latter. Our aim is not to explain away one framework in terms of the other, but rather to stress how the interaction between these two fields can bring about novel and fruitful insights. Such cross-pollination can occur at multiple levels; in this research we dive into two possibilities.

The first is to study hybrid structures, namely mathematical objects that have a clear connection to Category Theory and at the same time are amenable to a Modal Logic treatment. Not only such models turn out to be interesting on their own account, their model theory exhibits peculiar features when classical logical issues, e.g. completeness, expressivity and decidability, are injected with category-theoretic elements. The first three chapters of the thesis constitute an exploration of this territory.

A second possibility is to import questions and methodologies from one side to the other, exploiting one discipline as an heuristic tool. The suggestion we put forward, developed in the last two chapters, is that classes of models of modal languages can be conveniently seen as categories, where the appropriate notion of morphism is given by the bisimulation matching the language under consideration. This perspective offers a uniform way of defining basic notions and raising basic questions, but also sheds new light on classical issues in the field.

In what follows we offer two brief introductions to the main actors in our play, Category Theory and Modal Logic. We then touch on the existing work on the interface between the two disciplines. We conclude this preamble with an outline of the content of the chapters, followed by a list of publications from which said content is taken.

Category Theory. Stemming from the tradition of Felix Klein's Erlangen Program, the main concepts of Category Theory were introduced in the 40's by Eilemberg and MacLane in the seminal paper "General theory of natural equivalences" [52]. As the title suggests, the main interest of the authors was the study of natural transformations. As the theory developed, it quickly became clear that the most innovative aspect of these investigations was ingrained in the very definition of a category: rather than focusing on the internal structure of mathematical objects, category theory turns the attention to the relevant notion of transformation between certain kinds of structures. A category is thus composed of two parts, elements and arrows connecting them. Arrows are endowed with a partial operation of composition and are required to satisfy some general laws.

Important classes of mathematical objects, such as groups or vector spaces, can be arranged into a category by selecting a meaningful notion of arrows (sometimes called morphisms); the latter are often taken to be some 'structure-preserving' functions, e.g. group homomorphisms in the case of groups. Category Theory can thus be used as a powerful abstraction tool to systematically address questions that arise in different parts of mathematics within one general framework.

The original fields of application of Category Theory were algebraic topology and abstract algebra. From the second half of the 20th century to the present day, the connections of Category Theory with other fields within and outside Mathematics have increased steadily, to the extent that now category-theoretic tools are used in Mathematical Logic, Theoretical Computer Science and Mathematical Physics. For a structured account of the history of Category Theory and a glimpse at the breadth of its applications we refer the reader to [90] and [91].

Modal Logic. Although the core ideas of Modal Logic have deeper roots, the origin of the field is often taken to be C.I. Lewis' "A survey of symbolic logic" [82], where the author enriched propositional classical logic with an additional operator I meant to capture the concept 'it is impossible that'. The approach to modal languages in the early days had a distinctive syntactic flavor and authors mainly employed such languages to axiomatize specific philosophical concepts.

This scenario changed radically in the 1960s when semantics entered the stage. The introduction of relational semantics, more than any other later development, shaped the field in such a profound way that nowadays modal languages can be tentatively defined as apt languages for talking about relational structures.[1] The 'aptness' part of this definition refers to the two virtues of expressivity and tractability: one wants a language that is able to express important features of the semantics while at the same time preserve computational tractability. Unfortunately there is a trade off between these two desiderata, as more expressive languages tend to have high computational complexity or be undecidable. Modal languages strike a good balance between these two aspects, whence their attractiveness.

This point of view highlights the relationship between modal languages and more expressive languages such as first-order logic and extensions thereof. The

[1] This slogan welcomes the reader at the first page of the modern textbook [32].

idea of seeing modal languages as fragments of more complex languages, made precise by van Benthem's theorem in [24], brought to the foreground the intuition that modal languages could be characterized in terms of their invariants, against the background of a point-of-reference expressive language.

As for the applications of Modal Logic, we mentioned the early use of this formalism to capture and axiomatize philosophical notions such as 'necessity' and 'possibility'. Nowadays modal languages are employed for many different purposes: some are still connected with various areas of Philosophy such as epistemic and doxastic logics; some germane to Computer Science such as temporal logics, logics to reason about programs, logics for multi-agent systems and knowledge representation; some pertaining to Economics such as logics for games and strategic interactions. We redirect the reader to the historical section of [32] for a detailed overview of the general trends of research in this discipline.

Bridges between the two fields.

The interface between these two disciplines has been studied from a variety of angles. We give a short tour of the existing work in the area in order to properly place our own contribution.

An influential approach that touched on this connection is the one known as Coalgebra. Coalgebras are mathematical constructs that were introduced in Theoretical Computer Science to handle infinite data types; it later became clear that they could be conceived as abstract versions of systems. Coalgebras are dual of algebras in a precise, category-theoretic sense, and many important concepts in the realm of Universal Algebra have a natural dual in the theory of coalgebras.[2]

Subsequent developments in this research area unveiled the connection between coalgebras and particular logics, collectively grouped under the name 'coalgebraic logics'. It was shown that many modal logics could be recovered as coalgebraic logic for the right choice of functor, for example the standard modalities on relational and neighborhood structures. We refer the reader to [78] for an introduction to coalgebraic logic. The link between our work and the general theory of coalgebra can be made precise by noticing that presheaves, which constitute the models on which we interpret the logics studied in the first three chapters, can themselves be viewed as special coalgebras. This link is discussed at the end of Chapter 2, where we expand on the relationship between our logic and a coalgebraic logic associated to presheaves, and again in Chapter 3.

Another avenue was taken by the authors of the so-called 'presheaf approach to Concurrency Theory'. The first paper on the subject was [123], where many important models of concurrency such as transition systems, synchronization trees and event structures were organized into categories and systematically related via adjunctions. Upon realization that each of these models was associated to

[2]This is well-portrayed in the table in [108] p.7. For an introduction to coalgebras we point the reader to the classic [108] and the textbook [72].

a corresponding notion of path, in [74] Joyal, Winskel and Nielsen devised a representation of models of concurrency in terms of presheaves over suitable ‘path categories’, following the intuition that a model of concurrency consists of bundles of different paths glued together in a coherent way.

This perspective suggested to define a general categorical notion of behavioural equivalence solely in terms of path preservation and path ‘lifting’. While the former is usually inbuilt in the definition of morphism of the categories under examination, the latter had to be imposed, leading to the definition of *open maps*. The desired general concept of bisimilarity was then at hand: two models of concurrency are deemed bisimilar if their presheaf representations are connected by a span of open maps. In a follow-up paper [124] it was observed that presheaves can themselves be regarded as transition system via the construction usually known as ‘category of elements’. A notion of bisimulation for these transition systems, baptized ‘path bisimulation’, was proved equivalent to the bisimulation in terms of span of open maps. A modal logic called ‘path logic’ was shown to be characteristic for such path bisimulation. Given some conditions on the base category, presheaves can be thought of as generalized models of concurrency, with representables playing the role of path shapes. Path logic becomes then the natural choice of language for such models. The connection of our work with this line of research is described in detail in Chapter 3.

Last but not least, there is the body of work on the Curry-Howard correspondence for constructive modal logics. The core idea of this approach, elucidated in [47], was to establish a triangle of correspondences between constructive modal logics, type-theoretic formulations and categorical semantics (the latter in the spirit of Categorical Logic, see [87]). Although the program did not reach full completion, this line of research provided interesting insights into the categorical semantics for constructive modal logics. Examples are provided by [30], where the authors describe what structure needs to be added to a cartesian closed category in order to interpret the modalities of intuitionistic S4, or [6], where Kripke semantics and categorical semantics are related via algebraic semantics.

Since the background logic obtained via a categorical semantics is intuitionistic, this perspective is naturally geared towards constructive modal logics, namely logics based on the intuitionistic propositional calculus rather then the classical one. Intuitionistic modal logics lend themselves to computational interpretations of the modalities.[3] In this thesis we only consider modal logics based on classical propositional calculus, thus there does not seem to be a straightforward connection to our research program.

Beside these structured approaches, one can find category-theoretic motives in other fields that overlap with Modal Logic; a prominent example in this respect is Duality Theory, namely the systematic study of contravariant functors between different categories of mathematical objects. The dualities between classes of

[3]See [6] and [31] for pointers to the use of constructive modal logics in Computer Science.

relational structures, algebras and topological spaces offer deep insights into the semantics of modal languages and can be harnessed to prove important results, e.g. the algebraic proof of Goldblatt-Thomason theorem [62].

In this kind of investigations Category Theory comes in as a tool employed to achieve some independently motivated goals. The spirit of our own research is somewhat similar, in the sense that we want to develop machinery and intuitions that can be useful for modal logicians for a variety of purposes. In the next section we summarize our contributions.

1.1 Outline

In **Chapter 2** we introduce a relational structure called 'typed transition system', in which the labels of the relations are indexed by the arrows of a category. To highlight the bridge-like status of typed transition systems, we prove an equivalence with the corresponding presheaf category and later show an adjunction with the category of standard labeled transition systems.

We explain how to axiomatize these structures with a modal language, providing a Hilbert style calculus that is sound and strongly complete for these models. Such logic contains an infinitary rule, necessary to exclude the existence of untyped states. Although infinitary modal logics are known to be highly intractable, the logic we present is quite well behaved, in the sense that any derivation from a consistent set of formulas can be reduced to a finitary derivation. We proceed to show soundness and weak completeness results for the finitary fragment of the logic. Finally, we argue that our logic matches the coalgebraic logic arising from the coalgebras associated to presheaves.

Chapter 3 relates the work of the previous chapter to the 'presheaf approach' and explains how the logic for typed transition systems can be seen as a fragment of path logic. After reviewing some definitions and results from the literature, we prove a characterization result for path logic, in the fashion of van Benthem's theorem. We proceed to examine some properties of presheaves from the literature and see how to encode them in path logic. The main test cases are the sheaf of sections of a covering space, a construction used in topology, and a recent sheaf-theoretic analysis of non-locality and contextuality pioneered by Abramsky and Brandenburger in [4]. A core notion that one would like to capture is that of *sheaf*, but we observe that sheaves are not definable in path logic. We thus devote the second part of the chapter to the understanding of sheaves over topological spaces through their relational counterparts.

On one hand we enhance the semantic companion of path logic, namely path bisimulation, to preserve locality and gluing, the defining properties of a sheaf. Moving the first steps in this direction, we prove some basic results concerning the adequate notion of bisimulation in the context of sheaves, characterizing spans and co-spans of open maps. Pursuing a different avenue, we add enough expressive power to define the two properties, enhancing path logic with nominals. We conclude by showing that, when the category in the background is nice enough, the finitary fragment of this hybrid extension is decidable.

While previous chapters studied path logic and its expressivity at an abstract level, **Chapter 4** showcases what well-chosen fragments can achieve in concrete areas. The case study of this chapter is the formalization of Social Choice Theory. After introducing Social Choice Theory and the existing work on its formal foundations, we present a modal logic for social choice functions. Such logic is shown to be complete for the intended models; furthermore we explain how various concepts of interest for Social Choice Theory can be modelled in this logic. Most importantly, we describe how the logic can be used to prove three seminal impossibility results in this field. We discuss how this logic fares in comparison to other languages proposed for the same task and offer some remarks on the implementation of the logic by describing how to feed it to a SAT solver.

The modality in this language encodes the capability of a coalition to enforce the truth of a certain formula, given that the individuals outside the coalition maintain their course of action. This suggests that the central aspect of this theory is the possibility to track what happens to the preferences expressed by a coalition of agents when said coalition is expanded or shrunk. Following this intuition we explicate how a social choice function can be understood as a presheaf model, where the base category is the poset of all possible coalitions. These observations are followed by an explanation of how the logic for social choice function can be seen as a fragment of the relative path logic.

Leaving presheaf models behind, **Chapter 5** turns to a different question: what do we gain from conceiving a class of models for a modal language as a category, where the role of arrow is played by bisimulations? Since arrows are first-class citizens from the perspective of Category Theory, two questions become prominent: What is the right notion of bisimulation for a given modal operator? Is it closed under composition? It turns out that there is a group of well-known modalities for which these questions are not settled, namely conditional modalities. Chapter 5 provides a structured answer to these two issues for this class of operators, at the general level of conditional models. We provide a bisimulation for conditional modalities and prove the correspondence between bisimilarity and modal equivalence for the semantics on selection functions, along with other observations on the closure under unions and compositions.

In the rest of the chapter we demonstrate the versatility of this framework by applying it to areas of Modal Logic that have seen recent development. First we discuss the case of conditional belief on plausibility models, deriving some undefinability observations along the way. A similar analysis is conducted for conditional belief on evidence models, showing how we can handle the same operator interpreted on different semantics. Second, we prove that our approach covers more than just conditional belief by applying it to the operator of relativized common knowledge. Finally we explain how the central definition and results are amenable for a multi-agent generalization.

Chapter 6 continues to study models of modal languages from a categorical standpoint, building on the groundwork of the previous chapter. We focus on the case-study of plausibility and evidence models, explaining how these classes of models can be arranged into different categories by means of different choices of bisimulations. Since different bisimulations are linked to different modal languages, we can think of picking a notion of bisimulation as if selecting a language 'through which' we look at the models.

Regarding a class of models as a category whose arrows are bisimulations allows us to recast some known concepts and problems in categorical terms. An important notion is that of update, namely a model-changing operation that occurs after the model is fed with new information. Requiring an update to be functorial, for these particular categories of models, means to ask (among other things) whether bisimilar models are mapped to bisimilar models. This suggests a link between functoriality of an update and the existence of reduction laws for the associated dynamic operator. Another theme is the relationship between classes of models. A mapping between two different classes of models can have different properties when the classes are regarded as categories. For some choices of languages (read: bisimulations) such mapping will not be functorial, while for other languages the mapping will turn out to give a categorical equivalence.

Finally, both issues are composed in the problem known as tracking, namely the matching of information dynamics on different structures. One of the key aspects of tracking is the possibility to reduce an update on a complex structure to an update on a simpler construct. When tracking occurs we are able to transfer results from the updates on simpler structures to the updates on richer structures. The main result in this chapter is a characterization of the trackable updates in a certain class of "simple" updates: for the updates that fit the description we provide a procedure to construct the corresponding update on plausibility models; for the updates that do not meet the requirements we describe how to build a counterexample to tracking.

Prerequisites. We assume knowledge of the basics of Modal Logic, as well as the core notions of Category Theory such as category, functor and natural transformation. For quick reference we point the reader to [32] and [12].

1.2 Sources of the chapters

The content of this thesis

- Chapter 2 is partly based on an unpublished manuscript [16].
- Chapter 3 is partly based on:

  Giovanni Cinà and Sebastian Enqvist. Bisimulation and path logic for sheaves: contextuality and beyond. Technical report, ILLC Technical Notes X-2015-01, 2015.
- Chapter 4, with the exception of the connection to path logic, is based on two papers (where the second is an extended version):

  Giovanni Cinà and Ulle Endriss. A syntactic proof of Arrow's theorem in a modal logic of social choice functions. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1009–1017, 2015.

  Giovanni Cinà and Ulle Endriss. Proving classical theorems of social choice theory in modal logic. *Autonomous Agents and Multi-Agent Systems*, 30(5):963–989, 2016.
- Chapter 5 is based on two papers (where the former is an extended version):

  Alexandru Baltag and Giovanni Cinà. Bisimulation for conditional modalities. Technical report, ILLC Technical Notes X-2016-01, 2016.

  Alexandru Baltag and Giovanni Cinà. Bisimulation for conditional modalities. *Studia Logica*, 2017. Forthcoming.
- Chapter 6 is based on an unpublished manuscript [43].

Chapter 2

Typed modal logics

2.1 Introduction

We begin our investigation by analyzing the connection between presheaves and labeled transition systems, highlighting how the former correspond to particular relational structures with an associated modal logic.

Labeled transition systems (LTS henceforth) are widely used as mathematical representations of processes, where the latter are encoded as transitions between possible states of the system. This analysis can be refined by restricting the class of LTSs to capture a specific notion of process. The core idea of this chapter is to introduce and study a structure that is designed to capture *typed* processes. A typed process differs from an un-typed one in one important respect: every state (or possible world, or object) has a unique type and transition only connect states of a certain, predetermined type to states of another predetermined type.

Everyday examples of typed processes include cooking recipes, where one has to perform different operations depending on the ingredients, and instructions to assemble furniture. For more formal examples the reader can think of programs (that do not require interaction with the user) written in an object-oriented language: there the classes of objects are the types while the functions send objects of one class to objects of another predetermined class. In general, we want to consider any complex procedure involving several sorts of objects and sort-related operations.

We take these types and operations and generate a category, by adding a 'do nothing' operation for each type and understanding composition as performing one operation after the other.[1] Our structure of choice, which with a leap of fantasy we call *typed transition systems* (TTS), is a relational structure that incorporates the typing given by a category. In contrast with regular LTSs, where all the states are of the same kind, in a TTS we can have states of different 'sorts' or 'types' (we

[1]The choice of term 'type' is not accidental: its use to refer to objects of a category sinks its roots in the connections between type theory and category theory, as explained in [23] and [70].

will use these two terms interchangeably). The transitions are also type-dependent: for every label f there are two types, a "domain" type C and "codomain" type C', such that all the transitions labelled by f connect C-states to C'-states. The category-theoretic features help us handle the types, while the relational nature of the structure allows for the application of the Modal Logic toolbox.

The idea of enriching the structure of the set of labels is already problematized in [112], where the author explicitly argues in favor of the study of LTS with additional structure on labels. Important examples are labelled transition systems with a monoidal structure on labels [39] and those whose set of labels is a particular category, e.g. in [81,95]. The coalgebraic point of view on LTSs [108] does not seem to be suitable for such a task, in that the set of labels is fixed within the signature functor. The same criticism is offered for the approach of Winskel and Nielsen in [74] and [124]. Sobocinski's solution in [112] is the study of such richer LTSs by means of relational presheaves. Our approach is inspired by the work of Winskel and Nielsen on the categorical study of models of concurrency ([74, 123, 124]); indeed some of the functors we propose in the following sections are variations of their constructions. We are however not interested in encompassing one framework into the other, but rather we intend to explore the interplay between Category Theory and Modal Logic; indeed this chapter revolves around a structure that sits in between these two theories.

The chapter is structured as follows. In the next Section we introduce some preliminary notions and give a formal definition of typed transition systems. To highlight the bridge-like status of TTSs, in Section 2.3 we prove an equivalence with the corresponding presheaf category, a pivotal concept of Category Theory, while in Section 2.4 we show an adjunction with standard LTSs.

In order to axiomatize such structures, in Section 2.5 we provide a modal language, also parameterized by a category $\mathbf{C}$, and a Hilbert style calculus that is sound and strongly complete for the class of TTSs labeled by $\mathbf{C}$. Such logic is called $\mathbf{LTTS^C}$. It contains an infinitary rule, necessary to exclude the existence of untyped states. It is worth remarking that infinitary modal logics are known to be highly intractable, regardless of whether the infinitary character is due to infinitary connectives or infinitary rules. Nevertheless, the logic we present is quite well behaved, in the sense that any derivation from a consistent set of formulas can be reduced to a finitary derivation. This enables a Lindenbaum-like construction, which is typically where the standard completeness proof goes awry in the case of infinitary modal logics.[2] In Section 2.6 we proceed to show that the calculus can be made finitary, preserving all the theorems but losing strong completeness; we derive soundness and weak completeness results for this finitary logic.

Since the category of presheaves can be equivalently described as a particular category of coalgebras (see [5]), in Section 2.7 we argue that our logic matches

[2]In the case of Hilbert-style axiomatizations of modal logics with infinitary connectives one can resort to consistency properties; see [107] and [114].

the one arising from the coalgebras associated to presheaves. We generalize Pattinson's technique to extract a coalgebraic logic from a collection of natural transformations, transporting it to the realm of multi-typed **Set**-coalgebras; the application of this procedure to the right class of natural transformations is shown to yield the language and semantics of Section 2.5. Since Section 7 touches on the connection with Coalgebra and coalgebraic logic, a familiarity with the topic is helpful in appreciating the results therein.

2.2 Typed transition systems

We start introducing some terminology and basic observations.[3] A *labelled transition system* (LTS) is a tuple $\mathcal{T} = \langle W, \{R_i\}_{i \in L}\rangle$ where W and L are sets and $R_i \subseteq S \times S$ are relations on W indexed by labels in L.[4] A transition system is *deterministic* if every relation in it is a partial function.

A *bisimulation* between two LTSs $\mathcal{T}_1 = \langle W_1, \{R^1_i\}_{i \in L}\rangle$ and $\mathcal{T}_2 = \langle W_2, \{R^2_i\}_{i \in L}\rangle$ is a non-empty relation $Z \subseteq W_1 \times W_2$ such that, if $(w, w') \in Z$:

1. if $wR^1_i v$ then there is $v' \in W_2$ such that $(v, v') \in Z$ and $w'R^2_i v'$;
2. if $w'R^2_i v'$ then there is $v \in W_1$ such that $(v, v') \in Z$ and $wR^1_i v$.

A bisimulation is *functional* if B is a function; functional bisimulations are sometimes called p-morphisms or bounded morphisms. Note that we have defined bisimulations only between transition systems with the same set of labels.

2.2.1. Definition. Call $\mathbf{TS}_L$ the category having as objects transition systems with labels in L and as morphisms functional bisimulations.[5]

2.2.2. Proposition. *$\mathbf{TS}_L$ is a category.*

Proof:
The identity function is a bisimulation and the composition of functional bisimulations is the composition of the underlying functions; the forward and backward condition follow from the fact that the components are bisimulations. Associativity and identity laws follow from the corresponding laws on functions. □

We call a relation $I \subseteq W \times W$ a *pseudo-identity* if it is a partial identity function on W, that is, the identity function for a subset of W. We now introduce

[3] We refer to [32] for an overview on notions related to Modal Logic, and to [84] for the category-theoretic concepts.

[4] Labelled transition systems are often introduced as tuples $\langle W, L, Tr\rangle$, where $Tr \subseteq W \times L \times W$ is a relation specifying which pairs of states are related and how the edges are labelled. Clearly the two presentations are equivalent; we chose ours because it helps the intuition underlying our construction. We do not consider LTSs with initial state.

[5] See [123] for a general categorical study of labelled transition systems.

the special kind of LTS that we will study in the next sections. Note that, for the rest of the chapter, we will take the category $\mathbf{C}$ to be small, i.e. the collections of objects and arrows are both sets. We indicate with $\mathbf{C}_0$ and $\mathbf{C}_1$ the sets of objects and arrows respectively.

2.2.3. Definition. [Typed transition system] A *typed transition system* (TTS) is a tuple $\langle W, \mathbf{C}, \{R_f\}_{f \in \mathbf{C}_1} \rangle$ such that $\mathbf{C}$ is a small category, W is a set and $\{R_i\}_{i \in \mathbf{C}_1} \rangle$ is a family of relations indexed by the arrows of the category. The relations are deterministic and moreover satisfy the following properties:

1. The relations in the family $\{R_{Id_C}\}_{C \in \mathbf{C}_0}$ are pseudo-identities on W.
2. The domains (and thus the codomains) of the relations in the family $\{R_{Id_C}\}_{C \in \mathbf{C}_0}$ form a partition of W.
3. For all $f \in \mathbf{C}_1$, if $f : C' \to C$ then
   - if $(x, y) \in R_f$ then $(y, y) \in R_{Id_{C'}}$,
   - the domain of R_f coincides with the domain of R_{Id_C}.
4. If $f : C \to C'$ and $g : C' \to C''$ in $\mathbf{C}$ then $R_{g \circ f} = R_g; R_f$, where the symbol ; on the left is relational composition.

Note that by condition 3 and 4 two partial functions R_f and R_g are composable only if the codomain of R_f and the domain of R_g agree.[6]

2.2.4. Definition. Suppose two TTSs M_1 and M_2 are indexed by the same category $\mathbf{C}$. A *bisimulation* between them is a relation $Z \subseteq W_1 \times W_2$ such that if $(w, w') \in Z$ then

- for every $f \in \mathbf{C}_1$, if $(w, v) \in R^1_f$ then there is $v' \in W_2$ such that $(w', v') \in R^2_f$ and $(v, v') \in Z$, and vice versa.

When we consider TTSs equipped with a valuation on atomic propositions, we take a bisimulation to additionally satisfy $V_1(w) = V_2(w')$ for all $(w, w') \in Z$.

Typed transition systems can also be arranged into a category: call $\mathbf{TTS_C}$ the category of typed transition systems with labels in $\mathbf{C}_1$ and functional bisimulations.

[6] If the set of object of the category is a singleton $\{*\}$ then the conditions enforce a monoidal structure on the labels. Deterministic LTSs with a monoidal structure on labels - one of the examples mentioned in [112] - are thus an example of TTSs, namely those where the background category $\mathbf{C}$ has only one object.

2.3 Presheaves as TTSs

In this section we study the relationship between the category $\mathbf{TTS_C}$ and the presheaf category $\mathbf{Set}^{\mathbf{C}^{op}}$, concluding that the two are equivalent. This allows for the transfer of important results from presheaves to TTSs.

First some definitions and terminology. Given two categories $\mathbf{C}$ and $\mathbf{D}$, the *functor category* $\mathbf{D}^{\mathbf{C}}$ is a category having as objects the functors from $\mathbf{C}$ to $\mathbf{D}$ and as arrows natural transformations. A *presheaf* on a category $\mathbf{C}$ is a functor from $\mathbf{C}^{op}$, the opposite category of $\mathbf{C}$, to the category $\mathbf{Set}$ of sets and functions. The functor category $\mathbf{Set}^{\mathbf{C}^{op}}$ for these special functors is called the *presheaf category* over $\mathbf{C}$.

Presheaf categories are widely used in Category Theory; they are an important example of cartesian closed categories and of topoi. Presheaves have been employed on a variety of fronts, from applications to Topology to current models of quantum computation [55, 88]. For the fundamentals on presheaves we redirect the readers to the classic texts in Category Theory and Topos Theory, [84] and [85].

2.3.1. DEFINITION. Given a functor $F : \mathbf{C} \to \mathbf{Set}$, we can construct the *category of elements* $\mathcal{E}(F)$ as follows:

- objects are pairs (x, C), for C object of $\mathbf{C}$ and $x \in F(C)$
- there are arrows $\overline{f} : (x, C) \to (x', C')$ for all morphisms $f : C \to C'$ in $\mathbf{C}$ such that $F(f)(x) = x'$

This construction is used, for example, in proving that presheaves categories are the cocompletion of the corresponding base category, see [12]. It is a special case of the Grothendieck construction. Note that if the base category is small, we can transform the category of elements $\mathcal{E}(F)$ into a LTS $\langle W, \{R_f\}_{f \in \mathbf{C}_1}\rangle$ as follows:

- $W = \{(x, C) | C \in \mathbf{C}_0, x \in F(C)\} = \mathcal{E}(F)_0$
- $R_f = \{((x, C), (x', C')) | f : C \to C', F(f)(x) = x'\}$

Since $\mathbf{C}$ is small, the carrier W is a union of set-many sets and thus a set, while $\{R_f\}_{f \in \mathbf{C}_1}$ is a family indexed by a set.

2.3.1 The functor T

We now address the question: how can we apply the construction of the category of elements to a presheaf $F : \mathbf{C}^{op} \to \mathbf{Set}$? After a moment of reflection, it becomes clear that there are two alternatives, depending on how we want to encode the contravariance of the functor into the relations of the LTS:

1. $R_f := \{((F(f)(x), C)(x, C')) | x \in F(C')\}$ for $f : C \to C'$

2. $R_f := \{((x, C)(F(f)(x), C'))|x \in F(C)\}$ for $f : C' \to C$

The first construction yields relations that are the inverses of functions, while the second construction always yields a deterministic LTS. The first option is studied in [124] and will be addressed in the next chapter; here we focus on the second alternative. A LTS obtained with the second procedure turns out to be a TTS.

2.3.2. Definition. Given a small category $\mathbf{C}$ and a functor $F : \mathbf{C}^{op} \to \mathbf{Set}$, define a TTS $T_F = \langle W, \{R^F_f\}_{f\in\mathbf{C}_1}\rangle$ as follows:

- $W^F := \{(x, C)|x \in F(C), C \in \mathbf{C}_0\}$
- $R^F_f := \{((x, C)(F(f)(x), C'))|x \in F(C)\}$ for $f : C' \to C$

Thus each R^F_f is a partial function on W^F.

2.3.3. Proposition. *For every small category* $\mathbf{C}$*,* T_F *is a TTS labeled by* $\mathbf{C}$*.*

Proof:
We immediately have that T_F is deterministic, since every R^F_f is a partial function. The partial functions in T_F are indexed by the arrows of $\mathbf{C}$. The elements indexed by the identities are pseudo-identities on W^F and by construction their domains form a partition of it.

If $(a, b) \in R^F_f$ and f is not an identity then by construction it must be that $a = (x, C)$, $b = (F(f)(x), C')$ for $f : C' \to C$. Clearly $(a, a) = ((x, C), (x, C)) \in R^F_{Id_C}$ and $(b, b) = ((F(f)(x), C'), (F(f)(x), C')) \in R^F_{Id_{C'}}$. We automatically have that R^F_f is total on the domain of $R^F_{Id_C}$ because $F(f)$ is total on $F(C)$. Conditions 4 is given by functoriality and the definition of the transitions in T_F. □

This construction can be adapted to a functor $T : \mathbf{Set}^{\mathbf{C}^{op}} \to \mathbf{TTS_C}$ defined as:

$$F : \mathbf{C}^{op} \to \mathbf{Set} \mapsto T_F$$
$$\theta : F \to G \mapsto B_\theta \subseteq T_F \times T_G$$

where T_F is defined as above and $B_\theta = \{((x, C), (\theta_C(x), C))|x \in F(C), C \in \mathbf{C}_0\}$ is a functional bisimulation from T_F to T_G.

To show that this construction is well-defined we check that B_θ is an arrow in the target category.

2.3.4. Proposition. B_θ *is a functional bisimulation.*

Proof:
We start with functionality. For every (x, C) there is a corresponding pair $(\theta_C(x), C)$ because of the functionality of each component θ_C of the natural transformation, so every element in W has a unique image under B_θ.

For two presheaves $F, G : \mathbf{C}^{op} \to \mathbf{Set}$, consider the corresponding models T_F and T_G and let $B_\theta \subseteq T_F \times T_G$. Suppose given the pairs $((x, C), (\theta_C(x), C)) \in B_\theta$ and $((x, C)(F(f)(x), C')) \in R^F_f$, for some R^F_f partial function in T_F. Therefore $x \in F(C)$, $\theta_C(x) \in G(C)$ and $F(f)(x) \in F(C')$. Taking $(\theta_{C'} \circ F(f)(x), C')$ we can see that, by naturality of θ, it is equal to $(G(f) \circ \theta_C(x), C')$, and hence it makes the following diagram commute

$$\begin{array}{ccc} (x, C) & \xmapsto{B_\theta} & (\theta_C(x), C) \\ {\scriptstyle R^F_f}\big\downarrow & & \big\downarrow{\scriptstyle R^G_f} \\ (F(f)(x), C') & \xmapsto[B_\theta]{} & (\theta_{C'} \circ F(f)(x), C') \end{array}$$

For the backward condition suppose now that $((x, C), (\theta_C(x), C)) \in B_\theta$ and $((\theta_C(x), C), (G(f) \circ \theta_C(x), C')) \in R^G_f$ for some R^G_f relation in T_G. Since $F(f) : F(C) \to F(C')$ is a function we know that $((x, C), (F(f)(x), C')) \in R^F_f$, and by definition $((F(f)(x), C'), (\theta_{C'} \circ F(f)(x), C')) \in B_\theta$. Again by the naturality of θ we obtain that $(\theta_{C'} \circ F(f)(x), C') = (G(f) \circ \theta_C(x), C')$, thus we have the commutation of the above diagram. □

Notice that this proof does not carry over if we construct the LTS from a presheaf in the first way, as in [124], since the 'back' condition of bisimulation might fail.

2.3.5. PROPOSITION. *T is a functor.*

Proof:
The preservation of source and target is given by construction. For the identity Id_F, which is the identity natural transformation, we get that $T(Id_F) = \{((x, C), (Id_{F_C}(x), C)) | x \in F(C), C \in \mathbf{C}_0\} = \{((x, C), (x, C)) | x \in F(C), C \in \mathbf{C}_0\}$ which is the identity relation on T_F. For composition, take $\theta \circ \eta$. The corresponding functional bisimulation is $T(\theta \circ \eta) = \{((x, C), (\theta \circ \eta_C(x), C)) | x \in F(C), C \in \mathbf{C}_0\}$, which is the composition of the functional bisimulations $T(\eta) = \{((x, C), (\eta_C(x), C)) | x \in F(C), C \in \mathbf{C}_0\}$ and $T(\theta) = \{((x, C), (\theta_C(x), C)) | x \in F(C), C \in \mathbf{C}_0\}$. □

2.3.6. PROPOSITION. *T is full, faithful and injective on objects.*

Proof:
For injectivity on objects consider two functors $F \neq G$. If $F(C) \neq G(C)$ for some

$C \in \mathbf{C}_0$ then there must be $x \in F(C)$ such that $x \notin G(C)$ thus the pair (x, C) will be in the carrier of T_F but not in the carrier of T_G. If the functors coincide on objects but $F(f) \neq G(f)$ for some $f \in \mathbf{C}_1$ then we will have that R^F_f is a different relation from R^G_f, and so the transition systems T_F and T_G will have the same carrier but different partial functions.

Now suppose $\theta, \eta : F \to G$ and $\theta \neq \eta$. Then there must be a C and an $x \in F(C)$ such that $\theta_C(x) \neq \eta_C(x)$. But this means that $((x, C), (\theta_C(x), C)) \in T(\theta)$ but it cannot be in $T(\eta)$, because $((x, C), (\eta_C(x), C)) \in T(\eta)$ and there can be only one image of (x, C) in $T(\eta)$. So $T(\theta) \neq T(\eta)$.

To see that it is full, consider a functional bisimulation B between two objects in the image of T, say from T_F to T_G. We construct a natural transformation $\theta : F \to G$ such that $T(\theta) = B$.

Define $\theta_C(x) = y$ iff $((x, C), (y, C')) \in B$. We know by the functionality of B that there is only one such pair, hence θ_C is well defined for every C. We now claim that $C' = C$. Consider the fact that, by construction, $((x, C)(x, C)) \in R^F_{Id_C}$. By the forward condition on bisimulation we must have that

$$\begin{array}{ccc}
(x, C) & \xrightarrow{B} & (y, C') \\
{\scriptstyle R^F_{Id_C}} \downarrow & & \downarrow {\scriptstyle R^G_{Id_C}} \\
(x, C) & \xrightarrow[B]{} & (y, C')
\end{array}$$

Therefore it must be that $((y, C'), (y, C')) \in R^G_{Id_C}$, and this in turn entails that $C = C'$ by the construction of $R^G_{Id_C}$. So we have the right typing for the components of θ: $\theta_C(x) = y \in G(C)$. It remains to show naturality.

For each object x in $F(C)$, we have that $((x, C), (\theta_C(x), C)) \in B$ by definition of θ_C. Given an arrow $f : C' \to C$ in $\mathbf{C}$, the functor G will output the partial function R^G_f. Since $\theta(x) \in G(C)$, by construction we know that $((\theta_C(x), C), (G(f) \circ \theta_C(x), C')) \in R^G_f$. Applying the backward condition on bisimulations we conclude that there must be a pair (z, C'') such that $((x, C), (z, C'')) \in R^F_f$ and $((z, C''), (G(f) \circ \theta_C(x), C')) \in B$.

By the first item we can infer that $C'' = C'$ and $z = F(f)(x)$. From these facts, together with the second item and the definition of $\theta_{C'}$ we can conclude that $\theta_{C'} \circ F(f)(x) = \theta_{C'}(F(f)(x)) = \theta_{C'}(z) = G(f) \circ \theta_C(x)$. As both $x \in F(C)$ and f were generic, we can conclude that θ is a natural transformation. Clearly we have $T(\theta) = B$. □

2.3.2 The functor Pre

The next step is to describe a functor from TTSs to presheaves and then prove the equivalence. Given a typed transition system $\mathcal{T}$ labeled by $\mathbf{C}$ we can construct a presheaf $Pre(\mathcal{T}) : \mathbf{C}^{op} \to \mathbf{Set}$ as

$$C \in \mathbf{C}_0 \mapsto dom(R_{Id_C})$$
$$f : C' \to C \in \mathbf{C}_1 \mapsto R_f : dom(R_{Id_C}) \to dom(R_{Id_{C'}})$$

It takes a simple check to see that this is indeed a presheaf. We can make this construction functorial by defining $Pre : \mathbf{TTS_C} \to \mathbf{Set}^{\mathbf{C}^{op}}$

$$\mathcal{T} \mapsto Pre(\mathcal{T})$$
$$B : \mathcal{T}_1 \to \mathcal{T}_2 \mapsto \theta : Pre(\mathcal{T}_1) \to Pre(\mathcal{T}_2)$$

where the components of θ are defined as follows: for $x \in dom(R^1_{Id_C})$ put $\theta_C(x) = y$ iff $(x, y) \in B$ (the subscript 1 indicates that the partial function is in the transition system $\mathcal{T}_1$).

2.3.7. Proposition. *$Pre : \mathbf{TTS_C} \to \mathbf{Set}^{\mathbf{C}^{op}}$ is a functor.*

Proof:
We know that $Pre(\mathcal{T})$ is a well defined presheaf and that Pre preserves source and target. We need to show that $Pre(B) = \theta : Pre(\mathcal{T}_1) \to Pre(\mathcal{T}_2)$ is a natural transformation. Let C be an object of the category $\mathbf{C}$, $F_1 = Pre(\mathcal{T}_1)$ and $F_2 = Pre(\mathcal{T}_2)$. Suppose $\theta_C(x) = y$; first we want to ensure that the typing is right, namely that $y \in F_2(C)$. Since $(x, x) \in F_1(Id_C) = R^1_{Id_C}$ and $(x, y) \in B$ by the forward condition on bisimulation we know that there must be a y' such that $(x, y') \in B$ and $(y, y') \in R^2_{Id_C}$. By the fact that $R^2_{Id_C}$ is a pseudo-identity we can conclude that $y \in F_2(C) = dom(R^2_{Id_C})$.

Now for naturality. Suppose given $f : C' \to C$ in $\mathbf{C}$. Take $x \in F_1(C)$ and consider $F_2(f) \circ \theta_C(x) = z$. By the definitions of θ and F_2 we can infer that $(x, \theta_C(x)) \in B$ and $(\theta_C(x), z) \in F_2(f) = R^2_f$. By the backward condition on bisimulation there must be a y' such that $(x, y') \in R^1_f = F_1(f)$ and $(y', z) \in B$. But this means that $F_2(f) \circ \theta_C(x) = \theta_{C'} \circ F_1(f)(x)$, hence we have the commutation of the naturality diagram. □

2.3.8. Theorem. *For every small category $\mathbf{C}$ the categories $\mathbf{Set}^{\mathbf{C}^{op}}$ and $\mathbf{TTS_C}$ are equivalent.*

Proof:
Consider $T : \mathbf{Set}^{\mathbf{C}^{op}} \to \mathbf{TTS_C}$ and $Pre : \mathbf{TTS_C} \to \mathbf{Set}^{\mathbf{C}^{op}}$. We show two natural isomorphisms $\eta : T \circ Pre \to Id_{\mathbf{TTS_C}}$ and $\epsilon : Pre \circ T \to Id_{\mathbf{Set}^{\mathbf{C}^{op}}}$.

We begin by showing that $T(Pre(\mathcal{T}))$ is isomorphic to $\mathcal{T}$. First consider the carrier of $\mathcal{T}$, call it W. By definition of T the carrier W' of $T(Pre(\mathcal{T}))$ is

$$\{(x, C)|x \in Pre(\mathcal{T})(C), C \in \mathbf{C}_0\} = \{(x, C)|x \in dom(R_{Id_C}), C \in \mathbf{C}_0\}$$

So W' is essentially W, with the difference that each element is turned into a pair consisting of the element itself and its 'object label' or 'type'. There exists an obvious bijection $\eta_{\mathcal{T}} : W' \to W$ sending (x, C) to x. Clearly this bijection respects the cells of the partition.

Now consider a partial function R_f on $\mathcal{T}$, for $f : C \to C'$. If $(x, y) \in R_f$ then $(x, x) \in R_{Id_{C'}}$ and $(y, y) \in R_{Id_C}$. By definition we have that $Pre(\mathcal{T})(f) = R_f$, so $(x, y) \in Pre(\mathcal{T})(f)$. Finally, applying T we get

$$\begin{aligned} R_f^{Pre(\mathcal{T})} &:= \{((x, C')(Pre(\mathcal{T})(f)(x), C))|x \in Pre(\mathcal{T})(C')\} \\ &= \{((x, C')(y, C))|x \in Pre(\mathcal{T})(C')\} \\ &= \{((x, C')(y, C))|x \in dom(R_{Id_{C'}})\} \end{aligned}$$

So we can see that $R_f^{Pre(\mathcal{T})}$ contains the same pairs as R_f if we disregard the associated types, that is, we apply $\eta_{\mathcal{T}}$. Thus the bijection $\eta_{\mathcal{T}}$ respects all the partial functions on $T \circ Pre(\mathcal{T})$. Being both a bijection and respecting the partial functions $\eta_{\mathcal{T}}$ is a functional bisimulation that is also an isomorphism $\eta_{\mathcal{T}} : T \circ Pre(\mathcal{T}) \cong \mathcal{T}$ in the category $\mathbf{TTS_C}$.

To conclude the first half of the proof we need to show that η is natural. Applying the definition we can see that given a functional bisimulation $B : \mathcal{T}_1 \to \mathcal{T}_2$ the result of applying $T \circ Pre$ is (in stages):

$$\begin{aligned} &Pre(B)_C = \{(x, y)|y = B(x), (x, x) \in R^1_{Id_C}\} \\ &T \circ Pre(B)_C = \{((x, C), (y, C))|y = B(x), (x, x) \in R^1_{Id_C}\} \end{aligned}$$

where the superscript in $R^1_{Id_C}$ indicates that the partial functions lives in $\mathcal{T}_1$. This makes the following naturality diagram commute: given (x, C) in $T \circ Pre(\mathcal{T}_1)$, applying $T \circ Pre(B)_C$ we obtain (y, C) for $y = B(x)$, finally applying η_2 we get y; for the other half of the diagram, $\eta_1(x, C) = x$ and $B(x) = y$.

$$\begin{array}{ccc} T \circ Pre(\mathcal{T}_1) & \xrightarrow{\eta_1} & \mathcal{T}_1 \\ {\scriptstyle T \circ Pre(B)} \downarrow & & \downarrow {\scriptstyle B} \\ T \circ Pre(\mathcal{T}_2) & \xrightarrow[\eta_2]{} & \mathcal{T}_2 \end{array}$$

Now for the second part. Consider the presheaf $F : \mathbf{C}^{op} \to \mathbf{Set}$. Applying the functor T we obtain T_F, the transition system described in Section 2.3. Applying

Pre we obtain a presheaf $Pre(T_F) : \mathbf{C}^{op} \to \mathbf{Set}$. Note that now

$$Pre(T_F)(C) = dom(R^F_{Id_C}) = \{(x, C) | x \in F(C)\}$$

Again we can see that $Pre \circ T(F)$ is almost F, the only difference is that the set associated to each object of the category is a set of pairs where the second element is the object itself. We can thus define a natural transformation $\epsilon_F : Pre \circ T(F) \to F$, a family of functions $\epsilon_{F_C} : Pre \circ T(F)(C) \to F(C)$ defined as $\epsilon_{F_C}((x, C)) = x$. Notice that each of these functions is a bijection, even though we may have that $x \in F(C)$ and $x \in F(C')$ for different C and C': since the natural transformation acts component-wise, the components ϵ_{F_C} are injective and surjective.

We now prove the naturality of ϵ_F. For $f : C \to C'$, the definition of $Pre \circ T(F)(f)$ is $Pre \circ T(F)(f : C \to C') = R^F_f$. It is then easy to see that the diagram

$$\begin{CD} Pre \circ T(F)(C') @>{\epsilon_{F_{C'}}}>> F(C') \\ @V{Pre \circ T(F)(f)}VV @VV{F(f)}V \\ Pre \circ T(F)(C) @>>{\epsilon_{F_C}}> F(C) \end{CD}$$

commutes: given (x, C'), applying $\epsilon_{F_{C'}}$ and $F(f)$ we will get $F(f)(x)$; applying $Pre \circ T(F)(f : C \to C') = R^F_f$ we get $(F(f)(x), C)$ and applying ϵ_{F_C} we also get $F(f)(x)$. Therefore ϵ_F is a natural bijection and thus an isomorphism in the category $\mathbf{Set}^{\mathbf{C}^{op}}$.

We now need to show that ϵ_F is natural in F, that is, it is a natural transformation $\epsilon : Pre \circ T \to Id_{\mathbf{Set}^{\mathbf{C}^{op}}}$. Consider the natural transformation $\theta : F \to G$. We want to show the commutation of the diagram

$$\begin{CD} Pre \circ T(F) @>{\epsilon_F}>> F \\ @V{Pre \circ T(\theta)}VV @VV{\theta}V \\ Pre \circ T(G) @>>{\epsilon_G}> G \end{CD}$$

and to this end we show the commutation of the diagram pointwise:

$$\begin{array}{ccc} Pre \circ T(F)(C) & \xrightarrow{\epsilon_{F_C}} & F(C) \\ {\scriptstyle Pre \circ T(\theta)_C = \theta_{B_\theta, C}} \Big\downarrow & & \Big\downarrow {\scriptstyle \theta_C} \\ Pre \circ T(G)(C) & \xrightarrow[\epsilon_{G_C}]{} & G(C) \end{array}$$

We can build up the definition of the natural transformation $Pre \circ T(\theta)$ as follows:

$$\begin{aligned} T(\theta) = B_\theta &= \{((x, C), (\theta_C(x), C) | \dots\} \\ Pre \circ T(\theta) = \theta_{B_\theta} & \quad \text{s.t.} \\ & \theta_{B_\theta, C}(z) = z' \quad \text{iff} \quad (z, z') \in B_\theta, (z, z) \in R^F_{Id_C} \\ & \text{iff} \quad ((x, C), (\theta_C(x), C)) \in B_\theta, z = (x, C), z' = (\theta_C(x), C) \end{aligned}$$

where the superscript in $R^F_{Id_C}$ indicates that the partial function lives in $T(F)$. Using all the definitions we can check that, given (x, C),

$$\begin{aligned} \theta_C(\epsilon_{F_C}((x, C))) &= \theta_C(x) \\ &= \epsilon_{G_C}((\theta_C(x), C)) \\ &= \epsilon_{G_C}(\theta_{B_\theta, C}(x, C)) \end{aligned}$$

This concludes the proof of the naturality of ϵ and the second half of the proof of the theorem. □

This theorem allows for a transfer of results and information from presheaves to TTSs.

2.3.9. Proposition. *For every small category* $\mathbf{C}$*, the following facts hold.*

1. *The category* $\mathbf{TTS_C}$ *is a topos.*
2. *There is an embedding* $j : \mathbf{C} \to \mathbf{TTS_C}$ *that is full and faithful.*
3. *For every object* $C \in \mathbf{C}$*, consider the image of the representable presheaf* $Hom(C, -)$ *under the equivalence, call such a structure 'representable TTS for C' and denote it with* $TTS(C)$*. Then:*
   - *Every TTS is a colimit of representable TTSs.*
   - *For any TTS* M*, there is a bijection between objects of type* C *in* M *and functional bisimulations from* $TTS(C)$ *to* M*.*

Proof:
Items 1 and 3(1) are given by the fact that equivalences preserve such properties

and item 2 is given by the composition of the Yoneda embedding with the equivalence. Item 3(2) is a consequence of the Yoneda Lemma: given an element of type C in a TTS M, this will be an object in $P(C)$, where P is the pre-image of M under the equivalence; the Yoneda Lemma states that there is a bijection between elements of $P(C)$ and natural transformations $\theta : Hom(C, -) \to P$, while the equivalence entails that there is a bijection between the natural transformations $\theta : Hom(C, -) \to P$ and the functional bisimulations $TTS(C) \to M$. Thus composing the two bijections we obtain the desired correspondence. □

The class of TTS over $\mathbf{C}$ is thus extremely rich of structure.

2.4 Transition systems as TTSs

In this section we expand on the connection between typed transition systems and standard transition systems. The former are a richer version of the latter, so one obvious connection is that, given a TTSs labeled by $\mathbf{C}$, one can forget the additional structure and take $\mathbf{C}_0$ just as a set of labels, obtaining a plain transition system. A more interesting observation, due to Joyal, Winskel and Nielsen, is that we can encode a LTS labeled by a set L into a presheaf over the 'category of paths', i.e., describe a LTS as a bundle of chains of transitions glued together at certain points. From the previous section we know that presheaves correspond to TTSs, thus this gives us a procedure to turn every LTS into a TTS. Interestingly, this extends to an adjunction between LTS and TTS, where the functor from TTSs to LTSs is not a forgetful functor.

For a set of labels L, call $\mathbf{TS}'_L$ the category having as objects transition systems with labels in L and relation-preserving functions as morphisms. Take $\mathbf{L}^*$ to be the full subcategory of $\mathbf{TS}'_L$ consisting of only linear transition systems, i.e., chains of transitions. For all intents and purposes we can envision this category as having for objects the finite strings of labels in L and as arrows the substring inclusions.[7] For example, a string $llkl$ is a set of five points connected in a chain by the corresponding labeled edges, while, for $l, k \in L$, an arrow $i : lk \to llkl$ will be the obvious relation-preserving function mapping one chain into the other. The empty string, denoted by ε, is taken to indicate the transition system with only one object, denoted by $\star$, and no transitions. We use $\bar{l}$ to denote a string in $\mathbf{L}^*$.

In this section we show how, generalizing a construction given in [74], we can embed $\mathbf{TS}'_L$ into $\mathbf{Set}^{\mathbf{L}^{*op}}$ and successively into $\mathbf{TTS}_{\mathbf{L}^*}$. It is worthwhile to remark that this embedding offers a technique to transform a transition system into a deterministic one. Moreover, we show an adjunction between the two categories, where the embedding functor is the right adjoint.

[7] Note that in [74] the same notation is used to denote the category that has *initial* substring inclusions as morphisms.

2.4.1. Theorem. *For any set of labels L, the category $\mathbf{TS}'_L$ embeds into $\mathbf{TTS}_{\mathbf{L}^*}$.*

Proof:
By our previous theorem we know that $\mathbf{Set}^{\mathbf{L}^{*op}}$ is equivalent to $\mathbf{TTS}_{\mathbf{L}^*}$, hence we only need to show that $\mathbf{TS}'_L$ embeds into the presheaf category $\mathbf{Set}^{\mathbf{L}^{*op}}$. In order to achieve this we generalize the embedding of pointed LTSs described in [74] and [124], obtaining a variation of the Yoneda embedding.

Given a LTS $\mathcal{T}$, construct a presheaf $G_{\mathcal{T}} : \mathbf{L}^{*op} \to \mathbf{Set}$ as follows:

$$G_{\mathcal{T}}(\bar{l}) = Hom_{\mathbf{TS}'_L}(\bar{l}, \mathcal{T})$$

where $\bar{l}$ is a finite string of labels. Note that in particular $G_{\mathcal{T}}(\varepsilon)$is just the set of states of $\mathcal{T}$. The action on morphisms is defined as

$$G_{\mathcal{T}}(i : \bar{l} \to \bar{l}')(f : \bar{l}' \to \mathcal{T}) = f \circ i$$

To make this construction into a functor $G : \mathbf{TS}'_L \to \mathbf{Set}^{\mathbf{L}^{*op}}$ we define the action on morphisms as follows. Given a relation preserving function $f : \mathcal{T}_1 \to \mathcal{T}_2$, define

$$G(f : \mathcal{T}_1 \to \mathcal{T}_2) = \eta^f : G_{\mathcal{T}_1} \to G_{\mathcal{T}_2}$$

such that on a component $\bar{l}$ we have

$$\eta^f_{\bar{l}}(h : \bar{l} \to \mathcal{T}_1) = f \circ h : \bar{l} \to \mathcal{T}_2$$

This is routinely proven to be a natural transformation: for $h : \bar{l} \to \mathcal{T}_1$ and $i : \bar{l}' \to \bar{l}$, $G_{\mathcal{T}_2}(i)(\eta^f_{\bar{l}}(h)) = G_{\mathcal{T}_2}(i)(f \circ h) = f \circ h \circ i = \eta^f_{\bar{l}'}(h \circ i) = \eta^f_{\bar{l}'}(G_{\mathcal{T}_1}(i)(h))$.

We now show that the functor G is injective on objects. If two transition systems $\mathcal{T}_1$ and $\mathcal{T}_2$ differ on the states then we will have $G_{\mathcal{T}_1}(\varepsilon) \neq G_{\mathcal{T}_2}(\varepsilon)$. If one of the two has an edge with a label l that does not appear in the other one then $G_{\mathcal{T}_1}(l) \neq \emptyset = G_{\mathcal{T}_2}(l)$. Suppose the two differ because one of the edges is put in different positions in the two systems, let l be the label. Then the two inclusion $i_1, i_2 : \varepsilon \to l$ will be mapped to different functions: either $G_{\mathcal{T}_1}(i_1) \neq G_{\mathcal{T}_2}(i_1)$ or $G_{\mathcal{T}_1}(i_2) \neq G_{\mathcal{T}_2}(i_2)$ or both. Therefore if two transition systems are different then their images under G are different.

We now show G is faithful. Suppose $f, f' : \mathcal{T}_1 \to \mathcal{T}_2$ but $f \neq f'$. Then there is x such that $f(x) \neq f'(x)$. Call i_x the morphism in $\mathbf{TS}'_L$ of type $\varepsilon \to \mathcal{T}_1$ such that $i_x(\star) = x$. Then we have

$$G(f)_\varepsilon(i_x) = \eta^f_\varepsilon(i_x) = f \circ i_x = f(x) \neq f'(x) = f' \circ i_x = \eta^{f'}_\varepsilon(i_x) = G(f')_\varepsilon(i_x)$$

Hence the images of f and f' are different natural transformations.

It remains to show that G is full. Take $\eta : G_{\mathcal{T}_1} \to G_{\mathcal{T}_2}$. Define the function $f : \mathcal{T}_1 \to \mathcal{T}_2$ as $f(x) = \eta_\varepsilon(x)$, that is, following the action of the natural transformation on the empty paths. To see that it preserve edges, suppose there is an edge with label l from x to y in $\mathcal{T}_1$. Call w the morphisms $l \to \mathcal{T}_1$ that singles out this edge. We have $w \in G_{\mathcal{T}_1}(l)$ by construction. Recall the names of the two inclusion $i_1, i_2 : \varepsilon \to l$. Consider the following diagram:

$$\begin{array}{ccc} G_{\mathcal{T}_1}(\varepsilon) & \xrightarrow{\eta_\varepsilon} & G_{\mathcal{T}_2}(\varepsilon) \\ {\scriptstyle G_{\mathcal{T}_1}(i_1)}\uparrow & & \uparrow{\scriptstyle G_{\mathcal{T}_2}(i_1)} \\ G_{\mathcal{T}_1}(l) & \xrightarrow{\eta_l} & G_{\mathcal{T}_2}(l) \end{array}$$

This diagram, commuting by naturality of η, represents the fact that there is an edge labeled by l in $\mathcal{T}_2$, namely $\eta_l(w)$, and that the starting point of such edge is $\eta_\varepsilon(G_{\mathcal{T}_1}(i_1)(w)) = \eta_\varepsilon(x) = f(x)$. The analogous diagram where i_1 is replaced with i_2 captures the fact that the ending point of $\eta_l(w)$ is $f(y)$. We thus have showed that there is an edge with label l from $f(x)$ to $f(y)$, and thus the function f is relation-preserving.

Finally, we must show that $G(f) = \eta$. Given $\bar{l}$ object of $\mathbf{L}^*$, by definitions we have $G(f)_{\bar{l}}(h : \bar{l} \to \mathcal{T}_1) = f \circ h = \eta_\varepsilon \circ h$. Let $\eta_{\bar{l}}(h) = h'$, we want to show that $h' = \eta_\varepsilon \circ h$. Let x be a point in $\bar{l}$, call $i_x : \varepsilon \to \bar{l}$ the morphism that has x in its image. Consider

$$\begin{array}{ccc} G_{\mathcal{T}_1}(\varepsilon) & \xrightarrow{\eta_\varepsilon} & G_{\mathcal{T}_2}(\varepsilon) \\ {\scriptstyle G_{\mathcal{T}_1}(i_x)}\uparrow & & \uparrow{\scriptstyle G_{\mathcal{T}_2}(i_x)} \\ G_{\mathcal{T}_1}(\bar{l}) & \xrightarrow{\eta_{\bar{l}}} & G_{\mathcal{T}_2}(\bar{l}) \end{array}$$

This diagram commutes by naturality of η and showcases that $G_{\mathcal{T}_2}(i_x)(\eta_{\bar{l}}(h)) = G_{\mathcal{T}_2}(i_x)(h') = h' \circ i_x = h'(x)$ is the same as $\eta_\varepsilon \circ G_{\mathcal{T}_1}(i_x)(h) = \eta_\varepsilon \circ (h \circ i_x) = \eta_\varepsilon \circ h(x)$. Hence $G(f) = \eta$ and G is full. This concludes the proof that G is an embedding. □

Recall that $T : \mathbf{Set}^{\mathbf{L}^{*op}} \to \mathbf{TTS}_{\mathbf{L}^*}$ was the functor used in the previous section to prove the equivalence between $\mathbf{Set}^{\mathbf{L}^{*op}}$ and $\mathbf{TTS}_{\mathbf{L}^*}$.

2.4.2. THEOREM. *For any set of labels L, there is an adjunction between the categories $\mathbf{TS}'_L$ and $\mathbf{TTS}_{\mathbf{L}^*}$, where the functor $T \circ G$ is the right adjoint.*

Proof:
Define the left adjoint functor $H : \mathbf{TTS}_{\mathbf{L}^*} \to \mathbf{TS}'_L$ as follows. Given a TTS $\mathcal{T} = \langle W, \{R_i\}_{i \in \mathbf{L}_1^*} \rangle$, construct a LTS $H(\mathcal{T}) = \langle W', \{R_l\}_{l \in L} \rangle$ such that

- $W' = domR_{Id_\varepsilon}$, that is, all the elements of 'type' empty string

- $R_l = \{(a,b) \in W' \times W' | \exists x \in dom(R_{Id_l})\,(x,a) \in R_{i_1}, (x,b) \in R_{i_2}\}$, where $l \in L$ and $i_1, i_2 : \varepsilon \to l$ are the only two inclusion of the empty string into the string consisting of only the l element.

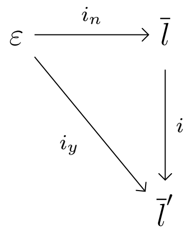
The intuition behind this definition is that R_l is the set of pairs such that there is an object of 'type' l connecting a to b. For the action on arrows, suppose $Z : \mathcal{T}_1 \to \mathcal{T}_2$ is a functional bisimulation. Define $H(Z) : H(\mathcal{T}_1) \to H(\mathcal{T}_2)$, for $a \in domR^1_{Id_\varepsilon}$, as $H(Z)(a) = Z(a)$. Since $b = Z(a)$ and $(a,a) \in R^1_{Id_\epsilon}$, by the forward condition on bisimulation we know that $(b,b) \in R^2_{Id_\varepsilon}$, so $H(Z)$ is indeed well defined.

We show $H(Z)$ is relation-preserving. Suppose $(a,b) \in R^1_l$, then by construction $\exists x \in dom(R^1_{Id_l})$ such that $(x,a) \in R^1_{i_1}$ and $(x,b) \in R^1_{i_2}$. Let $Z(x) = x'$; by the previous argument we know that $x' \in dom(R^2_{Id_l})$. By $Z(x) = x'$ and $(x,a) \in R^1_{i_1}$, again employing the forward condition on bisimulation, we conclude that there is a' such that $Z(a) = a'$ and $(x',a') \in R^2_{i_1}$. A similar argument yields the conclusions $Z(b) = b'$ and $(x',b') \in R^2_{i_2}$. Now we have $x' \in dom(R^2_{Id_l})$ such that $(x',a') \in R^2_{i_1}$ and $(x',b') \in R^2_{i_2}$, thus by construction $(a',b') \in R^2_l$. Since $a' = Z(a) = H(Z)(a)$ and $b' = Z(b) = H(Z)(b)$, we get $(H(Z)(a), H(Z)(b)) \in R^2_l$.

We now define a natural transformation $\eta : 1_{\mathbf{TTS_{L^*}}} \to T \circ G \circ H$ and show that it serves as the unit of the adjunction. For a TTS $\mathcal{T} = \langle W, \{R_f\}_{f \in \mathbf{L}^*}\rangle$, note that $TGH(\mathcal{T}) = \langle \bigcup_{\bar{l} \in \mathbf{L}^*_0} \{(i,\bar{l}) | i \in Hom_{\mathbf{TS}'_L}(\bar{l}, H(\mathcal{T}))\}, \{R_i\}_{i \in \mathbf{L}^*_1}\rangle$. For $(a,a) \in R_{Id_{\bar{l}}}$, $\eta_{\mathcal{T}}(a) = (\tilde{a}, \bar{l})$ such that $\tilde{a} : \bar{l} \to H(\mathcal{T})$; $\tilde{a}$ is defined as $\tilde{a}(n) = R_{i_n}(a)$, where n is a node in the string $\bar{l}$ and $i_n : \varepsilon \to \bar{l}$ is the inclusion that selects said node.

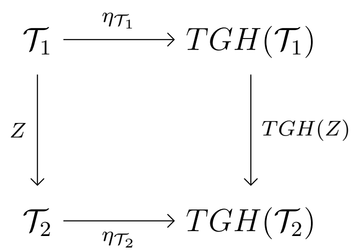
It remains to show that $\tilde{a}$ is relation-preserving and thus a legitimate element of $TGH(\mathcal{T})$. Suppose in $\bar{l}$ there is an edge (n,n') labeled by k. Since k is a label in the string $\bar{l}$, there must be an embedding $i_k : k \to \bar{l}$ that singles out the pair (n,n'), that is: $i_n : \varepsilon \to \bar{l}$ is such that $i_n = i_k \circ i_1$ and $i_{n'} : \varepsilon \to \bar{l}$ is such that $i_{n'} = i_k \circ i_2$. Now consider $x = R_{i_k}(a)$ in $\mathcal{T}$. By the typing of i_k and contravariance it must be that $x \in dom(R_{Id_k})$. By $i_n = i_k \circ i_1$ and contravariance we get $R_{i_n}(a) = R_{i_1}(R_{i_k}(a))$, which is tantamount to $R_{i_n}(a) = R_{i_1}(x)$, so $(x, R_{i_n}(a)) \in R_{i_1}$. By $i_{n'} = i_k \circ i_2$ and an analogous reasoning we obtain $(x, R_{i_{n'}}(a)) \in R_{i_2}$. Finally, recall that we defined $\tilde{a}(n) = R_{i_n}(a)$ and $\tilde{a}(n') = R_{i_{n'}}(a)$: substituting in what we just obtained we get that there is an $x \in dom(R_{Id_k})$ such that $(x, \tilde{a}(n)) \in R_{i_1}$ and $(x, \tilde{a}(n')) \in R_{i_2}$. By the construction of H, this is the definition of $(\tilde{a}(n), \tilde{a}(n')) \in R_k$, that is, there is an edge labeled by k between $\tilde{a}(n)$ and $\tilde{a}(n')$ in $H(\mathcal{T})$.

We proceed to argue that $\eta_{\mathcal{T}}$ is a functional bisimulation, i.e. a legitimate arrow in $\mathbf{TTS_{L^*}}$. We start with the forward condition. Suppose $(a,b) \in R_i$ for some inclusion $i : \bar{l} \to \bar{l}'$ in $\mathbf{L}^*$. So a is of type $\bar{l}'$ and b of type $\bar{l}$. The effect of the function R_i after the application of TGH is that $\eta_{\mathcal{T}}(a) = (\tilde{a}, \bar{l}')$ is mapped to the precomposition $(\tilde{a} \circ i, \bar{l})$ where $\tilde{a} \circ i : \bar{l} \to H(\mathcal{T})$. On the other hand, b is mapped by $\eta_{\mathcal{T}}$ to $(\tilde{b}, \bar{l})$. The two results are indeed the same function: for $n \in \bar{l}$, let $y = i(n)$, then $\tilde{b}(n) = R_{i_n}(b) = R_{i_n}(R_i(a)) = R_{i_y}(a) = \tilde{a}(y) = \tilde{a}(i(n))$. The equation $R_{i_n}(R_i(a)) = R_{i_y}(a)$ is a consequence of the commutation of the following diagram in $\mathbf{L}^*$



The backward condition is proved similarly: given a, $\tilde{a}$ and $R_i(\tilde{a})$, we take $R_i(a) = b$ and repeat the proof above for $\tilde{b}$.

Now for naturality. Suppose given a functional bisimulation $Z : \mathcal{T}_1 \to \mathcal{T}_2$. We show the commutation of the diagram



Given a such that $(a, a) \in R_{Id_{\bar{l}}}$, we have by definition $\eta_{\mathcal{T}_1} = (\tilde{a}, \bar{l})$. Applying $TGH(Z)$ we get the postcomposition $H(Z) \circ \tilde{a} : \bar{l} \to H(\mathcal{T}_1) \to H(\mathcal{T}_2)$. Since $H(Z) = Z$, this simplifies to $(Z \circ \tilde{a}, \bar{l})$. For a given input $n \in \bar{l}$, $Z(\tilde{a}(n)) = Z(R^1_{i_n}(a))$. We indicate with $R^1_{i_n}$ the relation in $\mathcal{T}_1$. On the other side of the diagram we have $\eta_{\mathcal{T}_2}(Z(a)) = \widetilde{Z(a)} : \bar{l} \to H(\mathcal{T}_2)$. For a given input $n \in \bar{l}$, $\widetilde{Z(a)}(n) = R^2_{i_n}(Z(a))$. Since Z is a functional bisimulation we have $R^2_{i_n}(Z(a)) = Z(R^1_{i_n}(a))$, thus we conclude that $\widetilde{Z(a)}(n) = R^2_{i_n}(Z(a)) = Z(R^1_{i_n}(a)) = Z(\tilde{a}(n))$. Hence the two functions coincide and the diagram commutes.

Finally, we show that η works as the unit of the adjunction. Suppose given a functional bisimulation $Z : \mathcal{T} \to TG(\mathscr{T})$, where $\mathcal{T}$ is a TTS and $\mathscr{T}$ is a LTS. Define $\Theta(Z) : H(\mathcal{T}) \to \mathscr{T}$ as follows. For $n \in dom(R_{Id_\varepsilon})$, put $\Theta(Z)(n) = \pi_1(Z(n))(\star)$, that is, take the pair $Z(n) = (i_n : \varepsilon \to \mathscr{T}, \varepsilon)$ (that lives in $TG(\mathscr{T})$) and obtain a point in $\mathscr{T}$ by applying the first component of $Z(n)$ to the only object in ε, namely $\star$. We now show that $Z = TG(\Theta(Z)) \circ \eta_{\mathcal{T}}$. Consider $a \in dom(R_{Id_{\bar{l}}})$. Applying the definitions we get: $TG(\Theta(Z))(\eta_{\mathcal{T}}(a)) = TG(\Theta(Z))((\tilde{a}, \bar{l})) = (\Theta(Z) \circ \tilde{a}, \bar{l})$. The first item is a function of type $\bar{l} \to H(T) \to \mathscr{T}$; the pair lives in $TG(\mathscr{T})$. We claim it coincides with $Z(a)$, which is also a pair $(f, \bar{l})$ where $f : \bar{l} \to \mathscr{T}$, we only need to check that $\Theta(Z) \circ \tilde{a}$ and f are the same function.

Let $R_{i_n}(a) = n'$. Since Z is a functional bisimulation, we have that the function $Z(a)$ is mapped to $Z(n')$ by the corresponding of the relation R_{i_n} in $TG(\mathscr{T})$: the latter is just precomposition with $i_{n'}$, thus $Z(n') = (f \circ i_{n'}, \varepsilon)$. Now we can see that, for any $n \in \bar{l}$, $\Theta(Z) \circ \tilde{a}(n) = \Theta(Z)(R_{i_n}(a)) = \Theta(Z)(R_{i_n}(a)) = \Theta(Z)(n) = \pi_1(Z(n))(\star) = f \circ i_{n'}(\star) = f(n)$. This shows that the two functions coincide, hence the natural transformation η has the universal mapping property of the unit. □

Note that this adjunction is not an equivalence of categories. LTSs constitute an optimized version of the corresponding TTSs, in the sense that the latter may contain redundant information: a TTS could contain two objects of type $l \in L$ that have the same source a and target b. After applying the functor H these two objects generate the same edge (a, b), i.e., the functor H identifies TTSs that have multiple copies of the same edges.

2.5 A logic for TTSs

Presheaves have been used as a semantics of first-order modal logics, e.g. in [59,60], and extensions of intuitionistic and modal logic, see [86]. We have seen that presheaves are equivalent to TTSs, when the base category is fixed, thus we may interpret some of these logics onto TTSs. However, none of these authors was interested in a language with modalities that explicitly referred to the arrows of the category. We on the other hand want a language that captures typed processes: since the typing of the processes is encoded in the arrows of our background category **C** and the arrows are the labels of the transitions, the natural choice is a language with a modality for each arrow. This means that such logic is parametric on a given category **C**, just like TTSs were defined parametrically on a category **C**. We use $\mathbf{LTTS^C}$ to refer to the logic of TTSs for **C**. Its language is a fragment of the so called *path logic* from [74]; we will contrast this logic with path logic in the next chapter, where the latter is introduced in full details.

Given a set of atomic propositions At, define the formulas $\mathcal{F}_{\mathbf{LTTS^C}}$ as:

$$\phi ::= \quad p \,|\, \neg\phi \,|\, \phi \wedge \phi \,|\, \langle f \rangle \phi$$

where $p \in At$ and $f \in \mathbf{C}_1$. Suppose $|\mathbf{C}_1| = \kappa$. If κ is finite then $\mathcal{F}_{\mathbf{LTTS^C}}$ has size ω, otherwise it has size κ.

A *model* for this logic is a tuple $\mathcal{M} = \langle W, \mathbf{C}, \{R_f\}_{f \in \mathbf{C}_1}, V \rangle$, where the first three items constitute a TTS labeled by **C** and $V : At \to \wp W$ is a valuation. We define the satisfaction of the formulas as usual for the propositional case, while for the modalities we put:

- $\mathcal{M}, w \vDash \langle f \rangle \phi \quad \text{iff} \quad \exists (w, w') \in R_f \wedge \mathcal{M}, w' \vDash \phi$

When Γ is a set of formulas in $\mathcal{F}_{\mathbf{LTTS^C}}$ we write $\Gamma \vDash \phi$ to mean that, for all models $\mathcal{M}$, if $\mathcal{M}, w \vDash \psi$ for all $\psi \in \Gamma$ then $\mathcal{M}, w \vDash \phi$.

We are interested in axiomatizing typed transition systems in such a language. The main difference with standard Kripke models, beside the functionality of the transitions, are two key facts:

- the transitions are type-dependent, that is, a transition only connects worlds of a given type to worlds of another given type, according to $\mathcal{C}$;
- every element in a model has a unique type.

The former feature can be captured by a set of axioms that is parametric on $\mathcal{C}$. The axioms for the logic $\mathbf{LTTS^C}$ are:

1. tautologies of classical propositional logic
2. K-axioms for all $f \in \mathcal{C}_1$
3. *Dual*-axioms for all $f \in \mathcal{C}_1$
4. $\langle g \circ f\rangle p \leftrightarrow \langle g\rangle\langle f\rangle p$ for all $g \circ f \in \mathcal{C}_1$ (axiom for composition)
5. $\langle f\rangle p \to [f]p$ for all $f \in \mathcal{C}_1$ (partial functions)
6. $p \to [Id_C]p$ for all $C \in \mathcal{C}_0$ (partial identities)
7. $\neg(\langle Id_C\rangle\top \wedge \langle Id_{C'}\rangle\top)$ for all $C, C' \in \mathcal{C}_0, C \neq C'$
   (every object has at most one type)
8. $\langle f\rangle\top \leftrightarrow \langle Id_C\rangle\top$ for all $f \in \mathcal{C}_1$ and such that $f : C' \to C$
   (actions respect types I)
9. $\langle f\rangle\top \to \langle f\rangle\langle Id_{C'}\rangle\top$ for all $f \in \mathcal{C}_1$ such that $f : C' \to C$
   (actions respect types II)

where $[f]$ is defined as $\neg\langle f\rangle\neg$. The problematic aspect is enforcing the existence of a type for each world, there could be infinitely many objects in $\mathcal{C}_0$ and thus infinitely many types. In order to deal with infinitely many types we introduce an infinitary inference rule that excludes the possibility of a world without a type. The inference rules of the calculus are Modus Ponens, Uniform Substitution and Generalization for all the modalities involved plus the following infinitary rule:

$$Rule - Id\,\frac{\langle Id_C\rangle\top \to \phi \quad \text{for all } C \in \mathcal{C}_0}{\phi}$$

with the proviso that ϕ is a non-modal formula, that is, it does not contain any modal operator. We will see later in this section that despite the infinitary rule the logic $\mathbf{LTTS^C}$ is rather well-behaved.

A $\mathbf{LTTS^C}$-*proof* of ϕ from Γ is defined as a pair (α, g) such that α is an ordinal and $g : \alpha \to \mathcal{F}_{\mathbf{LTTS^C}}$ with the following property: if $g(\beta) = \phi$, for $\beta \leq \alpha$, then ϕ is either an instance of one of the axioms, a premise in Γ or it is derived by means of an application of an inference rule to the formulas indexed by smaller ordinals, that is, indexed by γ's such that $\gamma < \beta$. We write $\Gamma \vdash_{\mathbf{LTTS^C}} \phi$ if there is a $\mathbf{LTTS^C}$-proof of ϕ from Γ; if ϕ is proved without the use of any premise we call it a *theorem* and use the notation $\vdash_{\mathbf{LTTS^C}} \phi$.

The logic $\mathbf{LTTS^C}$ is *sound* if, for any Γ and ϕ,

$$\Gamma \vdash_{\mathbf{LTTS^C}} \phi \quad \Rightarrow \quad \Gamma \models \phi$$

and *strongly complete* if, for any Γ and ϕ,

$$\Gamma \models \phi \quad \Rightarrow \quad \Gamma \vdash_{\mathbf{LTTS^C}} \phi$$

A logic is said to be *weakly complete* if the latter statement holds for $\Gamma = \emptyset$.

2.5.1. Theorem. *The logic* $\mathbf{LTTS^C}$ *is sound with respect to the class of TTSs arising from* $\mathbf{C}$*.*

Proof:
Assume Γ holds in a model and suppose we have a proof $\Gamma \vdash_{\mathbf{LTTS^C}} \phi$. The proof proceeds as usual by induction on the length of the derivation: let α be such length and suppose the claim holds for $\beta < \alpha$. By definition, ϕ can be an axiom, a premise or the result of the application of an inference rule. If ϕ is a premise in Γ then we are done. We inspect the axioms.

The first three axioms are valid by standard results, so we only check the remaining ones. Suppose given a model $\mathcal{M}$ based on a typed transition system $\mathcal{F} = \langle W, \{R_i\}_{i \in \mathbf{C}_1} \rangle$. Then Axiom 4 is valid by the fact that $R_{g \circ f} = R_g; R_f$, where the latter symbol is relational composition. Axiom 5 is valid by the partial functionality of the relations in $\mathcal{F}$ and Axiom 6 because the relations Id_C are pseudo-identities. Axiom 7 is ensured by the fact that the domains of the relations Id_C form a partition of W.

For Axiom 8 and 9, suppose that $f : C' \to C$. Since we are in a typed transition system the first conjunct is true at every world, since the domain of R_{Id_C} and R_f coincide. If $\langle f \rangle \top$ is true at a world w then there exists w' such that $(w, w') \in R_f$. By the second item of condition 4 we know that $(w', w') \in R_{Id_{C'}}$, and thus w makes true the formula $\langle f \rangle \langle Id_{C'} \rangle \top$.

For the inference rules, Modus Ponens, Uniform Substitution and Generalization work just like in the standard argument. For $Rule - Id$, suppose in a world w the formula $\langle Id_C \rangle \top \to \phi$ is true for all $C \in \mathbf{C}_0$. We know that in each world w a formula $\langle Id_C \rangle \top$ will be true for exactly one C. Hence ϕ will be true in w. $\square$

We now go on to prove a strong completeness result for $\mathbf{LTTS^C}$ for the class of TTSs arising from $\mathbf{C}$. The proof follows the routine argument employing a Lindenbaum construction and the canonical model. We highlight where our argument differs from the standard proof. The reader with a background in infinitary modal logic may be surprised by this result, since completeness proofs for such logics typically require more involved techniques such as adaptations of consistency properties.[8] The special feature of our logic is that every derivation from a consistent set of formulas can be turned into a finite derivation. The first step is to realize that every consistent set can be extended with a type, i.e., a formula $\langle Id_C \rangle \top$ for some $C \in \mathbf{C}_0$.

2.5.2. Lemma. *Every* $\mathbf{LTTS^C}$*-consistent set of formulas* Γ *can be consistently extended with a formula in the set* $\{\langle Id_C \rangle \top\}_{C \in \mathbf{C}_0}$.

Proof:
Suppose not. Then Γ entails $\neg\langle Id_C \rangle \top$ for all $C \in \mathbf{C}_0$. Therefore it also entails all the formulas in $\{\langle Id_C \rangle \top \to \bot\}_{C \in \mathbf{C}_0}$, so by the infinitary rule $Rule - Id$ we can infer that Γ entails $\bot$, which contradicts the fact that it is consistent. □

Notice that this also entails that every element in a model of the logic *must* have a type. Now Consider maximally $\mathbf{LTTS^C}$-consistent sets of formulas; we call them MCS for short.

2.5.3. Corollary. *Every MCS contains exactly one of the formulas in the set* $\{\langle Id_C \rangle \top\}_{C \in \mathbf{C}_0}$.

This is an immediate consequence of axiom 7. Now we show that if a type is added to the premises of a derivation then such derivation can be made finitary.

2.5.4. Lemma. *If* $\Gamma \vdash_{\mathbf{LTTS^C}} \phi$ *then for any* $C \in \mathbf{C}_0$ *there is a finitary proof* $\{\langle Id_C \rangle \top\} \cup \Gamma \vdash_{\mathbf{LTTS^C}} \phi$.

Proof:
Suppose $\Gamma \vdash_{\mathbf{LTTS^C}} \phi$ is of length α, where α is an ordinal, and consider a generic C. Proceeding by ordinal induction, suppose the claim holds for all $\beta < \alpha$. The formula at step α can be either an instance of one of the axioms, or a premise or it is derived from previous formulas in the sequence by means of an inference rule. In the first two cases we can immediately reduce the length of the derivation to 1: if ϕ is an instance of an axiom or a premise we can just introduce it and we get a proof from $\{\langle Id_C \rangle \top\} \cup \Gamma$.

If the formula at step α has been introduced via MP then the two premises must appear in the sequence before step α. The first of the two premises, call it ϕ_1, will be associated to an ordinal, say β_1. By induction hypothesis the proof up to β_1 can be turned into a finitary proof of length n of ϕ_1 from $\{\langle Id_C \rangle \top\} \cup \Gamma$. Similarly we can get a proof of length m of the second premise ϕ_2. Combining these proofs and applying MP we obtain a proof of length $n + m + 1$ of ϕ from $\{\langle Id_C \rangle \top\} \cup \Gamma$. A similar argument works for the other finitary rules.

If the formula at step α has been introduced via $Rule - Id$ then we know that the premise $\langle Id_C \rangle \top \to \phi$ must occur at some point in the sequence before α, say

[8] See for example [107].

at step β. By induction hypothesis we know that we can get a finitary proof of $\langle Id_C\rangle\top \to \phi$ from $\{\langle Id_C\rangle\top\} \cup \Gamma$; assume the length of such proof is n. Then we can directly introduce $\langle Id_C\rangle\top$, which is now one of the premises, and apply MP to obtain ϕ in a derivation of length $n+2$. $\square$

These two lemmas are key to the success of the Lindenbaum Lemma: given a consistent set of formulas Γ, we build a MCS by first adding a type to it and then performing the usual inductive construction on the rest of the formulas; due to Lemma 2.5.4, all derivations from the consistent sets that we build will be finitary, thanks to the presence of the type.

2.5.5. Lemma. *Every consistent set of formulas can be extended to a MCS of* $\mathbf{LTTS^C}$.

Proof:
Given a consistent set of formulas Γ, label the formulas of the language with the elements of κ in the following way. First, pick any ordering of the formulas (e.g. lexicographic). Second, shift the ordering so that the formulas in the set $\{\langle Id_C\rangle\top\}_{C\in\mathbf{C}_0}$ appear first. So if $|\mathbf{C}_0| = \kappa'$ we will have that the first κ' formulas are all the formulas $\{\langle Id_C\rangle\top\}_{C\in\mathbf{C}_0}$. Notice that by Lemma 2.5.2 there exists a C such that $\langle Id_C\rangle\top$ is consistent with Γ. Take the first formula $\langle Id_C\rangle\top$ in the ordering that is consistent with Γ and define $\Gamma_{\kappa'} = \Gamma\cup\{\langle Id_C\rangle\top\}\cup\{\neg\langle Id_{C'}\rangle\top\}_{C'\neq C}$.

Then for $\kappa' \leq \alpha$ proceed as follows:

- for the successor step put $\Gamma_{\alpha+1} = \Gamma_\alpha \cup \{\phi_\alpha\}$ if it is $\mathbf{LTTS^C}$-consistent, or $\Gamma_\alpha \cup \{\neg\phi_\alpha\}$ otherwise;
- for the limit step put $\Gamma_\alpha = \bigcup_{\beta<\alpha}\Gamma_\beta$, for $\alpha = \bigcup_{\beta<\alpha}\beta$ and $\kappa' \leq \beta$.

Finally, take $\Gamma_\kappa = \bigcup_{\alpha<\kappa}\Gamma_\alpha$. We now show by ordinal induction that Γ_α is consistent for all α such that $\kappa' \leq \alpha \leq \kappa$, that is, we start the induction from κ'. The base step is ensured by Lemma 2.5.2 and axiom 7: $\Gamma \cup \{\langle Id_C\rangle\top\}$ is consistent by the Lemma and the axiom allows us to deduce every formula in the set $\{\neg\langle Id_{C'}\rangle\top\}_{C'\neq C}$. The successor step is given. For the limit step, suppose $\Gamma_\alpha = \bigcup_{\beta<\alpha}\Gamma_\beta$ is inconsistent. Thus there are $\Sigma \cup \{\psi\} \subseteq \bigcup_{\beta<\alpha}\Gamma_\beta$ such that $\Sigma \vdash_{\mathbf{LTTS^C}} \neg\psi$. Say ψ has been added at step γ.

Since $\Gamma_{\kappa'} \subseteq \Gamma_\alpha$ we can conclude that $\{\langle Id_C\rangle\top\} \cup \Sigma \vdash_{\mathbf{LTTS^C}} \neg\psi$ is still a derivation from premises in Γ_α. By Lemma 2.5.4 we know that there is a finitary proof $\{\langle Id_C\rangle\top\} \cup \Sigma \vdash_{\mathbf{LTTS^C}} \neg\psi$. Since only finitely many premises can be used in a finite derivation, this entails that $\{\langle Id_C\rangle\top\} \cup \Sigma' \vdash_{\mathbf{LTTS^C}} \neg\psi$ with Σ' finite subset of Σ. Hence all the premises in $\{\langle Id_C\rangle\top\} \cup \Sigma'$ appear at some stage before the limit, call it β'. From this we can conclude that either Γ_γ or Γ'_β are already inconsistent, which contradicts the IH. We can thus conclude that Γ_κ is a MCS extending Γ. $\square$

Consider the relations between MCS as usually defined in the canonical model:

$$R_f\Delta\Delta' \text{ iff, for every formula } \phi,\ \phi \in \Delta' \text{ entails } \langle f\rangle\phi \in \Delta$$

We write R_f to indicate the relation associated to $f \in \mathbf{C}_1$.

2.5.6. LEMMA (EXISTENCE). *For all $f \in \mathbf{C}_1$ and all MCS Δ, if $\langle f\rangle\psi \in \Delta$ then there is an MCS Δ' such that $R_f\Delta\Delta'$ and $\psi \in \Delta'$.*

Proof:
Let f have the typing $f : C \to C'$. Consider $\{\psi\} \cup \{\theta | [f]\theta \in \Delta\}$. Suppose it is inconsistent. Then $\Phi = \{\theta | [f]\theta \in \Delta\} \vdash_{\mathbf{LTTS^C}} \neg\psi$. Since $\langle f\rangle\psi \in \Delta$ we know that $\langle f\rangle\top \in \Delta$, so by axiom 8 $\langle f\rangle\langle Id_C\rangle\top \in \Delta$. By axiom 5 we have $[f]\langle Id_C\rangle\top \in \Delta$ and therefore by definition $\langle Id_C\rangle\top \in \Phi$.

By Lemma 2.5.4 we know that there is a proof of $\neg\psi$ from finitely many premises $\phi_1, \ldots, \phi_n$ in Φ. We can then run the usual argument: applying necessitation and axiom K to the implication $\phi_1 \wedge \cdots \wedge \phi_n \to \neg\psi$ we obtain a contradiction with the fact that $\langle f\rangle\psi \in \Delta$.

Since $\{\psi\} \cup \{\theta | [f]\theta \in \Delta\}$ is consistent, by the previous Lemma we can take the MCS that extends it; such MCS satisfies the required conditions by construction. □

The Truth Lemma works as usual: we derive that in the canonical model, for any MCS Δ and any formula ϕ, $\Delta \vDash \phi$ iff $\phi \in \Delta$.

2.5.7. THEOREM. *The logic $\mathbf{LTTS^C}$ is strongly complete with respect to the class of typed transition systems arising from $\mathbf{C}$.*

Proof:
The proof of completeness is the standard one: given a consistent set of formulas we extend it to a MCS and by the Truth Lemma we know such MCS satisfies the formulas of the set.

We check that the canonical model is indeed an typed transition systems labeled by $\mathbf{C}$. Consider the canonical model $\mathcal{M}_{\mathbf{LTTS^C}} = \langle W, \{R_f\}_{f \in \mathbf{C}_1}, V\rangle$. The axioms 1,2,3,5 and 6 are canonical for the corresponding properties by standard results. So the canonical model is deterministic and satisfies conditions 1 of typed transition systems.

Consider a MCS Δ. By Corollary 2.5.3 we know that there is a $C \in \mathbf{C}_0$ such that $\langle Id_C\rangle\top \in \Delta$, while by Axiom 7 we know that Δ can have at most one type. Therefore every MCS has one and only one type. This takes care of condition 2. For condition 3, suppose that $f : C' \to C$. If $(\Delta, \Delta) \in R_{Id_C}$ then by the semantics and axiom 8 we must have $\langle f\rangle\top \in \Delta$, so by the Truth Lemma there is Δ' such that $(\Delta, \Delta') \in R_f$. The converse works analogously, hence the domains of f and Id_C coincide. The remaining implication also holds due to axiom 9.

For the last property, suppose $(\Delta, \Delta') \in R_{g\circ f}$. Hence for every formula ϕ if $\phi \in \Delta'$ then $\langle g \circ f\rangle\phi \in \Delta$. Due to axiom 4, the latter is the case iff $\langle g\rangle\langle f\rangle\phi \in \Delta$.

Since $\top \in \Delta'$, we obtain $\langle g \rangle \langle f \rangle \top \in \Delta$. By Lemma 2.5.6 we know that this entails the existence of a MCS Δ'' such that $(\Delta, \Delta'') \in R_g$ and $\langle f \rangle \top \in \Delta''$. So by assuming $\phi \in \Delta'$ we obtain $\langle g \rangle \langle f \rangle \phi \in \Delta$, which by axiom 5 entails $[g]\langle f \rangle \phi \in \Delta$. The latter together with $(\Delta, \Delta'') \in R_g$ allows us to conclude that $\langle f \rangle \phi \in \Delta''$. Therefore by definition we can conclude $(\Delta'', \Delta') \in R_f$. This shows that $R_{g \circ f} \subseteq R_g; R_f$. The converse is proved directly with the other direction of axiom 4. Hence $R_{g \circ f} = R_g; R_f$ and the canonical model satisfies the last property of TTSs. This is enough to establish that $\mathcal{M}_{\mathbf{LTTS}^{\mathbf{C}}}$ is a typed transition system labeled by $\mathbf{C}$. □

2.6 A finitary logic for TTSs

If the set of objects of the category $\mathbf{C}$ is finite then the infinitary rules become finitary, in which case we have soundness and completeness for a finitary calculus. Now the question is: can we design a finitary calculus also for the categories with infinitely many objects?

Consider the same language and the same satisfaction relation. The axioms and the rules for the logic $\mathbf{LTTS}^{\mathbf{C}}_{fin}$ are the same as before, except for the infinitary rule which is now absent. Soundness follows from the earlier proof.

2.6.1. THEOREM. *The logic $\mathbf{LTTS}^{\mathbf{C}}_{fin}$ is sound for the class of typed transition systems labeled by $\mathbf{C}$.*

Proof:
As usual by induction on the length of the proof. The base cases of the axioms are the same as those treated in the soundness of the infinitary logic. Modus Ponens, Substitution and Necessitation are dealt with as usual. □

We remark that, even though the infinitary rule is missing, another similar finitary rule is admissible in the system.

2.6.2. LEMMA. *The following rule is admissible in $\mathbf{LTTS}^{\mathbf{C}}_{fin}$, for any $C \in \mathbf{C}_0$:*

$$Rule - Id_{fin} \frac{\vdash \langle Id_C \rangle \top \to \phi}{\vdash \phi}$$

again with the proviso that ϕ is not a modal formula.

Proof:
Suppose $\vdash \langle Id_C \rangle \top \to \phi$ is the case, that is, we have a derivation of $\langle Id_C \rangle \top \to \phi$ from the axioms. There can be two cases. Suppose that ϕ is a propositional tautology. In this case we have $\vdash \phi$ directly by the propositional part of the calculus. If ϕ is not a propositional tautology then there exists a classical valuation $V : At \to \{0, 1\}$ that falsifies the formula: since ϕ is a non-modal formula, its truth depends solely on the valuation. Consider a frame having as carrier the

set $W = \{w_C | C \in \mathbf{C}_0\}$, namely the set having only one object for every type. We thus have $R_{Id_C} = \{(w_C, w_C)\}$ for every $C \in \mathbf{C}_0$. The partial functions on this carrier are completely determined, since all the domains and codomains are singletons. It is a straightforward to check that this is indeed a TTS.

On top of this frame build a model $\mathcal{M}$ with a valuation V' such that $w_C \in V'(p)$ iff $V(p) = 1$, that is, every world has the same valuation V. By construction we have that $\mathcal{M}, w_C \models \langle Id_C \rangle \top$ but $\mathcal{M}, w_C \not\models \phi$. This contradicts the assumption $\vdash \langle Id_C \rangle \top \to \phi$ and soundness, therefore ϕ must be a propositional tautology. □

With the help of such admissible rule we can show that every theorem of $\mathbf{LTTS^C}$ is also a theorem of $\mathbf{LTTS}^{\mathbf{C}}_{fin}$, this allows us to infer the weak completeness of $\mathbf{LTTS}^{\mathbf{C}}_{fin}$.

2.6.3. THEOREM. *The logic* $\mathbf{LTTS}^{\mathbf{C}}_{fin}$ *is weakly complete for the class of typed transition systems labeled by* $\mathbf{C}$.

Proof:
Suppose ψ is a theorem of $\mathbf{LTTS^C}$, we prove that ψ is a theorem of $\mathbf{LTTS}^{\mathbf{C}}_{fin}$ by induction on the length of the derivation $\vdash_{\mathbf{LTTS^C}} \psi$. Let α be the length and suppose that for all $\beta < \alpha$ the claim is proved.

If at step α the formula ψ is introduced as an axiom or it is proved by one of the finitary rules (MP, Necessitation, Substitution) then the same step can be copied in $\mathbf{LTTS}^{\mathbf{C}}_{f}$ so together with the IH we have that $\vdash_{\mathbf{LTTS^C}{fin}} \psi$. Consider the case in which the step α is an application of the rule

$$Rule - Id \, \frac{\langle Id_C \rangle \top \to \psi \quad \text{for all } C \in \mathbf{C}_0}{\psi}$$

Since such proof has no other premises except the axioms, in order to apply $Rule - Id$ at step α it must be that $\vdash_{\mathbf{LTTS^C}} \langle Id_C \rangle \top \to \psi$ appears in some step before α, for every $C \in \mathbf{C}_0$. Consider the first appearance of those formulas, say at step $\beta < \alpha$. From IH on β we can deduce that there is a proof $\vdash_{\mathbf{LTTS}^{\mathbf{C}}_{fin}} \langle Id_C \rangle \top \to \psi$ in the finitary logic. But then we can apply rule $Rule - Id_{fin}$ directly at step β to obtain a proof $\vdash_{\mathbf{LTTS}^{\mathbf{C}}_{fin}} \psi$. This concludes the induction and proves the claim.

From these considerations it follows that if ψ is not a theorem of $\mathbf{LTTS}^{\mathbf{C}}_{fin}$ then it is not a theorem of $\mathbf{LTTS^C}$. By the completeness of the latter calculus, there is a TTS labeled by $\mathbf{C}$ that refutes the formula, hence $\mathbf{LTTS}^{\mathbf{C}}_{fin}$ is weakly complete. □

2.7 Coalgebraic perspective on TTSs

It is a known fact in the literature that the category of presheaves can be equivalently described as a particular category of coalgebras; see [5] for a description of the construction and [71] p.900 for the genesis of the main ideas. Coalgebras come with associated logics, known as *coalgebraic logics*, thus it is natural to wonder what is the connection between $\mathbf{LTTS^C}$ and the coalgebraic logic arising from the class of TTSs seen as coalgebras. We will see that $\mathbf{LTTS^C}$ is indeed a coalgebraic logic *à la* Pattinson, when we generalize the setting of coalgebraic logic to accommodate typed structures.

2.7.1 Presheaves as coalgebras

Given an endofunctor $E : \mathbf{C} \to \mathbf{C}$ on a category $\mathbf{C}$, a coalgebra for E is a pair $(C, \xi : C \to E(C))$, where C is an object of $\mathbf{C}$ and $\xi : C \to E(C)$ a morphism. When $\mathbf{C}$ is the category $\mathbf{Set}$ of sets and functions, the coalgebras for an endofunctor $E : \mathbf{Set} \to \mathbf{Set}$ are pairs $(X, \xi : X \to E(X))$, where X is a set and ξ a function.

Now consider $\mathbf{Set}^S$, the category consisting of S-indexed families of sets and S-indexed families of functions: an object of this category is a family $\{X_s\}_{s\in S}$ and an arrow is $\{f_s\}_{s\in S} : \{X_s\}_{s\in S} \to \{Y_s\}_{s\in S}$ such that $f_s : X_s \to Y_s$ for every $s \in S$. One obtains 'many-sorted' or 'typed' coalgebras by replacing sets with S-indexed sets: for an endofunctor $E : \mathbf{Set}^S \to \mathbf{Set}^S$ we take coalgebras to be pairs $(\{X_s\}_{s\in S}, \xi : \{X_s\}_{s\in S} \to E(\{X_s\}_{s\in S}))$.

It is shown in [5] (improving on the results of [125]) that presheaves over the base category $\mathbf{C}$ can be seen as coalgebras for the endofunctor $\Xi : \mathbf{Set}^{\mathbf{C}_0} \to \mathbf{Set}^{\mathbf{C}_0}$ defined as follows:[9]

$$\Xi(\{X_C\}_{C\in\mathbf{C}_0}) = \{\prod_{C'\in\mathbf{C}_0} X_{C'}^{Hom_{\mathbf{C}}(C',C)}\}_{C\in\mathbf{C}_0}$$

This functor acts on arrows by taking a family $\{f_C\}_{C\in\mathbf{C}_0} : \{X_C\}_{C\in\mathbf{C}_0} \to \{Y_C\}_{C\in\mathbf{C}_0}$ and returning a family $\Xi(\{f_C\}_{C\in\mathbf{C}_0})$ such that its component C maps an input $\langle succ_{C'}, succ_D, \dots\rangle \in \prod_{C'\in\mathbf{C}_0} X_{C'}^{Hom_{\mathbf{C}}(C',C)}$ (where $succ_{C'} : Hom_{\mathbf{C}}(C', D) \to X'_C$ and so on) to an output $\langle f_{C'} \circ succ_{C'}, f_D \circ succ_D, \dots\rangle \in \prod_{C'\in\mathbf{C}_0} Y_{C'}^{Hom_{\mathbf{C}}(C',C)}$.

The core idea behind this construction is that each presheaf over $\mathbf{C}$ is described by the images of the objects plus the specification of the functions associated to the arrows. The former is encoded in a family such as $\{X_C\}_{C\in\mathbf{C}_0}$, the latter is captured by the coalgebra structure for the functor Ξ: for any element $x \in X_C$, $\xi(x)$ returns an element in $\prod_{C'\in\mathbf{C}_0} X_{C'}^{Hom_{\mathbf{C}}(C',C)}$, that is, for every arrows of type $f : C' \to C$ it returns the image of x under the function associated to f (which is flipped because of contravariance). Hence an element in $\prod_{C'\in\mathbf{C}_0} X_{C'}^{Hom_{\mathbf{C}}(C',C)}$ can be thought of as the collection of all the successors for an element of type C, where said successors might be of different types. A presheaf (and therefore a TTS) will thus look like a coalgebra $(\{X_C\}_{C\in\mathbf{C}_0}, \{\xi_C\}_{C\in\mathbf{C}_0})$. We refer the reader to the literature for the details of the equivalence.

[9] Recall that we are considering the base category to be small.

2.7.2 Coalgebraic logic for TTSs

In the influential paper [100], Pattinson introduced the idea of extracting a modal logic for the coalgebras of a functor $E : \mathbf{Set} \to \mathbf{Set}$ from a collection of natural transformations of type $E \to \wp$, where the latter is the covariant powerset functor.[10] More precisely, given $E : \mathbf{Set} \to \mathbf{Set}$, $\mu : E \to \wp$, a coalgebra $(X, \xi : X \to E(X))$ and $x \in X$ we can define a modal operator $[\mu]$ with semantics

$$(X, \xi), x \vDash [\mu]\phi \text{ iff } (X, \xi), y \vDash \phi \text{ for all } y \in \mu_X(\xi(x))$$

For example, when E is the covariant powerset functor then a coalgebra is equivalent to an ordinary Kripke frame and $\mu_X(\xi(x))$ is the set of successors of x in X. With an analogous methodology we can capture atomic propositions as given by a natural transformation $\nu : E \to \wp(At)$, where At is a fixed set of atomic propositions and $\wp(At)$ is the constant functor mapping every set to $\wp(At)$. Then

$$(X, \xi), x \vDash p \text{ iff } p \in \nu_X(\xi(x))$$

In order to apply this idea to the coalgebras associated to presheaves we must first generalize this idea to 'typed' coalgebras. The first observation concerns the generalization of the powerset functor. Given a family $\{X_C\}_{C \in \mathbf{C}_0}$, a first option would be to apply the powerset functor component-wise, defining $\wp^{\mathbf{C}_0}(\{X_C\}_{C \in \mathbf{C}_0}) = \{\wp(X_C)\}_{C \in \mathbf{C}_0}$. But this solution falls short when we couple it with the fact that arrows in $\mathbf{Set}^{\mathbf{C}_0}$ act component-wise. Suppose $x \in X_C$, then the tuple $\xi^C(x) \in \prod_{C' \in \mathbf{C}_0} X_{C'}^{Hom_{\mathbf{C}}(C',C)}$ describes the successors of x along the functions indexed by the arrows of $\mathbf{C}$. With the current definition of $\wp^{\mathbf{C}_0}$, $\mu_X^C(\xi^C(x)) \in \wp(X_C)$ is a subset of X_C, which is at odds with the fact that the successors of x might have different types.

We thus propose to generalize the powerset construction in a different way: for $\{X_s\}_{s \in S}$ family of sets, define $\wp^S(\{X_s\}_{s \in S}) = \{\prod_{s' \in S} \wp(X_{s'})\}_{s \in S}$; now at each s-component we have a product of all the powersets of the elements of the family. For an arrow $\{f_s\}_{s \in S} : \{X_s\}_{s \in S} \to \{Y_s\}_{s \in S}$ the image under $\wp^S$ is a family of functions such that at component s the function

$$(\wp^S(\{f_s\}_{s \in S}))_s : \prod_{s' \in S} \wp(X_{s'}) \to \prod_{s' \in S} \wp(Y_{s'})$$

sends a tuple $\langle A_s, A_{s'}, \dots \rangle$, where $A_s \subseteq X_s$, $A_{s'} \subseteq X_{s'}$ and so on, to the tuple $\langle f_s(A_s), f_{s'}(A_{s'}), \dots \rangle$.

This alternative definition of $\wp^{\mathbf{C}_0}$, where we have replaced S with the set of 'types' given by the objects of the category, allows us to close the gap and extract a coalgebraic logic following Pattinson's technique. For $f : C' \to C$ an arrow in $\mathbf{C}$,

[10] See [78] for an overview of coalgebraic logic.

we construct a natural transformation $\mu[f] : \Xi \to \wp^{\mathbf{C}_0}$ along the following lines. Given a family $\overline{X} = \{X_C\}_{C \in \mathbf{C}_0}$ and a component $D \in \mathbf{C}_0$, define

$$\mu^D_{\overline{X}}[f] : \prod_{C' \in \mathbf{C}_0} X_{C'}^{Hom_{\mathbf{C}}(C',D)} \to \prod_{C' \in \mathbf{C}_0} \wp(X_{C'})$$

on the input $\langle succ_{C'}, succ_D, \dots \rangle$, where $succ_{C'} : Hom_{\mathbf{C}}(C', D) \to X_{C'}$ and so on, as follows:

$$\begin{aligned} \mu^D_{\overline{X}}[f](\langle succ_{C'}, \dots, succ_{D'}, \dots \rangle) = &\langle succ_{C'}[Hom_{\mathbf{C}}(C', D) \cap \{f\}], \dots \\ &\dots, succ_{D'}[Hom_{\mathbf{C}}(D', D) \cap \{f\}], \dots \rangle \end{aligned}$$

Notice that every intersection $Hom_{\mathbf{C}}(D', D) \cap \{f\}$ is non-empty only if D' and D are the domain and codomain of f, respectively. If $Hom_{\mathbf{C}}(D', D) \cap \{f\}$ is empty then $succ_{D'}[Hom_{\mathbf{C}}(D', D) \cap \{f\}]$ is also empty. Thus one of two cases must occur. If D is the codomain of f then the tuple $\langle succ_{C'}[Hom_{\mathbf{C}}(C', D) \cap \{f\}], \dots, succ_{D'}[Hom_{\mathbf{C}}(D', D) \cap \{f\}], \dots \rangle$ will consist of empty sets with the exception of the component corresponding to the domain of f, there we will have a singleton $succ_{dom(f)}(f)$. In other words, when the codomain of f coincides with the component of μ then the function returns the successor for the function labeled by f, embedded in a tuple that establishes the correct typing. If D is not the codomain of f then the tuple will consist of only empty sets. Note that the tuple on the right is indeed an element in $\prod_{C' \in \mathbf{C}_0} \wp(X_{C'})$, since the singleton $succ_{dom(f)}(f)$ is a subset of $X_{dom(f)}$.

2.7.1. PROPOSITION. *The above definition turns $\mu[f] : \Xi \to \wp^{\mathbf{C}_0}$ into a natural transformation.*

Proof:
Let $\{f_C\}_{C \in \mathbf{C}_0} : \{X_C\}_{C \in \mathbf{C}_0} \to \{Y_C\}_{C \in \mathbf{C}_0}$ be an arrow in $\mathbf{Set}^{\mathbf{C}_0}$. We need to show the commutation of the corresponding diagram for each component D.

$$\begin{array}{ccc} \prod_{C' \in \mathbf{C}_0} X_{C'}^{Hom_{\mathbf{C}}(C',D)} & \xrightarrow{\mu^D_{\overline{X}}[f]} & \prod_{C' \in \mathbf{C}_0} \wp(X_{C'}) \\ {\scriptstyle \Xi(\{f_C\}_{C' \in \mathbf{C}_0})} \downarrow & & \downarrow {\scriptstyle \wp^{\mathbf{C}_0}(\{f_C\}_{C' \in \mathbf{C}_0})} \\ \prod_{C' \in \mathbf{C}_0} Y_{C'}^{Hom_{\mathbf{C}}(C',D)} & \xrightarrow[\mu^D_{\overline{Y}}[f]]{} & \prod_{C' \in \mathbf{C}_0} \wp(Y_{C'}) \end{array}$$

An object in the top-left corner is a tuple $\langle succ_{C'}, succ_D, \dots \rangle$, where $succ_{C'} : Hom_{\mathbf{C}}(C', D) \to X_{C'}$ and so on. Applying $\mu^D_{\overline{X}}[f]$ we obtain a tuple

$$\langle succ_{C'}[Hom_{\mathbf{C}}(C', D) \cap \{f\}], \dots, succ_{D'}[Hom_{\mathbf{C}}(D', D) \cap \{f\}], \dots \rangle$$

Applying $\wp^{\mathbf{C}_0}(\{f_C\}_{C'\in\mathbf{C}_0})$ to the latter we get

$$\langle f_{C'} \circ succ_{C'}[Hom_{\mathbf{C}}(C', D) \cap \{f\}], \ldots, f_{D'} \circ succ_{D'}[Hom_{\mathbf{C}}(D', D) \cap \{f\}], \ldots\rangle$$

Following the other path in the diagram, from the initial tuple we obtain

$$\langle f_{C'} \circ succ_{C'}, \ldots, f_{D'} \circ succ_{D'}, \ldots\rangle$$

by applying $\Xi(\{f_C\}_{C'\in\mathbf{C}_0})$; an application of $\mu^D_{\overline{Y}}[f]$ results in the string

$$\langle f_{C'} \circ succ_{C'}[Hom_{\mathbf{C}}(C', D) \cap \{f\}], \ldots, f_{D'} \circ succ_{D'}[Hom_{\mathbf{C}}(D', D) \cap \{f\}], \ldots\rangle$$

proving the commutation. □

Pattinson's blueprint for the semantics of modal operators can now be used successfully to define the semantics of the operator $[f]$, where f is an arrow of $\mathbf{C}$. Let $f : C' \to C$, use $(\overline{X}, \overline{\xi})$ as a shortcut for the coalgebra $(\{X_C\}_{C\in\mathbf{C}_0}, \{\xi_C\}_{C\in\mathbf{C}_0})$ and denote with (x, D) the fact that $x \in X_D$:

$$(\overline{X}, \overline{\xi}), (x, D) \vDash [f]\phi \text{ iff } (\overline{X}, \overline{\xi}), (y, C') \vDash \phi \text{ for all } y \in \pi_{C'}(\mu^D_{\overline{X}}[f](\xi_D(x)))$$

This definition encodes the fact that $[f]\phi$ is true at x iff ϕ holds at every f-successor of x. Compared to the single-sorted version, the main differences are that the elements are now typed and that we have to pick the right component of the tuple $\mu^D_{\overline{X}}[f](\xi_D(x))$ with the projection $\pi_{C'}$. The dual operator $\langle -\rangle$ will then be defined as

$$(\overline{X}, \overline{\xi}), (x, D) \vDash \langle f\rangle\phi \text{ iff there is } y \in \pi_{C'}(\mu^D_{\overline{X}}[f](\xi_D(x))) \text{ such that } (\overline{X}, \overline{\xi}), (y, C') \vDash \phi$$

When $D \neq C$ then there should be no f-successors, because the function associated to f only applies to C-objects; indeed the tuple $\mu^D_{\overline{X}}[f](\xi_D(x))$ will consist of only empty sets and thus $\langle f\rangle\top$ will be false. Note that this definition matches the one of Section 2.5 when the relation R_f is unpacked from the definition of the coalgebra structure $\overline{\xi}$.

As for atomic propositions we can again mimic the single-sorted approach, taking care to relativize to the correct type. Clearly to interpret atomic propositions we have to go from frames to models, thus add a valuation. In the single sorted case this means adding $- \times \wp(At)$ to the functor $E : \mathbf{Set} \to \mathbf{Set}$, so that each element is mapped to a set of propositions. In the multi sorted case we will follow this procedure for every type, thus for example in the case of Ξ we will have

$$\Xi(\{X_C\}_{C\in\mathbf{C}_0}) = \{(\prod_{C'\in\mathbf{C}_0} X_{C'}^{Hom_{\mathbf{C}}(C',C)}) \times \wp(At)\}_{C\in\mathbf{C}_0}$$

Finally, consider the family $\{At_C\}_{C\in\mathbf{C}_0}$ consisting of $|\mathbf{C}_0|$ copies of At and define the constant functor $\wp^{\mathbf{C}_0}(\{At_C\}_{C\in\mathbf{C}_0})$ mapping every family of sets to the family

$\{\prod_{C' \in \mathbf{C}_0} \wp(At)\}_{C \in \mathbf{C}_0}$. Given a natural transformation $\nu : \Xi \to \wp(At)$ (with the new definition of Ξ), we can define ν_X^D as

$$\nu_X^D(\langle succ_{C'}, \ldots, succ_{D'}, \ldots, Prop \rangle) = \langle \emptyset, \ldots, Prop, \ldots, \emptyset, \ldots \rangle$$

where $Prop \subseteq At$ is a subset of atomic propositions and the tuple on the right contains only empty sets with the exception of component D, where it features the set $Prop$. Now the semantics of the atomic propositions becomes

$$(\overline{X}, \overline{\xi}), (x, D) \vDash p \text{ iff } p \in \pi_D(\nu_X^D(\xi_D(x)))$$

which is again a generalization of Pattinson's definition. In conclusion, given that presheaves over $\mathbf{C}$ can be seen as coalgebras over $\mathbf{Set}^{\mathbf{C}_0}$, we have

- proposed a way to generalize the coalgebraic logic for **Set** functors to the case of $\mathbf{Set}^S$ functors, thus covering the case of presheaves-as-coalgebras;
- defined natural transformations $\mu[f]$ to capture modal operators labeled by the arrows of $\mathbf{C}$;
- shown that these operators match those of the logic $\mathbf{LTTS}^{\mathbf{C}}$.

The relation of this proposal with general coalgebraic logic, that is, for endofunctors on generic categories, remains to be investigated.

2.8 Conclusions

We started by introducing a peculiar version of transition systems, called "typed": they are deterministic LTS with a superimposed typing structure given by a small category. We have seen that the presheaf category $\mathbf{Set}^{\mathbf{C}^{op}}$ is equivalent to $\mathbf{TTS}_{\mathbf{C}}$, the category of typed transition systems labeled by $\mathbf{C}$ that has functional bisimulations as arrows. Furthermore, we have investigated the connection with standard LTSs, proving an adjunction between $\mathbf{TS}'_L$, the category of transition systems with labels in L and relation-preserving functions, and $\mathbf{TTS}_{\mathbf{L}^*}$, the category of TTSs labeled by $\mathbf{L}^*$, the subcategory of $\mathbf{TS}'_L$ consisting of only linear paths. We provided an infinitary logic that is sound and strongly complete for the class of TTSs (parametrically on $\mathbf{C}$) and also proposed a finitary version of the calculus that enjoys soundness and weak completeness. Finally, we investigated the link with Coalgebra by showing how our logic arises from the coalgebras associated to presheaves. We generalized Pattinson's technique to extract a coalgebraic logic from natural transformations, transporting it to the realm of multi-typed coalgebras; the application of this procedure is shown to yield the language and semantics of Section 2.5.

The hybrid character of TTSs makes is such that enquiries on these structures have double significance, both from a Modal Logic and from a category-theoretic perspective. For instance, the study of bisimulations for TTSs leads to categorical notions that are studied on presheaves, such as *open maps* from [74]. Another example is the investigation on the expressivity of the language we introduced; the centrality of presheaf categories can hardly be overestimated, whence the interest in capturing some of their properties with (suitable extensions of) the logic $\mathbf{LTTS}^{\mathbf{C}}$. We come back to both themes in the next chapter.

We conclude suggesting two possible developments of this line of work. A first task is to strengthen the links with other frameworks. The connection with coalgebras and coalgebraic logic could be expanded beyond what is sketched in the previous section. Another pivotal concept is that of profunctor studied in [40]: it is easy to notice that our functor $T : \mathbf{Set}^{\mathbf{C}^{op}} \to \mathbf{TTS}_{\mathbf{C}}$ is parametric on $\mathbf{C}$, thus a natural question to ask is whether we can extend this construction to a functor from the category of presheaf categories and profunctors to the category of **TTS** categories equipped with some suitable notion of morphism.

A second issue concerns the structure of the category $\mathbf{TTS}_{\mathbf{C}}$. We have seen in Section 2.3 that such category inherits many interesting features from the category of presheaves; in particular, this gives a number of concrete constructions (limits and colimits) as well as particular models such as representables. This also entails that $\mathbf{TTS}_{\mathbf{C}}$ is a topos and thus comes with an internal first order logic, hence one would like to understand how such internal logic relates to the modal logic of Section 2.5.

Chapter 3

Expressivity

3.1 Introduction

We have seen in the previous chapter that the logic of typed transition systems is complete for presheaves seen as relational structures. What remains to be shown is that the language is expressive enough to capture relevant properties of presheaves. When the answer to such questions is negative, it is also natural to wonder whether an extension of the language can give the desired result. These two issues are the focus of this chapter.

As we mentioned above, the language we introduced is a fragment of a language known as *path logic*; we thus take the existing work on path logic as the starting point of our investigation. Path logic was first formalized by the authors of the so called 'presheaf approach to Concurrency Theory', which in turn originated from the categorical outlook towards models of concurrency described in [123]. In this paper many important models such as transition systems, synchronization trees and event structures were organized into categories and systematically related via adjunctions. Upon realization that each of these models was associated to a corresponding notion of path, in the seminal paper [74] Joyal, Winskel and Nielsen devised a representation of models of concurrency in terms of presheaves over suitable 'path categories', following the intuition that a model of concurrency consists of bundles of different paths glued together in a coherent way.

This perspective unveiled the possibility to define a general categorical notion of behavioural equivalence solely in terms of path preservation and path 'lifting'. While the former is usually inbuilt in the definition of morphism of the categories under examination, the latter had to be imposed, leading to the definition of *open maps*. The desired general concept of bisimilarity was then at hand: two models of concurrency are deemed bisimilar if their presheaf representations are connected by a span of open maps. In a follow-up paper [124] it was observed that presheaves can themselves be regarded as transition system via the construction usually known as category of elements. A notion of bisimulation for these transition systems, baptized *path bisimulation*, was proved equivalent to the bisimulation in terms of span of open maps. A modal logic called *path logic* was proposed and shown to be characteristic for such path bisimulation. Given some conditions on the base category, presheaves can be thought of as generalized models of concurrency, with representables playing the role of path shapes. Path logic becomes then the natural choice of language for such models. As we already remarked, path logic is an extension of the logic presented in the previous chapter.

After reviewing some definitions and results from the literature, we start by proving a characterization result for path logic, in the fashion of van Benthem's theorem, for the case in which the background category **C** has finitely many objects. In Section 3.4 we examine some properties of presheaves from the literature and see how to encode them in path logic. The main test cases are the sheaf of section of a covering space and the recent sheaf-theoretic analysis of non-locality and contextuality pioneered by Abramsky and Brandenburger in [4]. Contextuality has proven to be a crucial feature of quantum phenomena and this line of research has since then been developed in a series of papers.[1] We will show how the pivotal concepts of that analysis can be captured by path logic; this in turn entails that such properties are invariant under path bisimulation.

A core notion that one would like to capture is that of *sheaf*, a concept that is widely used in geometry and Topology.[2] Alas, we observe that sheaves are not definable in path logic. We thus devote the rest of the chapter to the understanding of sheaves over topological spaces through their relational counterparts.

One possibility, investigated in Section 3.5, is to enhance the semantic companion of path logic, namely path bisimulation, to preserve locality and gluing, the defining properties of a sheaf. Moving the first steps in this direction, we prove some basic results concerning the apt notion of bisimulation in the context of sheaves, characterizing spans and co-spans of open maps. Another approach, developed in Section 3.6, is to add enough expressive power to define the two properties; this is achieved by adding nominals to the language. We conclude by showing that, when the category in the background is nice enough, the finitary fragment of this hybrid extension is decidable.

3.2 Preliminaries

We first review the definitions and results of the presheaf approach to Concurrency Theory by Winskel and Nielsen. Let **C** be a fixed small category.

3.2.1. Definition. A presheaf $P : \mathbf{C}^{op} \to \mathbf{Set}$ is *rooted* if **C** has an initial object, denoted with 0, and $P(0)$ is a singleton. The unique object in $P(0)$ is called the *root* and will be denoted by r.

[1] We will especially refer to [3, 4, 75].

[2] See [85] for a classic text.

Note that, due to the universal property of the initial objects, all representable presheaves are rooted. Given a cardinal κ and a set At of propositional variables, the syntax of *path logic* $\mathbf{PL}_\kappa(\mathbf{C}, At)$ is defined by the grammar

$$\varphi ::= p \mid \neg\varphi \mid \bigvee_{i\in I}\{\varphi_i\} \mid \langle f\rangle\varphi \mid \overline{\langle f\rangle}\varphi$$

where p ranges over At, $f \in \mathbf{C}_1$ and the cardinality of I is less than κ. We define $\bigwedge := \neg\bigvee\neg$, $\top = \bigwedge\emptyset$ and $\bot = \bigvee\emptyset$. The syntax of this logic is amenable for many interpretations, depending on the nature of the category $\mathbf{C}$; we will see in later sections how, when the base category is a poset, we can think of the modalities $\langle f\rangle\varphi$ and $\overline{\langle f\rangle}\varphi$ as extension and restriction of contexts.

In order to evaluate path logic on a presheaf, we first turn the presheaf into a labelled transition system:

3.2.2. Definition. Given a presheaf $P : \mathbf{C}^{op} \to \mathbf{Set}$, we can define a labelled transition system $\langle W, \{R_f\}_{f\in\mathbf{C}_1}\rangle$ via a variation of the category of elements, as described in [124]:

- $W := \{(x, C) | x \in P(C), C \in \mathbf{C}_0\}$
- $R_f := \{((x, C), (y, C')) | f : C \to C', P(f)(y) = x\}$

We previewed this construction in the previous Chapter, where we argued that the other encoding of contravariance allowed for further results when the arrows were functional bisimulations. At the object level the two constructions are obviously equivalent: one structure is obtained from the other by taking the converse of all the relations.

3.2.3. Definition. A *presheaf model M over* $\mathbf{C}$ is a presheaf P together with a valuation $V : At \to \mathcal{P}W$, where $\langle W, \{R_f\}_{f\in\mathbf{C}_1}\rangle$ is the LTS associated with P. The model M is said be rooted if P is a rooted presheaf.

We can now define the satisfaction relation for formulas of $\mathbf{PL}_\kappa(\mathbf{C}, At)$ on a presheaf model $M = (P, V)$ over $\mathbf{C}$, essentially by doing standard Kripke semantics over the LTS associated with the presheaf P and treating $\overline{\langle f\rangle}$ as a backwards modality. For atomic propositions we have $M, (x, C) \vDash p$ iff $(x, C) \in V(p)$ and the clauses for connectives are as usual, while for the modalities put

- $M, (x, C) \vDash \langle f\rangle\varphi$ iff there is (y, C') such that $((x, C), (y, C')) \in R_f$ and $M, (y, C') \vDash \varphi$
- $M, (x, C) \vDash \overline{\langle f\rangle}\varphi$ iff there is (y, C') such that $((y, C'), (x, C)) \in R_f$ and $M, (y, C') \vDash \varphi$

Since a presheaf model is just a TTSs with the relations flipped, forward modalities $\langle f \rangle$ in $\mathbf{LTTS^C}$ corresponds precisely to backward modalities $\overline{\langle f \rangle}$ in $\mathbf{PL}_\kappa(\mathbf{C}, At)$.

3.2.4. FACT. The logic $\mathbf{LTTS^C}$ is the fragment of the path logic $\mathbf{PL}_\kappa(\mathbf{C}, At)$ consisting of the finitary propositional calculus plus backward modalities.

The syntax and semantics of path logic were originally introduced in [74] to characterize the notion of strong path bisimulation:[3]

3.2.5. DEFINITION. [Path Bisimulation] A *path bisimulation* Z between two rooted presheaf models $M_1 = (Q_1, V_1)$ and $M_2 = (Q_2, V_2)$ over $\mathbf{C}$ is a family $(Z_C)_{C \in \mathbf{C}_0}$ in which each Z_C is a set of pairs of objects (x, y) such that $x \in Q_1(C)$ and $y \in Q_2(C)$ satisfying the following conditions:

1. roots are related: $(r_1, r_2) \in Z_I$;
2. if $(x, y) \in Z_C$ then $x \in V_1(p)$ iff $y \in V_2(p)$
3. (forward) for $(x, y) \in Z_C$, if there is $x' \in Q_1(C')$ such that $Q_1(f)(x') = x$ for $f : C \to C'$ then there must be $y' \in Q_2(C')$ such that $Q_2(f)(y') = y$ and $(x', y') \in Z_{C'}$, and conversely reversing the role of the presheaves;
4. (backward) if $(x, y) \in Z_C$ and $f : C' \to C$ then $(Q_1(f)(x), Q_2(f)(y)) \in Z_{C'}$.

In case $At = \emptyset$ we refer to Z simply as a path bisimulation between the presheaves Q_1 and Q_2.

Note that the two directions of this condition are sometimes called "zig" and "zag" or "forth" and "back" conditions in Modal logic; here they are clustered together in the (forward) item. Again, note that item 2 and 4 in this definition correspond to the conditions for bisimulations between TTSs, hence path bisimulation is a stronger notion. Path logic is expressive for path bisimulations:

3.2.6. THEOREM (SEE [74]). *There is a path bisimulation between two rooted presheaf models M_1, M_2 over $\mathbf{C}$ iff the respective roots satisfy the same formulas of the path logic $\mathbf{PL}_{|\mathbf{C}_0|}(\mathbf{C}, At)$.*

Since path bisimulation is stronger that bisimulation for TTSs and the latter correspond to natural transformations, it is reasonable to expect that path bisimulations will be matched by special natural transformations.[4]

[3] We will always consider strong path bisimulation, hence we will drop the adjective 'strong' henceforth.

[4] On presheaves this definition coincides with the one in [73].

3.2.7. Definition. [Open map] Given two presheaves $Q_1, Q_2 : \mathbf{C}^{op} \to \mathbf{Set}$, a natural transformation $\eta : Q_1 \to Q_2$ is an *open map* if, for every $f : C \to C'$ in $\mathbf{C}$, the following commuting square

$$\begin{array}{ccc} Q_1(C') & \xrightarrow[\eta_{C'}]{} & Q_2(C') \\ {\scriptstyle Q_1(f)}\downarrow & & \downarrow{\scriptstyle Q_2(f)} \\ Q_1(C) & \xrightarrow{\eta_C} & Q_2(C) \end{array}$$

is a quasi-pullback, that is, if $x \in Q_1(C)$ and $y \in Q_2(C')$ are such that $\eta_C(x) = Q_2(f)(y)$ then there exist $z \in Q_1(C')$ for which $\eta_{C'}(z) = y$ and $Q_1(f)(z) = x$.[5]

3.2.8. Definition. [Span, Co-Span] Given two objects C, D in a category $\mathbf{C}$, a *span* between them is a triple (C', f, g) such that C' is an object of $\mathbf{C}$ and $f : C' \to C$ and $g : C' \to D$ are two morphisms in $\mathbf{C}$. A *co-span* between them is a triple (C', f, g) such that C' is an object of $\mathbf{C}$ and $f : C \to C'$ and $g : D \to C'$ are two morphisms in $\mathbf{C}$.

3.2.9. Theorem (See [74]). *A pair of rooted presheaves are path bisimilar if, and only if, they are related by a span of open maps.*

3.3 Correspondence theory for path logic over presheaves

We have seen that path logic is known to characterize the notion of path bisimulation. A classic result in modal logic, van Benthem's theorem, states that basic modal logic is the bisimulation invariant fragment of first-order logic. We shall see that an analogous result holds for finitary path logic: this logic can be characterized as the fragment of a many-sorted first-order language that is invariant for path bisimulation, provided that the background category $\mathbf{C}$ has finitely many objects. The finitary path logic $\mathbf{PL}_\omega(\mathbf{C}, At)$ is presented concretely by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \langle f \rangle \varphi \mid \overline{\langle f \rangle} \varphi$$

The first step is to show how the tools of Correspondence Theory can be adapted to this setting. The main difference to account for is that presheaf models are *sorted* structures: each element comes with an associated object from the category $\mathbf{C}$ that we can regard as a sort. Hence the yardstick against which we measure the expressivity of $\mathbf{PL}_\omega(\mathbf{C}, At)$ is a many-sorted first order logic whose sorts are given by $\mathbf{C}_0$ and whose relational symbols are given by $\mathbf{C}_1$. We assume a countably

[5]This definition is equivalent to the one in term of path lifting, see [40].

infinite supply of variables for each object C. We use a subscript to indicate the sort, e.g. x_C is a variable of sort C.

The syntax of $\mathbf{FOL}(\mathbf{C}, At)$ is defined by the following grammar:

$$\varphi ::= P^p x_C \mid x_C R_f x_{C'} \mid x_C =_C y_C \mid \neg\varphi \mid \varphi \vee \varphi \mid \exists x_C(\varphi)$$

for $f : C \to C'$ in $\mathbf{C}_1$, $C, C' \in \mathbf{C}_0$ and $p \in At$. The other connectives and $\forall$ are defined as usual; we use $\exists! x_C$ as shorthand for $\exists x_C \forall y_C(y_C = x_C)$. The intended models for this first order language are many-sorted relational structures.

3.3.1. DEFINITION. A $\mathbf{FOL}(\mathbf{C}, At)$-*model* is a tuple $M = \langle \{D_C\}_{C\in\mathbf{C}_0}, \{R_f\}_{f\in\mathbf{C}_1}, \{P^p\}_{p\in At}\rangle$ where D_C is a set for every C, if $f : C \to C'$ then R_f is a relation $R_f \subseteq D_C \times D_{C'}$ and P^p is a unary predicate over the disjoint union of the family $\{D_C\}_{C\in\mathbf{C}_0}$.

Clearly presheaf models are particular instances of such structures, with the purely cosmetic difference that all the objects are lumped together into a disjoint union. The satisfaction relation $\vDash_\sigma$ relative to a variable assignment σ is defined in the customary way. When a model M, an object u in M and a variable assignment σ are clear from the context we write $\varphi(x_C)[u]$ to mean that the variable assignment is modified in order to map the variable x_C to u, leaving the rest of the assignment unchanged. Elementary equivalence and elementary extension of a model are defined in the usual way.

The usual notion of standard translation can also be made parametric in a sorted variable: define by recursion ST_{x_C}

- $ST_{x_C}(p) = P^p x_C$
- $ST_{x_C}(\neg\varphi) = \neg ST_{x_C}(\varphi)$
- $ST_{x_C}(\varphi \vee \psi) = ST_{x_C}(\varphi) \vee ST_{x_C}(\psi)$
- $ST_{x_C}(\langle f\rangle\varphi) = \exists_{y_{C'}}(x_C R_f y_{C'} \wedge ST_{y_{C'}}(\varphi))$, if $f : C \to C'$
- $ST_{x_C}(\langle f\rangle\varphi) = \perp$, otherwise.
- $ST_{x_C}(\overline{\langle f\rangle}\varphi) = \exists_{y_{C'}}(y_{C'} R_f x_C \wedge ST_{y_{C'}}(\varphi))$, if $f : C' \to C$
- $ST_{x_C}(\overline{\langle f\rangle}\varphi) = \perp$, otherwise.

Call $\mathcal{M}(\mathbf{C})$ the class of models of $\mathbf{FOL}(\mathbf{C}, At)$ arising from presheaves over $\mathbf{C}$. The main observation we need for our characterization result is the following:

3.3.2. LEMMA. *The class of* $\mathbf{FOL}(\mathbf{C}, At)$*-models arising from presheaves is elementary. The axiom schemas are:*

1. $\forall x_C \forall y_C\,(x_C R_{\mathsf{Id}_C} y_C \leftrightarrow x_C =_C y_C)$, *for* $C \in \mathbf{C}_0$

2. $\forall x_C \forall y_{C''} \, (x_C R_{f \circ g} y_{C''} \leftrightarrow \exists z_{C'} (x_C R_f z_{C'} \wedge z_{C'} R_g y_{C''}))$, *for* $h : C' \to C''$ *and* $f : C \to C'$
3. $\forall x_C \exists! y_{C'} \, y_{C'} R_f x_C$, *for* $f : C' \to C$

Proof:
Given a model M of the axioms, define a presheaf $F : \mathbf{C}^{op} \to \mathbf{Set}$ as follows:

- $F(C) = D_C$
- $F(f : C \to C') = \{(y, x) | (x, y) \in R_f\}$

The axioms ensure that Id_C is an identity for every C, that every $F(f)$ is a total function and that composition behaves as it should. For the valuation, interpret each $p \in At$ to the extension of the corresponding predicate P^p. □

Note that it is also possible to impose rootedness via the formula

$$\exists! y_0 \, (y_0 =_0 y_0)$$

With this observation in place the argument essentially adapts van Benthem's original proof to the present setting. We make essential use of the fact that some general results of first-order logic also hold for the multi-sorted version, e.g. the Compactness Theorem (see e.g. [53]). First we refresh two known definitions.

3.3.3. Definition. [Modal saturation] Given a $\mathbf{FOL}(\mathbf{C}, At)$-model, we say that M is *modally saturated* if, for every element $x \in D_C$ and each label f with $dom(f) = C$, we have that every theory in $\mathbf{PL}_\omega(\mathbf{C}, At)$ which is finitely satisfiable among the R_f-successors of x is satisfiable at some R_f-successor of x.

3.3.4. Definition. [ω-saturation] For X be a finite subset of a model M, let $\mathbf{FOL}(\mathbf{C}, At)[X]$ be the language expanded with constants $\underline{a}$ for elements $a \in X$. Call M_X be the expansion of M to a model of $\mathbf{FOL}(\mathbf{C}, At)[X]$ where each constant $\underline{a}$ in interpreted to the corresponding a. A model M is *ω-saturated* if every set of formulas Γ such that

- formulas in Γ are in the language $\mathbf{FOL}(\mathbf{C}, At)[X]$,
- formulas in Γ have one open variable,
- Γ is consistent with the many-sorted first order theory of M_X,

there is an element b such that $\psi(x_C)[b]$ holds in M_X for all $\psi(x_C) \in \Gamma$. We say in this case that Γ is realized in M.

Also in multi-sorted logic it is easy to show that any ω-saturated $\mathbf{FOL}(\mathbf{C}, At)$-model is modally saturated.

3.3.5. Lemma. *Any ω-saturated* $\mathbf{FOL}(\mathbf{C}, At)$*-model is modally saturated.*

Proof:
Suppose a model M is ω-saturated, let w be an element of the domain and let Γ be a theory in $\mathbf{PL}_\omega(\mathbf{C}, At)$ that is finitely satisfiable among the R_f-successors of w. Let $f : C \to C'$. Define $\Gamma' = \{wR_f x_{C'}\} \cup \{ST_{x_{C'}}(\phi) | \phi \in \Gamma\}$.

We claim Γ' is consistent with the (many-sorted) first order theory of M, $\mathcal{T}(M)$. If not, by compactness there must be finitely many formulas $\psi_1, \dots, \psi_n$ that entail the negation of a formula $\xi \in \mathcal{T}(M)$. But we assumed Γ is finitely satisfiable among the R_f-successors of w, thus there is a successor satisfying the counterparts of $\psi_1, \dots, \psi_n$. This would entail that said successor satisfies $\neg\xi$, contradiction. By ω-saturation we have that Γ' must be realized at some world v. Hence v is a R_f-successor of w and satisfies Γ. □

The following step is where we need the category to have finitely many objects. We exploit what is known as the 'fundamental translation' in order to harness results on single-sorted models; this and other standard results on many-sorted logic can be found in [53] and [119].

3.3.6. Lemma. *Any* $\mathbf{FOL}(\mathbf{C}, At)$*-model has a modally saturated elementary extension.*

Proof:
Given a $\mathbf{FOL}(\mathbf{C}, At)$-model M, we can transform it into a single-sorted model via a canonical construction: we take the disjoint union of all the sets in the family $\{D_C\}_{C \in \mathbf{C}_0}$ and add one predicate $P_C = D_C$ for each sort C, in order to retain the information about the sorts. Call the new single-sorted model M^*. We then translate $\mathbf{FOL}(\mathbf{C}, At)$ into single-sorted first-order logic as follows:

- $(x_C)^* = x$
- $(x_C =_C y_C)^* = (x^*_C = y^*_C)$
- $(x_C R_f y_{C'})^* = x^*_C R_f y^*_{C'}$
- $(\neg\phi)^* = \neg\phi^*$
- $(\psi \wedge \phi)^* = \psi^* \wedge \phi^*$
- $(\exists x_C \phi)^* = \exists x(P^C(x) \wedge \phi^*)$

For any assignment σ of sorted variables to elements of M, define σ^* as $\sigma^*(x) = \sigma(x_C)$. A standard proof by induction shows that, for every many-sorted formula ϕ, $M \vDash_\sigma \phi$ iff $M^* \vDash_{\sigma^*} \phi^*$. Moreover, the following will be true in M^*:

$$\forall x(\bigvee_{C \in \mathbf{C}_0} P^C(x)) \wedge \bigwedge_{C \neq C'} \neg\exists y(P^C(y) \wedge P^{C'}(y)) \tag{3.1}$$

This formula states that every element in the model has one and only one sort. This can be encoded in a formula since the objects of the category, qua sorts, are finitely many.

By standard results in Model Theory, every first-order structure has an ω-saturated elementary extension (see for example [69] Theorem 8.2.1): call $UP(M^*)$ the ω-saturated elementary extension of M^*. As a consequence of Łos Theorem, the formula 3.1 is also true in $UP(M^*)$. We can thus see $UP(M^*)$ as a $\mathbf{FOL}(\mathbf{C}, At)$-model. Given an ω-type Σ in $\mathbf{FOL}(\mathbf{C}, At)$, the type $\{(\phi)^*|\phi \in \Sigma\}$ will be realized in $UP(M^*)$ at some tuple $\overline{w}$ because of ω-saturation. By $M \vDash_\sigma \phi$ iff $M^* \vDash_{\sigma^*} \phi^*$ we know that the same $\overline{w}$ will satisfy the original Σ. □

This Lemma holds when restricted to any elementary class in $\mathbf{FOL}(\mathbf{C}, At)$, since the fundamental translation preserves the truth of the formulas in the theory and so do elementary extensions.

3.3.7. Corollary. *Let M be a $\mathbf{FOL}(\mathbf{C}, At)$-model arising from a presheaf. Then M has a modally saturated elementary extension that also arises from a presheaf.*

We can use the standard argument to prove the following:

3.3.8. Lemma. *Let M and M' be two $\mathbf{FOL}(\mathbf{C}, At)$-models arising from presheaves P and P' respectively, and let $u \in P(C)$, $u' \in P'(C)$. If M, u and M', u' satisfy the same formulas of finitary path logic, and M and M' are both modally saturated, then there is a path bisimulation between P and P' relating u to u'.*

A formula $\varphi(x_C)$ of $\mathbf{FOL}(\mathbf{C}, At)$ is *invariant for path bisimulations* if, whenever Z is a path bisimulation between presheaves P and P', M and M' are the structures induced by P and P' respectively and $u \in P(C)$ and $u' \in P'(C)$ are such that $(u, u') \in Z_C$, we have

$$M \vDash \varphi(x_C)[u] \text{ iff } M' \vDash \varphi(x_C)[u']$$

3.3.9. Theorem. *Let $\varphi(x_C)$ be any $\mathbf{FOL}(\mathbf{C}, At)$ formula open in one variable x_C which is invariant for path bisimulations. Then there exists a modal formula $\varphi^\dagger$ in $\mathbf{PL}_\omega(\mathbf{C}, At)$ such that, for every $M \in \mathcal{M}(\mathbf{C})$ and every u of sort C, we have*

$$(M, u) \vDash \varphi^\dagger \textit{ iff } M \vDash \varphi(x_C)[u]$$

Proof:
Suppose there is no such formula $\varphi^\dagger$. Define the set of path logic formulas that are consequence of $\varphi(x_C)$:

$$MOC(\varphi(x_C)) = \{ST_{x_C}(\psi)|\psi \text{ in } \mathbf{PL}_\omega(\mathbf{C}, At), \varphi(x_C) \vDash ST_{x_C}(\psi)\}$$

Since the Compactness Theorem holds for many-sorted first order logics, the usual argument will serve to show that $MOC(\varphi(x_C)) \vDash \varphi(x_C)$ entails that $\varphi(x_C)$ is equivalent to the translation of a finitary path logic formula.[6]

Assume a model $M \in \mathcal{M}(\mathbf{C})$ is such that $M \vDash MOC(\varphi(x_C))[u]$, we show $M \vDash \varphi(x_C)[u]$. Let

$$T(x_C) = \{ST_{x_C}(\psi)|M \vDash ST_{x_C}(\psi)[u]\}$$

Again by a standard compactness argument we can find a model N and an element w' such that $N \vDash T(x_C) \cup \{\varphi(x_C)[u']\}$. By construction w and w' satisfy the same formulas of path logic. Now consider the ω-saturated, and hence modally saturated, elementary extension of both models, which exist by Lemma 3.3.7. Call them $UP(M)$ and $UP(N)$.

Since $\mathcal{M}(\mathbf{C})$ is an elementary class any modally saturated elementary extensions of these models will provide models in $\mathcal{M}(\mathbf{C})$. Moreover, (the images under the embeddings of) u and u' still satisfy the same path logic formulas. Since we are in modally saturated models, Lemma 3.3.8 entails that there is a path bisimulation connecting the two elements. Hence $N \vDash \varphi(x_C)[u']$ entails $UP(N) \vDash \varphi(x_C)[e(u')]\}$, where $e(u')$ is the image of the element u' under the elementary embedding $N \hookrightarrow UP(N)$. Invariance under bisimulation gives $UP(M) \vDash \varphi(x_C)[e(u)]\}$, while invariance under elementary embeddings allows us to conclude $M \vDash \varphi(x_C)[u]\}$. □

The converse of the Theorem is proved as usual by induction on the structure of the modal formula.

3.4 Expressing properties of presheaves in path logic

We now turn to interesting properties of presheaf that we can capture in the language. The first notions we investigate are that of 'generated subpresheaf' and 'co-generated subpresheaf'. Suppose given a presheaf $P : \mathbf{C}^{op} \to \mathbf{Set}$ and a subset A of the associated transition system: A can be seen as a family $\{A_C\}_{C \in \mathbf{C}_0}$ such that $A_C \subseteq P(C)$ for all $C \in \mathbf{C}_0$. When such an A is fixed, we can generate the *maximal subpresheaf contained in* A and the *minimal subpresheaf containing* A, sometimes called the 'presheaf co-generated by A' and the 'presheaf generated by A', respectively. The former is defined as

$$\overline{A}(C) = \{x|\forall C' \in \mathbf{C}_0 \, \forall f : C' \to C \quad P(f)(x) \in A_C\}$$

that is, one removes from A all the elements that are mapped outside of A by some function. The latter is defined as

$$\underline{A}(C) = \{y|\exists C' \in \mathbf{C}_0 \, \exists f : C \to C' \quad y = P(f)(x)\}$$

[6]See [32] for the standard argument.

thus we add to A all the images under the functions. These two notions were employed by Ghilardi and Meloni in [60] as semantics for their first order modal logic. It turns out that both structures can be captured in path logic. Given a presheaf P and its relational structure M, let p be a propositional variable interpreted on M. The interpretation of a propositional variable can be sliced into a family, $\{A_C\}_{C \in \mathbf{C}_0}$ such that $A_C \subseteq P(C)$, since M is composed by the disjoint union of the sets $P(C)$'s. Then the maximal subpresheaf contained in p is definable as

$$p \wedge \bigwedge_{f \in \mathbf{C}_1} \overline{[f]} p$$

When p is replaced with another formula ϕ, this in particular means that the logic can talk about the biggest submodel where ϕ is valid. The notion of minimal subpresheaf containing p is encoded in the formula

$$p \vee \bigvee_{f \in \mathbf{C}_1} \langle f \rangle p$$

3.4.1 Path logic and Topology

We now introduce sheaves and presheaves over a topological space and offer some examples of topological properties that can be captured by path logic.

As the name suggests, the concept of presheaf can be strengthened to obtain what is known as sheaf, a formal tool that was introduced in algebraic geometry to handle information attached to open sets of a topological space.[7] Even though they can be defined in general categorical terms, we shall only be interested in sheaves over a fixed topological space $\mathbb{X}$, that is, sheaves with base category the poset category of open sets $Open(\mathbb{X})$. In fact, we shall simply identify a topological space $\mathbb{X}$ with the associated poset category of open sets. Hence the path logic associated with a space $\mathbb{X}$ and a cardinal κ is just the logic $\mathbf{PL}_\kappa(\mathbb{X})$, where this notation is used just as before. Moreover, even though sheaves can have different categories as target, depending on which kind of information is attached to open sets, we shall restrict ourselves to presheaves over **Set**.

The key feature of a sheaf is the way in which it connects local and global data, that is, how it reconciles the information attached to an open (global data) with the information attached to a family of open sets 'covering it' (local data). We begin by making the notion of covering precise.

3.4.1. Definition. Given an open set $U \in Open(\mathbb{X})$, a *covering family* for U is a family of opens $(U_i)_{i \in I}$ such that $U_i \subseteq U$ for every i and $U \subseteq \bigcup_{i \in I} U_i$.

Given a presheaf $P : \mathbb{X}^{op} \to \mathbf{Set}$, an inclusion $\iota : U \subseteq U'$ and an element $x \in P(U')$ we sometimes denote $P(\iota)(x) \in P(U)$ with $x|^P_U$; this is called the

[7] For a classic in sheaf theory see [85].

restriction of x to U. Elements of $P(U)$ for an open set U will be referred to as the *sections* of P over U. Elements of $P(X)$, where X refers to the whole space, are called *global sections*.

3.4.2. DEFINITION. [Sheaf] A *sheaf* P over $\mathbb{X}$ is a presheaf over $Open(\mathbb{X})$ satisfying the following conditions, for any given covering family $(U_i)_{i\in I}$ of an open U:

1. If $x, y \in P(U)$ are such that $x|^P_{U_i} = y|^P_{U_i}$ for all $i \in I$ then $x = y$, that is, if two elements agree on their restrictions to the members of the covering family then they must coincide. This condition is often called **locality**.

2. If a given family $(x_i)_{i\in I}$ is such that $x_i \in P(U_i)$ and $x_i|^P_{U_i\cap U_j} = x_j|^P_{U_i\cap U_j}$ for all $i, j \in I$ (the elements of the family 'agree on the intersections' of the covering family) then there exists a 'gluing' of such family, an element $x \in P(U)$ such that $x|^P_{U_i} = x_i$. This condition is known as **gluing**.

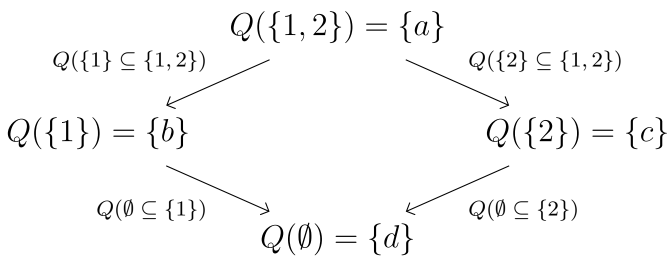
We denote by $\mathbf{Sh}(\mathbf{C})$ the category of sheaves over $\mathbf{C}$.

Note that every sheaf is a rooted presheaf: the initial object in this case is the empty set, which is an open; the image of the empty set must be a singleton due to the locality condition on a sheaf.

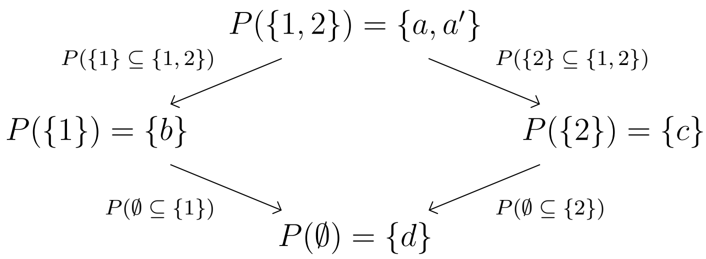
Alas, locality and gluing are not definable in path logic.

3.4.3. PROPOSITION. *The 'locality' condition is not definable in path logic.*

Proof:
The proof is a standard undefinability argument. Let us suppose by contradiction that locality is definable by a formula α in path logic, then by previous results the formula α is invariant for path bisimulation. If we can find a sheaf and a presheaf (that does not satisfy locality) and a path bisimulation between them then we obtain a contradiction. Consider the set $\{1, 2\}$ and the discrete topology $\wp(\{1, 2\})$. Define the sheaf Q as depicted in the figure.



The functions are defined in the obvious way and it is straightforward to check that this is a sheaf. The element d constitutes the root. Now define a presheaf P which is exactly as Q but for a little difference: now there is another element a' in the image of $\{1, 2\}$ and a, a' have the same restrictions.



Since the singletons constitute a covering family of $\{1, 2\}$ and a, a' have the same restrictions, this presheaf violates locality, which would require $a = a'$. Now define a relation between the two corresponding models connecting all the elements with themselves, that is, $Z = \{(a, a), (b, b), (c, c), (d, d)\}$. It is simple to check that this is a path bisimulations, thus we found the contradiction and we must conclude that locality is undefinable in path logic. □

Note that, since one of the two models is a sheaf, the last proof also shows that the conjunction of locality and gluing is not definable.

3.4.4. Proposition. *The 'gluing' condition is not definable in path logic.*

Proof:

The argument is analogous to the one used in the previous proposition. Consider the same sheaf Q and define P to be exactly as Q but with two copies of all the elements except the root. We thus have $P(\{1\}) = \{b_1, b_2\}$, $P(\{2\}) = \{c_1, c_2\}$ and $P(\{1, 2\}) = \{a_1, a_2\}$. All the b's and the c's are mapped to the root d, while a_i is mapped to b_i and c_i respectively (where $i \in \{1, 2\}$). We thus we have a family b_1, c_2 such that

- each element belongs to the image of a singleton (and the singletons form a covering family),
- they 'agree on intersections', that is, their restrictions coincide.

However, there is no element in $P(\{1, 2\})$ that is mapped to both b_1 and c_2, thus the gluing property fails for P. Nevertheless, $Z = \{(x, x_i) | x \in \{a, b, c\}, i \in \{1, 2\}\} \cup \{(d, d)\}$ is still a path bisimulation between the two models. □

These proofs reveal what the core issue is: the logic cannot name the elements in the image of the presheaf. We will see in Section 3.6 how to enhance path logic to capture gluing and locality. For now we restrict the notion of presheaf model to sheaves.

3.4.5. Definition. A *sheaf model* is a presheaf model (P, V) where P is a sheaf.

Since the morphisms in the base category are inclusion maps, and there is only one such inclusion for each pair of objects, in the corresponding path logic we denote with $\langle U, U' \rangle$ the modality associated to the inclusion $U \subseteq U'$ for U and U' opens in $\mathbb{X}$. A fairly natural way of interpreting the modalities in this setting, and more generally when the base category is a poset category, is in terms of *change of context*: the forward modality expresses the fact that a property holds when the context is extended from U to U', while the backward one handles the restriction from bigger to smaller contexts. This perspective on path logic is apt for the sheaf-theoretic analysis of contextuality, as we will see in the next subsection.

For a formula φ of path logic and a rooted presheaf P, we may write $P \vDash \varphi$ to say that φ is true at the root. An example of a sheaf of immediate topological interest is the sheaf of sections of a covering map:

3.4.6. Definition. [Covering Map, Sheaf of Sections] Let $\mathbb{X}$ be a topological space and let $\pi : \mathbb{Y} \to \mathbb{X}$ be any continuous map. Then π is called a *covering map* if, for every point u in $\mathbb{X}$, there is an open neighborhood U of u such that the inverse image $\pi^{-1}[U]$ is the union of disjoint sets $\{V_i\}_{i \in I}$ such that, for each $i \in I$, the restriction of π to V_i is a homeomorphism onto U.

Any covering map $\pi : \mathbb{Y} \to \mathbb{X}$ gives rise to a sheaf P, called the *sheaf of sections of* π. Given an open set U, the elements of $P(U)$ are continuous maps f mapping U into Y such that $\pi \circ f = Id_U$. Restrictions of sections are given by function restriction.

We can use path logic to describe how sections extend and restrict when we extend or restrict the corresponding open sets. In the case of a sheaf of sections P, calling X the carrier of the topological space $\mathbb{X}$, the statement

$$P \vDash [\emptyset, X] \perp \tag{3.2}$$

describes how the space $\mathbb{Y}$ is related to $\mathbb{X}$ by the covering map π: it states that the unique section over the empty set (the root) cannot be extended to a section over the whole set X. Since every section restricts to the root, this is tantamount to saying that there is no global section. Thus the space $\mathbb{X}$ may look like $\mathbb{Y}$ locally, that is, there could be sections from some opens of $\mathbb{X}$ into $\mathbb{Y}$, but not globally.

An example is the covering map $\pi : \mathbb{R} \to \mathbf{S}_1$, where $\mathbb{R}$ is the real line and $\mathbf{S}_1$ is the unit circle, such that $\pi(a) = (cos 2\pi a, sin 2\pi a)$. We take $\mathbb{R}$ to be equipped with the open interval topology and $\mathbf{S}_1$ with the subspace topology from $\mathbb{R}^2$. This covering map has no global section: for $\pi \circ f = Id_X$ to be the case f would have to be injective, but there is no injective continuous map from the unit circle into $\mathbb{R}$. However, there are infinitely many sections over every smaller open set in $\mathbf{S}_1$.

On the other hand, the statement

$$P \vDash \bigvee_{U \in Open(\mathbb{X})} \langle \emptyset, U \rangle [U, X] \perp \tag{3.3}$$

only forces the existence of *some* sections that cannot be extended to a global one; an example of a covering map that satisfies 3.3 but not 3.2 is $\pi : \mathbb{R} + \mathbb{R} \to \mathbb{R}$, the projections of two disjoint copies of $\mathbb{R}$ into $\mathbb{R}$ itself. Despite there being global sections, we can design a section over two disjoint open intervals in such a way that it cannot be extended to a section over the whole space. Take for example the two intervals $(1, 2)$ and $(3, 4)$ and call U the open interval resulting by their union. Define a section $f : U \to \mathbb{R} + \mathbb{R}$ that maps the first interval to the first copy of $\mathbb{R}$ and the second interval to the second copy of $\mathbb{R}$. Clearly $\pi \circ f = Id_U$, but f cannot be extended to a global section, i.e. a section over the entire $\mathbb{R}$.

Another example of a general notion that can be encoded in path logic is *flabby sheaf*: a sheaf $P : \mathbb{X}^{op} \to \mathbf{Set}$ is said to be flabby if, for every inclusion $\iota : U \to X$, the restriction map $P(\iota)$ is surjective. Flabby sheaves play a special role in homological algebra, see [38] for an overview. It is not hard to see that the class of flabby sheaves over $\mathbb{X}$ is captured by the path logic formula

$$\bigwedge_{U \in Open(\mathbb{X})} [\emptyset, U]\langle U, X\rangle \top$$

By previous results on path logic we can conclude that said properties are invariant under path bisimulation over presheaf models.

3.4.2 Path logic and Contextuality

In this section we outline the framework put forward in [4] and describe how to encode contextuality in its different variants. We adopt notation and definitions from [3]. As a proof of concept we describe how to apply path logic to one of the simplest frameworks for contextuality; we believe these intuition are transferrable to more complex models, for example ones incorporating simplicial complexes.

Suppose given a finite set X of variables, that in the quantum setting can be regarded as physical quantities, together with a set of possible outcomes O. Define a sheaf $\mathcal{E} : \wp(X)^{op} \to \mathbf{Set}$ mapping $U \subseteq X$ to O^U, the set of functions from U to O, while on arrows the function $\mathcal{E}(U \subseteq U')$ simply maps a function to the same function on the restricted domain. This functor is called the *sheaf of events*, as it associates to each set of variables all the possible assignments of outcomes to those variables.

Consider now a family of subsets of X, call it $\mathcal{M}$, such that the members of $\mathcal{M}$ form an antichain in $\wp(X)$ and $\bigcup \mathcal{M} = X$. Such family $\mathcal{M}$ is called *measurement cover* and represents the maximal sets of variables that can be tested together. For example, the variables associated to position and momentum in a quantum system cannot be tested together: this and analogous constraints motivate this definition. Note that our inability to test two variables together does not preclude a priori the existence of a simultaneous assignment of values to both. Given a triple $\langle X, \mathcal{M}, O\rangle$, a subpresheaf $\mathcal{S}$ of $\mathcal{E}$ is called an *empirical model* if

1. $\mathcal{S}(C) \neq \emptyset$ for all $C \in \mathcal{M}$: possible joint measurements give joint outcomes.

2. $\mathcal{S}(U \subseteq U')$ is surjective if $U \subseteq U' \subseteq C$ for $C \in \mathcal{M}$: the model satisfies the no-signalling principle.

3. For any family $\{s_C\}_{C\in\mathcal{M}}$ with $s_C \in \mathcal{S}(C)$ such that

$$\forall C, C' \in \mathcal{M} \quad s_C|_{C\cap C'} = s_{C'}|_{C\cap C'}$$

   there exists a unique global section in $\mathcal{S}(X)$. This is the same as the gluing condition for sheaves, relativized to $\mathcal{M}$.

It is worth remarking that the notion of empirical model cannot itself be encoded in basic path logic. The first condition can be captured by $\bigwedge_{C\in\mathcal{M}}\langle\emptyset, C\rangle\top$, a formula stating that there is a section assigning outcomes to all the measurements in C, for all $C \in \mathcal{M}$. The second requirement is recorded by the formula $\bigwedge_{C\in\mathcal{M}}\bigwedge_{U\subseteq U'\subseteq C}[\emptyset, U]\langle U, U'\rangle\top$, expressing the fact that every section over U has a section over U' extending it. Note that we do not need infinitary connectives to form these conjunctions, due to the finiteness of X. The third condition however cannot be rendered in path logic, since it is essentially a relativized version of gluing; this issue will be addressed in Section 3.6, where we will suggest how to express such properties with the use of nominals.

With respect to contextuality, the key properties of an empirical model are called *weak contextuality* and *strong contextuality*.[8] An empirical model is weakly contextual if there is a maximal context $C \in \mathcal{M}$ and a section $s \in \mathcal{S}(C)$ such that s cannot be extended to a global section in $\mathcal{S}(X)$. This means that there is a particular assignment of values to the measurements that cannot be reconciled with an assignment of values to all variables together.

Given an empirical model $\mathcal{S}$ we can capture weak contextuality in the path logic for $\wp(X)$ in a natural way:

$$\bigvee_{C\in\mathcal{M}}\langle\emptyset, C\rangle[C, X]\perp$$

An empirical model is said to be strongly contextual if there is no global section: $\mathcal{S}(X) = \emptyset$. This condition states that there cannot be a simultaneous assignment of values to all variables; it can be encoded with the formula:

$$[\emptyset, X]\perp$$

Notice that these formulas are akin to those discussed in the previous section in the context of covering spaces.

A similar treatment of these notions, also casted in a modal language, was offered by Kishida in [75]. In said paper the labels for the modalities are measurements contexts, that is, compatible sets of measurements, and propositional

[8] Weak contextuality is called *logical* contextuality in [3].

variables are used to specify which outcome is associated to which measurement. The notions of weak and strong contextuality are captured via a formula *Det* expressing determinacy, namely a big disjunction encoding all the pairs measurements-outcomes and stating that one of them is the case.

We believe path logic constitutes an improvement over this line of work for three reasons.[9] First, the modalities of path logic contain all the identities of the objects of $\wp(X)$, thus we can encode a formula $[a]\phi$ from [75], meaning that ϕ will be the case whenever measurement a is performed, into the formula $[\emptyset, \{a\}]\phi$, stating that every section over a, hence any assignment of outcome to a brought about by measuring a, will satisfy ϕ. We can then use the propositional variables to associate measurements and outcomes to reproduce Kishida's formulas within path logic. Second, path logic can express the *change* of measurement context: this allows for a characterization of contextuality in terms of the impossibility to extend to global sections, along the lines of the original paper [4]. Such characterization abstracts away from the particular specification of all the measurement-output pairs. Finally, path logic is not a logic designed specifically for this setting, but rather a very general language already studied in relation to concurrency. As known tools do, it not only solves the task at hand but also suggests further connections. In particular, since we know that path logic formulas are invariant under path bisimulations, expressing some of the central properties of contextuality in path logic leads to a simple but interesting observation:

3.4.7. COROLLARY. *The properties of weak and strong contextuality are preserved by path bisimulations (and hence by spans of open maps).*

This connection between contextuality and path bisimulation seems to merit further investigation.

3.4.3 Two ways forward

We have seen that we can capture some interesting features of presheaves and this seems to encourage further investigation into the expressivity of path logic over presheaf models. On the other hand, some crucial properties such as locality and gluing, the defining properties of a sheaf, fall outside of the scope of path logic. In the next sections we explore two possible ways to reason about sheaves.

One possibility is to enhance the semantic companion of path logic, namely path bisimulation, to preserve locality and gluing. The next section is devoted to this line of enquiry; we shall see that there are indeed other possible candidates for notions of behavioural equivalence of sheaves. Another approach, developed in Section 3.6, is to add enough expressive power to define the two properties in the language.

[9]These consideration address only the fragment without probabilities; we believe similar remarks can be made for the probabilistic case.

3.5 Bisimulations for sheaves

In the general case path bisimulations could be neatly characterized in terms of spans of open maps in the category of presheaves over a fixed base category: path bisimulations correspond to spans of open maps, in the sense that rooted presheaves are path bisimilar if and only if they are related by a span of open maps. Since open maps are special cases of coalgebra morphisms [80], spans of open maps correspond to what is known as an *Aczel-Mendler bisimulation*. Furthermore, it is not hard to show that two presheaves are related by a span of open maps iff they are related by a co-span of open maps.

3.5.1. PROPOSITION. *Existence of a co-span of open maps entails the existence of a span of open maps.*

This follows since open maps are stable under pullbacks (see [73]), so we can always obtain a span of open maps from a co-span of open maps by taking the pullback (which exists in $\mathbf{Sh}(\mathbb{X})$ since this is a topos).

3.5.2. PROPOSITION. *The pushout of a span of open maps in* $\mathbf{Set}^{\mathbb{X}^{op}}$ *is a co-span of open maps.*

Proof:
This follows from the usual description of pushouts as 'co-products followed by co-equalizers': the insertion maps into the co-product are clearly open, and the co-equalizing map is open since it is essentially a quotient map from the co-product to its quotient by the bisimulation induced by the open span. So the proof that this map is open follows the usual proof that the quotient map from a Kripke model to its bisimulation quotient is a p-morphism. □

As a corollary we have that a pair of rooted presheaves is connected by a span of open maps if, and only if, it is connected by a co-span of open maps.

Seeing open maps as coalgebra morphisms, a co-span is in fact an instance of the coalgebraic concept of behavioural equivalence as a co-span in a category of coalgebras. So we have three equivalent descriptions of bisimilarity of presheaf models: path bisimilarity, spans of open maps and co-spans of open maps.

The situation for sheaves is less straightforward: the proofs of the equivalences mentioned above are not valid when we restrict attention to the category of sheaves over a space, i.e. when "span of open maps" means a span *in the category of sheaves* over the given space. So it seems we have three genuinely distinct candidates for behavioural equivalence of sheaves. It is easy to see that spans of open maps give rise to path bisimulations: the proof that worked for presheaves (Theorem 3.2.9) also covers the case of sheaves. Furthermore, co-spans of open maps give rise to spans of open maps for the reason mentioned before: the category of sheaves over a fixed space has pullbacks and open maps are stable under pullbacks. So we have:

$$\text{co-spans} \;\Rightarrow\; \text{spans} \;\Rightarrow\; \text{path bisimulations}$$

In the following sections, we shall look more closely at spans and co-spans of open maps, and relate them to special kinds of path bisimulations.

3.5.1 Path bisimulations and spans of open maps

We start by investigating the connection between path bisimulations and spans of open maps. It is certainly true that, for any pair of path bisimilar sheaves, we can construct a span of open maps connecting these sheaves. However, the presheaf at the "vertex" of this span may not be a sheaf, so the characterization of path bisimulations as spans of open maps is not internal to the category of sheaves over a given space.

One way to think about the problem is to see the path bisimulation itself as a presheaf, where the image of an object C is Z_C and the image of a morphism is given by restriction on both components. From this point of view natural requirement for a path bisimulation is to satisfy the gluing condition. Let Z be a path bisimulation between Q_1 and Q_2, let $(U_i)_{i\in I}$ be a covering family for U: Z satisfies gluing if for every given family of pairs $\{(x_i, y_i) \in Z_{U_i}\}$ such that they agree on intersections, that is $(x_i|^{Q_1}_{U_i\cap U_j}, y_i|^{Q_2}_{U_i}) = (x_j|^{Q_1}_{U_i\cap U_j}, y_j|^{Q_2}_{U_i})$, there is a pair $(x, y) \in Z_U$ such that $(x, y)|^{Z}_{U_i} = (x|^{Q_1}_{U_i}, y|^{Q_2}_{U_i}) = (x_i, y_i)$.

3.5.3. Proposition. *Two sheaves on $\mathbb{X}$ are related by a path bisimulation satisfying the gluing condition iff they are related by a span of open maps.*

Proof:
From left to right: given $Q_1, Q_2 : \mathit{Open}(\mathbb{X})^{op} \to \mathbf{Set}$ and a path bisimulation Z between them, define a sheaf $P : \mathit{Open}(\mathbb{X})^{op} \to \mathbf{Set}$ by mapping

$$U \mapsto Z_U$$
$$\iota : U \to U' \mapsto P(\iota) : Z_{U'} \to Z_U$$

where $P(\iota)((x, y)) = (x|^{Q_1}_U, y|^{Q_2}_U)$. The back condition on path bisimulation ensures that $(x|^{Q_1}_U, y|^{Q_2}_U) \in Z_U$. Define the maps $f : P \to Q_1$ and $g : P \to Q_2$ with the projections. This immediately gives us naturality. The fact that these maps are open is given by the forward condition on path bisimulation.

We now show that P is a sheaf. Starting from locality, suppose that, for a given a covering $(U_i)_{i\in I}$ of an open set U, we have two elements $(x, y), (x', y') \in P(U) = Z_U$ that agree on all the restrictions, that is, such that $(x|^{Q_1}_{U_i}, y|^{Q_2}_{U_i}) = (x'|^{Q_1}_{U_i}, y'|^{Q_2}_{U_i})$ for all i. This in particular means that $x|^{Q_1}_{U_i} = x'|^{Q_1}_{U_i}$ for all i, so by locality on Q_1 we obtain that $x = x'$. Similarly we conclude that $y = y'$, and so $(x, y) = (x', y')$. The gluing condition is given by assumption.

From right to left, suppose there is a span of open maps, where P is the sheaf at the vertex and f, g are the two open maps. Define

$$(x, y) \in Z_U \text{ iff there is } z \text{ in } P(U) \text{ such that } f(z) = x \text{ and } g(z) = y$$

By Theorem 3.2.9 we know that this is a path bisimulation. Now observe that the presheaf Z so defined is essentially (is in bijection with) the quotient of P under the equivalence relation R defined as

$$(x, x') \in R \text{ iff } f(x) = f(x') \text{ and } g(x) = g(x')$$

This is because we can identify each equivalence class of R with a pair in Z. Then Z is the coequalizer of P (up to iso) for the two projections $R \to P$. Since sheaves are closed under small colimits, Z is a sheaf, thus it satisfies gluing. □

3.5.2 Path bisimulations and co-spans of open maps

Another important notion of bisimulation in the coalgebra literature is given by the dual of the span, i.e. the *co-span* of coalgebra morphisms. This is often called a *behavioural equivalence*. The same concept can be applied to sheaves, that is, we may consider co-spans of open maps rather than spans. It turns out that we can characterize the existence of a co-span of open maps precisely in terms of a concrete notion of path bisimulation. First, we introduce a two axioms for path bisimulations, mimicking the corresponding axiom for sheaves:

3.5.4. Definition. [Axiom 1] Suppose we are given a covering $(U_i)_{i \in I}$ of an open set U, two sheaves $Q_1, Q_2 : Open(\mathbb{X})^{op} \to \mathbf{Set}$ and a path bisimulation Z between them. We say Z satisfies *Axiom 1* if for all $x \in Q_1(U)$ and $y \in Q_2(U)$ such that $(x|_{U_i}^{Q_1}, y|_{U_i}^{Q_2}) \in Z_{U_i}$ for all $i \in I$, we have $(x, y) \in Z_U$.

3.5.5. Definition. [Axiom 2] Suppose given a covering $(U_i)_{i \in I}$ of an open set U, two presheaves $Q_1, Q_2 : Open(\mathbb{X})^{op} \to \mathbf{Set}$ and a path bisimulation Z between them. The relation Z satisfies *Axiom 2* if the following is the case: whenever there are two families $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$ such that $x_i \in Q_1(U_i)$ and $y_i \in Q_2(U_i)$ for all i and moreover for all i, j $(x_i|_{U_i \cap U_j}^{Q_1}, y_j|_{U_i \cap U_j}^{Q_2}) \in Z$ there exist two elements $x \in Q_1(U)$ and $y \in Q_2(U)$ such that $(x, y) \in Z$ and, for all i, $(x|_{U_i}^{Q_1}, y_i) \in Z$ and $(y|_{U_i}^{Q_2}, x_i) \in Z$.

Finally, we need a little technical side condition, that we borrow from [63]:

3.5.6. Definition. A path bisimulation Z is said to be *di-functional* if $(x, y) \in Z$, $(x', y) \in Z$ and $(x', y') \in Z$ entail $(x, y') \in Z$.

We can now state our characterization result:

3.5.7. Theorem. *Two sheaves Q_1 and Q_2 are related by a co-span of open maps*

$$Q_1 \to P \leftarrow Q_2$$

where P is a sheaf, if and only if they are related by a di-functional path bisimulation that satisfies Axioms 1 and 2.

Proof:
From left to right, assume there are a sheaf P and two open maps $f : Q_1 \to P$ and $g : Q_2 \to P$. Define

$$Z_U = \{(p, q) \in Q_1(U) \times Q_2(U) | f_U(p) = g_U(q)\}$$

Clearly $Z = \bigcup_U Z_U$. We start with the forward condition of path bisimulation. Suppose $(x, y) \in Z_U$, $\iota : U \to U'$ and there is $x' \in U'$ such that $x'|_U^{Q_1} = x$. We need to show that there is $y' \in Q_2(U')$ such that $y'|_U^{Q_2} = y$ and $(x', y') \in Z_{U'}$, that is, $f_{U'}(x') = g_{U'}(y')$. By naturality we know that $f_{U'}(x')|_U^P = f_U(x'|_U^{Q_1}) = f_U(x) = g_U(y)$, so by weak pullback we obtain $y' \in Q_2(U')$ such that $y'|_U^{Q_2} = y$ and $f_{U'}(x') = g_{U'}(y')$.

For the backward condition suppose $(x, y) \in Z_U$ and $\iota : U' \to U$. We need to show that $x|_{U'}^{Q_1} = x'$ and $y|_{U'}^{Q_2} = y'$ are in relation: $(x', y') \in Z_{U'}$, that is, $f_{U'}(x') = g_{U'}(y')$. This follows immediately by the naturality of f and g and $f_U(x) = g_U(y)$.

We proceed to check that Axiom 1 holds. Suppose given a covering $(U_i)_{i\in I}$ of U. Say there are $x \in Q_1(U)$ and $y \in Q_2(U)$ such that $(x|_{U_i}^{Q_1}, y|_{U_i}^{Q_2}) \in Z_{U_i}$ for all $i \in I$. We want to show that $(x, y) \in Z_U$, that is, that $f_U(x) = g_U(y)$. We have for every i that

$$f_U(x)|_{U_i}^P = f_{U_i}(x|_{U_i}^{Q_1}) \tag{3.4}$$
$$= g_{U_i}(y|_{U_i}^{Q_2}) \tag{3.5}$$
$$= g_U(y)|_{U_i}^P \tag{3.6}$$

where the first and last step are given by the naturality of f and g and the second is given by our assumption. Since $f_U(x)$and $g_U(y)$ agree on all the restrictions we can apply the locality property of the sheaf P to obtain $f_U(x) = g_U(y)$. Now for Axiom 2. Suppose given a covering $(U_i)_{i\in I}$ of U. Say there are two families $(x_i)_{i\in I}$ and $(y_i)_{i\in I}$ such that for all i $x_i \in Q_1(U_i)$ and $y_i \in Q_2(U_i)$ and moreover for all i, j $(x_i|_{U_i\cap U_j}^{Q_1}, y_j|_{U_i\cap U_j}^{Q_2}) \in Z$. We need to show that there are $x \in Q_1(U)$ and $y \in Q_2(U)$ such that $(x, y) \in Z$ and, for all i, $(x|_{U_i}^{Q_1}, y_i) \in Z$ and $(y|_{U_i}^{Q_2}, x_i) \in Z$. Our definition entails that $f_{U_i\cap U_j}(x_i|_{U_i\cap U_j}^{Q_1}) = g_{U_i\cap U_j}(y_j|_{U_i\cap U_j}^{Q_2})$. By naturality of f and g we get that $f_{U_i}(x_i)|_{U_i\cap U_j}^P = g_{U_j}(y_j)|_{U_i\cap U_j}^P$. When $i = j$ this means that

$f_{U_i}(x_i) = g_{U_i}(y_i)$, hence we have a family of objects $(f_{U_i}(x_i) = g_{U_i}(y_i))_{i\in I}$ such that $f_{U_i}(x_i) = g_{U_i}(y_i) \in P(U_i)$ and the elements of this family agree at the intersections. Thus we can apply gluing in P to obtain $t \in P(U)$ such that $t|^P_{U_i} = f_{U_i}(x_i) = g_{U_i}(y_i)$ for all i.

Pick an index i; by the fact that f and g are open maps we obtain by weak pullback $x \in Q_1(U)$ and $y \in Q_2(U)$ such that $f_U(x) = t = g_U(y)$, which means $(x, y) \in Z_U$, and $x|^{Q_1}_{U_i} = x_i$ and $y|^{Q_2}_{U_i} = y_i$. This entails that $f_{U_i}(x|^{Q_1}_{U_i}) = t|^P_{U_i} = g_{U_i}(y_i)$ and $g_{U_i}(y|^{Q_2}_{U_i}) = t|^P_{U_i} = f_{U_i}(x_i)$. Now take $j \neq i$. We have that $f_{U_j}(x|^{Q_1}_{U_j}) = f_U(x)|^P_{U_j} = t|^P_{U_j} = g_{U_j}(y_j)$, where the first step is by naturality and the last is a consequence of gluing. Similarly we can show that $g_{U_j}(y|^{Q_1}_{U_j}) = f_{U_j}(x_j)$.

Difunctionality is immediate by the definition of the relation: if x, y, x' and y' are all sent to the same object then the relation will hold between x' and y'.

From right to left, suppose there is a difunctional path bisimulation between Q_1 and Q_2 satisfying Axiom 1 and 2. Take Eq_U to be the smallest equivalence relation containing Z_U. Define the sheaf P as follows

$$U \mapsto Q_1(U) + Q_2(U) \backslash Eq_U$$
$$\iota : U \to U' \mapsto P(\iota) : Q_1(U') + Q_2(U') \backslash Eq_{U'} \to Q_1(U) + Q_2(U) \backslash Eq_U$$

where $P(\iota)([x]) = [Q_l(\iota)(x)]$ if $x \in Q_l(U')$, for $l \in \{1, 2\}$. Notice that this definition automatically makes $[-] : Q_l \to P$ a natural transformation for $l = 1$ and $l = 2$. We sometimes omit the subscript when it is clear from the context.

We first show that $P(\iota)$ is well defined. Suppose $x \neq y$ and $[x] = [y]$, we want to show that $P(\iota)([x]) = P(\iota)([y])$. Since $[x] = [y]$, only two scenarios can occur. Suppose $(x, y) \in Z_{U'}$. Then $P(\iota)([x]) = [Q_1(\iota)(x)]$ and $P(\iota)([y]) = [Q_2(\iota)(y)]$. By backward condition of path bisimulation we obtain from $(x, y) \in Z_{U'}$ that $(x|^{Q_1}_U, y|^{Q_2}_U) \in Z_{U'}$, so $[Q_1(\iota)(x)] = [Q_2(\iota)(y)]$ and we are done. Now suppose x and y are in relation because of a zig-zag of relations in $Z_{U'}$. Applying our previous argument to every pair in $Z_{U'}$ we get, by transitivity of equality, that $[Q_1(\iota)(x)] = [Q_2(\iota)(y)]$. We now show that $[-]$ is an open map. Suppose $\iota : U \to U'$ and say that there are $x_1 \in Q_1(U)$ and $[x_2] \in P(U')$ such that $[x_1] = P(\iota)([x_2])$. We know that $[x_1] = P(\iota)([x_2]) = [x_2|^{Q_l}_U]$, for some $l \in \{1, 2\}$. Hence there is a zig-zag of Z_U edges between x_1 and $x_2|^{Q_l}_U$. Starting from x_2, we apply the forward condition to all the edges of the zig-zag (this is an argument by induction, similar to the one in [74]); in this way we obtain an element $x' \in Q_1(U')$ such that there is a zig-zag of $Z_{U'}$ edges between x_2 and x', hence $[x_2] = [x']$, and $x'|^{Q_1}_U = x_1$.

We proceed to show that P is a sheaf, beginning with locality. Suppose given a covering $(U_i)_{i\in I}$ of U. Consider $[x], [y] \in P(U)$ such that $[x]|^P_{U_i} = [y]|^P_{U_i}$ for all i. Because the bisimulation includes the roots, we can always assume that each equivalence class $[x]_U$ contains at least a member of $Q_1(U)$ and a member of $Q_2(U)$. So we can take $x \in Q_1(U)$ and $y \in Q_2(U)$. By $[x]|^P_{U_i} = [y]|^P_{U_i}$ we can infer that $[x|^{Q_1}_{U_i}] = [y|^{Q_2}_{U_i}]$ for every i. By difunctionality we can conclude that $(x|^{Q_1}_{U_i}, y|^{Q_2}_{U_i}) \in Z_{U_i}$ for all i. Then Axiom 1 on Z allows us to infer that

$(x, y) \in Z_U$, hence $[x] = [y]$. Finally we prove that P has the gluing property. Suppose given a covering $(U_i)_{i\in I}$ of U. Suppose there is a family $([x_i])_{i\in I}$ with $[x_i] \in P(U_i)$ such that, for all $i, j \in I$, $[x_i]|^P_{U_i\cap U_j} = [x_j]|^P_{U_i\cap U_j}$. We want to find $[x] \in P(U)$ such that $[x]|^P_{U_i} = [x_i]$ for all i. We know $[x]_U$ contains at least a member of $Q_1(U)$ and a member of $Q_2(U)$. So we can infer that there are two families $(p_i \in Q_1(U_i))_{i\in I}$ and $(q_i \in Q_2(U_i))_{i\in I}$ such that $[p_i] = [q_i] = [x_i]$. So from $[x_i]|^P_{U_i\cap U_j} = [x_j]|^P_{U_i\cap U_j}$ we infer that $[p_i|^{Q_1}_{U_i\cap U_j}] = [q_j|^{Q_2}_{U_i\cap U_j}]$. By difunctionality it must be that $(p_i|^{Q_1}_{U_i\cap U_j}, q_j|^{Q_2}_{U_i\cap U_j}) \in Z_{U_i\cap U_j}$, and this for all i and j. By Axiom 2 we conclude that there are $p \in Q_1(U)$ and $q \in Q_2(U)$ such that $(p, q) \in Z_U$ and for all i $p|^{Q_1}_{U_i} = q_i$ and $q|^{Q_2}_{U_i} = p_i$. Take $[x] = [p] = [q]$: we have for all i that $[x]|^P_{U_i} = [p]|^P_{U_i} = [p|^{Q_1}_{U_i}] = [q_i] = [x_i]$. This concludes the proof. □

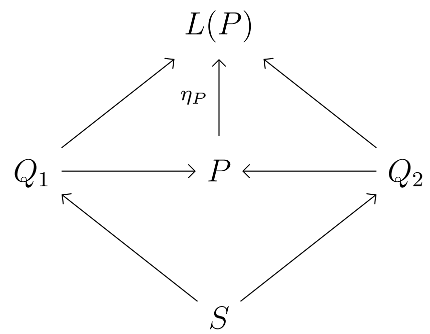
It is easy to see that every path bisimulation is contained in a difunctional path bisimulation, its "difunctional closure". But since we cannot assume that the difunctional closure operation preserves Axiom 1 and 2, we have to state difunctionality as an explicit premise of the previous theorem.

3.5.3 Spans versus co-spans

To sum up, we have studied two different categorical notions of behavioural equivalence for sheaves: existence of a span of open maps in $\mathbf{Sh}(\mathbb{X})$, and existence of a co-span of open maps in $\mathbf{Sh}(\mathbb{X})$. How are spans and co-spans related to each other in the category of sheaves? One direction is clear, co-spans entail spans. For the other direction, we can characterize exactly when the pushout of an open span is open. Let L be the left adjoint to the inclusion of $\mathbf{Sh}(\mathbb{X})$ into $\mathbf{PrSh}(\mathbb{X})$, usually known as the sheafification functor.

3.5.8. PROPOSITION. *Consider a span $Q_1 \leftarrow S \rightarrow Q_2$ in $\mathbf{Sh}(\mathbb{X})$. Then the pushout of this span in $\mathbf{Sh}(\mathbb{X})$ is an open co-span if and only if the unit map $\eta_P : P \rightarrow L(P)$ is open, where P is the pushout of the span in $\mathbf{PrSh}(\mathbb{X})$.*

Proof:
The pushout in $\mathbf{Sh}(\mathbb{X})$ is given by the following commutative diagram:



Here the four bottom arrows are the pushout square in $\mathbf{PrSh}(\mathbb{X})$, the horizontal arrows are open by Proposition 3.5.2 and the vertical arrow represents the unit map. So if the unit map is open, then the upper diagonal arrows are open since open maps are closed under composition. On the other hand, suppose that the upper diagonal arrows are open. We want to show that the vertical arrow is open: by the "quotient axiom" for open maps in [73] it suffices to show that the horizontal arrows are epimorphisms.

Since Q_1 and Q_2 are sheaves, and hence rooted presheaves, it is easy to see that P is also a rooted presheaf. Now, given any $x \in P(C)$, the restriction of x to the empty set will give the root of P; since the root of Q_1 is mapped to the root of P and the natural transformation $Q_1 \to P$ is open there must be $x' \in Q_1(C)$ that is mapped to x and restricts to the root. Thus the map $Q_1(C) \to P(C)$ is surjective for all C and thus the transformation $Q_1 \to P$ is an epimorphism. The same line of reasoning can be applied to show that the natural transformation $Q_1 \to P$ is an epimorphism. □

3.6 Hybrid path logic

In order to capture additional properties of presheaves, such as the sheaf conditions, we can enrich the path logic with extra expressive power. The suggestion that presents itself is to go to *hybrid logic*, which has been showed to be a powerful and yet well-behaved extension of standard modal logic.[10]

We define the syntax of *hybrid path logic* $\mathbf{HPL}_\kappa(\mathbb{X}, N, At)$ for $\mathbb{X}$ a topological space, a regular cardinal κ and over a set of nominals N, by the following grammar:

$$\varphi ::= p \mid i \mid @_i\varphi \mid \bigvee \Gamma \mid \neg\varphi \mid \langle U, V\rangle\varphi \mid \overline{\langle U, V\rangle}\varphi$$

Here i ranges over N, U and V range over open sets of $\mathbb{X}$, p ranges over At and Γ ranges over sets of formulas of size $< \kappa$.

A *presheaf model* for this language is a rooted presheaf P over $\mathbb{X}$ together with a map $A : N \to \biguplus\{P(U) \mid U \in Open(\mathbb{X})\}$ (where $\biguplus$ is the disjoint union) and a valuation V for At. Truth conditions of formulas in a model (P, A) at some $w \in P(U)$ are defined as before, with the added clauses:

- $(P, A, w) \vDash i$ if and only if $A(i) = w$
- $(P, A, w) \vDash @_i\varphi$ if and only if $(P, A, A(i)) \vDash \varphi$

3.6.1. Definition. We say that φ is true in (P, A, V), written $(P, A, V) \vDash \varphi$, if $(P, A, V, r) \vDash \varphi$ where r is the root of P. We say that φ is valid in P, written $P \vDash \varphi$, if $(P, A, V) \vDash \varphi$ for every A and every V.

[10] See Chapter 14 in [33].

3.6.1 Expressing locality and gluing

Now, given a space $\mathbb{X}$, which we assume to be infinite, let κ be a regular cardinal greater than 2^ξ where ξ is the number of open sets of $\mathbb{X}$. Assuming the axiom of choice we can take this to be the successor of 2^ξ. Let N be a set of nominals with $2^\xi \leq |N| < \kappa$. Then consider the following formulas of $\mathbf{HPL}_\kappa(\mathbb{X}, N)$:

Loc: For any cover $\{U_i\}_{i\in I}$ of an open set U of $\mathbb{X}$, pick nominals $j, k, \{l_i\}_{i\in I}$ and construct the formula:

$$\bigwedge_{i\in I} @_{l_i}\langle U_i, U\rangle j \wedge \bigwedge_{i\in I} @_{l_i}\langle U_i, U\rangle k \to @_j k$$

Then we define Loc to be the conjunction of all these formulas, corresponding to all the covers of open sets in $\mathbb{X}$. The conjunction is well defined since there are at most 2^ξ covers to consider.

Glu: For any cover $\{U_i\}_{i\in I}$ of an open set U of $\mathbb{X}$, pick nominals $\{l_i\}_{i\in I}$ and construct the formula:

$$\bigwedge_{i,j\in I} @_{l_i}\overline{\langle U_i \cap U_j, U_i\rangle}\langle U_i \cap U_j, U_j\rangle l_j \to \langle \emptyset, U\rangle \bigwedge_{i\in I} \overline{\langle U_i, U\rangle} l_i$$

We take Glu to be the conjunction of all these formulas.

These two formulas closely follow the conditions of locality and gluing. The former states that if two elements in $P(U)$ (labeled by nominals j and k) have the same restrictions (labeled by $\{l_i\}_{i\in I}$) then they must coincide. The latter encodes the fact that if a family of elements (labeled by $\{l_i\}_{i\in I}$) agrees on intersections - in the sense that if we restrict l_i to $U_i \cap U_j$ then we can extend again to U_j and obtain l_j - then there exists an element in $P(U)$ (which need not be labeled) which restricts to all the l_i's.

The proof of the following result is a simple check, but we list it as a theorem since we think it has some importance.

3.6.2. Theorem. *A rooted presheaf P is a sheaf if, and only if, $P \models \mathsf{Loc} \wedge \mathsf{Glu}$.*

It follows, of course, that validity of formulas in hybrid path logic is not preserved by path bisimulations. However, truth in a model is easily seen to be preserved by a natural extension of path bisimulations:

3.6.3. Definition. Let (P, A, V) and (P', A', V') be presheaf models. Then a *nominal path bisimulation* is a path bisimulation between (P, V) and (P', V') such that, for every nominal i, $A(i)$ is related to x by this path bisimulation if and only if $x = A'(i)$, and vice versa.

3.6.4. Proposition. *Formulas of* $\mathbf{HPL}_\kappa(\mathbb{X}, N, \mathsf{Var})$ *are invariant under nominal path bisimulations.*

Proof:
The proof is a routine induction on the structure of the formula; we only show the cases involving the nominals. Let Z be a nominal path bisimulation between models (P, A, V) and (P', A', V'), suppose $(w, w') \in Z$. If $(P, A, w) \vDash i$ then by definition $A(i) = w$, thus by the property of nominal path bisimulation we have that $A'(i) = w'$ and thus $(P', A', w') \vDash i$. Now assume that $(P, A, w) \vDash @_i\varphi$: by definition $(P, A, A(i)) \vDash \varphi$. Take y in the second model such that $A'(i) = y$, again by the property of nominal path bisimulation we must have $(A(i), y) \in Z$. By induction hypothesis $(P', A', y) \vDash \varphi$, thus $(P', A', w') \vDash @_i\varphi$. □

3.6.2 Decidability

In this subsection we show that the finitary fragment of the hybrid path logic is decidable on the class of presheaf models, when the underlying poset of opens is finite. This is achieved by proving the bounded model property: every formula of the language that is satisfiable in a presheaf model is satisfiable in a finite presheaf model. The core idea is that one such presheaf models is essentially a Kripke structure where relations have to satisfy a 'backward functionality' requirement. We thus design a filtration that preserves this property and exploit the typing given by the category to keep everything finite.

3.6.5. Proposition. *If the poset* $Open(\mathbb{X})$ *is finite then the logic* $\mathbf{HPL}_\omega(\mathbb{X}, N, At)$ *has the bounded model property.*

Proof:
Suppose P is a presheaf model and $P, A, x \vDash \varphi$. It is easy to see that φ is semantically equivalent to a formula φ' where we have removed any reference to identity arrows, since $\langle U, U\rangle\psi \leftrightarrow \psi \leftrightarrow \overline{\langle U, U\rangle}\psi$ is a validity for any open U. Thus it is enough to show that φ' is satisfied in a finite presheaf model.

The strategy of the proof is to take a quotient of P in order to turn it onto a finite model; this technique is known as filtration. Let $sub(\varphi')$ be the set of subformulas of φ'. Let $back(sub(\varphi'))$ be the smallest set containing $sub(\varphi')$ and such that:

- for any $U \neq U'$, if $\psi \in back(sub(\varphi'))$ then $\overline{\langle U, U'\rangle}\psi \in back(sub(\varphi'))$;
- for any open U, $\overline{\langle \emptyset, U\rangle}\top \in back(sub(\varphi'))$;
- $\top \in back(sub(\varphi'))$.

In other words, we close the set of subformulas under application of backward modalities and we add a typing mechanism. Note that this set is still closed under subformulas, although it might be infinite.

We now perform the smallest filtration over the relational structure associated to P[11]: we identify elements that satisfy the same formulas in $back(sub(\varphi'))$ and for any inclusion $U \to U'$ we define a new relation on equivalence classes as follows

$$([w],[v]) \in R^s_{U,U'} \quad \text{iff} \quad \exists w' \in [w], v' \in [v] \quad P(U \to U')(v') = w'$$

We claim this structure corresponds to a finite presheaf model P'. The presheaf P' is defined on objects as

$$P'(U) = \{[w] | \exists w' \in [w],\ P, A, w \vDash \overline{\langle \emptyset, U \rangle} \top\}$$

that is, U is mapped to the collection of equivalence classes containing objects in $P(U)$. On arrows, we map an inclusion $U \to U'$ to the relation $R^s_{U,U'}$ defined above. On this frame we define a model by taking a valuation $V'(p) = \{[w] | \exists w' \in [w]\, w' \in V(p)\}$ and an interpretation $A'(i) = [A(i)]$.

We need to check that this model is indeed a presheaf by showing the backward functionality of $R^s_{U,U'}$. The existence follows immediately by the functionality of $P(U \to U')$. For uniqueness suppose $([w],[v]) \in R^s_{U,U'}$ and $([t],[v]) \in R^s_{U,U'}$. By definition there are $w' \in [w], v', v'' \in [v]$ and $t' \in [t]$ such that $P(U \to U')(v') = w', P(U \to U')(v'') = t'$. If $P, A, w' \vDash \psi$ with $\psi \in back(sub(\varphi'))$ then by semantics $P, A, v' \vDash \overline{\langle U, U' \rangle} \psi$. By construction we also have $\overline{\langle U, U' \rangle} \psi \in back(sub(\varphi'))$ and thus $v', v'' \in [v]$ must entail $P, A, v'' \vDash \overline{\langle U, U' \rangle} \psi$. This in turn implies $P, A, t' \vDash \psi$. The converse argument shows that $P, A, t' \vDash \psi$ entails $P, A, w' \vDash \psi$, thus $[w] = [t]$.

Finally, the usual argument by induction shows that $P', A', [x] \vDash \varphi'$ holds; we only cover the cases with nominals. If $P, A, x \vDash i$ then by definition $x = A(i)$, so $[x] = [A(i)] = A'(i)$ and $P', A', [x] \vDash i$. If $P, A, x \vDash @_i \xi$ then by definition $P, A, A(i) \vDash \xi$ so by induction hypothesis $P', A', [A(i)] \vDash \xi$. By construction this entails $P', A', A'(i) \vDash \xi$ thus $P', A', [x] \vDash @_i \xi$. The fact that P' is finite and has an upper bound on the size is highlighted in the following Lemma. □

The finiteness of P' is a consequence of the typing imposed by the category: the set $back(sub(\varphi'))$ is 'finite up to semantic equivalence' because backward modalities can only be applied meaningfully in patterns that are constrained by the poset $Open(\mathbb{X})$.

3.6.6. Lemma. *The presheaf model P' defined in Proposition 3.6.5 is finite and the bound on its size is computable from φ'.*

Proof:
We show that P' is finite, that is, there are finitely many equivalence classes, despite

[11] See [32] for an introduction to the filtration method.

the fact that we are filtrating with an infinite set of formulas. Since we are using a finite poset category $Open(\mathbb{X})$, we can recursively define the height or distance $h(U)$ between an open U and the empty set: take $h(U) = max\{h(U')+1|U' \subset U\}$. This in particular entails $h(\emptyset) = 0$. Since the poset is finite every object has a finite height.

Let $num(n) := |\{U \in Open(\mathbb{X})|h(U) = n\}|$, that is, the number of opens of height n in the poset. Since we are using a finite poset, $num(n)$ is finite for every n. Let t_n be the amount of equivalence classes in the images of opens of height n:

$$t_n = |\bigcup\{P'(U')|h(U') = n\}|$$

Claim: for every natural number m, t_m is finite.

We proceed by induction on m. When $m = 0$ we have $t_0 = |\bigcup\{P'(U')|h(U') = 0\}| = |P'(\emptyset)|$. Note that, due to the typing of the modality, if $w \in P(\emptyset)$ then $w \vDash \overline{\langle U', U\rangle}\psi$ iff $U' = U = \emptyset$; in other words, all backward modalities beside the identity $\overline{\langle \emptyset, \emptyset\rangle}$ are false at worlds of type $\emptyset$. By construction, in $back(sub(\varphi'))$ there are no identity arrows featuring in any formulas. Since worlds in $P(\emptyset)$ all agree on backward modalities, two equivalence classes in $P'(\emptyset)$ can be distinct only if they assign different truth values to formulas in $sub(\varphi')$. Hence there are at most $2^{|sub(\varphi')|}$ equivalence classes in $|P'(\emptyset)|$.

Now suppose $m = n + 1$. By IH t_n is finite. We claim that t_{n+1} has upper bound $2^{|sub(\varphi')|} \times t_n^{num(n)} \times num(n + 1)$, which is finite due to IH. To show this we prove that if $h(U) = n + 1$ then $|P'(U)|$ is bounded by $2^{|sub(\varphi')|} \times t_n^{num(n)}$. The upper bound for t_{n+1} is then obtained from this figure multiplying by $num(n+1)$, the number of objects of height $n + 1$.

So suppose $h(U) = n + 1$. For $[w] \in P'(U)$, let $\Phi_{[w]} = \{\psi \in sub(\varphi')|[w] \vDash \psi\}$. For $\Phi \subseteq sub(\varphi')$ define $Eq_\Phi = \{[w] \in P'(U)|\Phi_{[w]} = \Phi\}$; this is the set of all equivalence classes in $P'(U)$ that satisfy exactly the same subset of formulas of $sub(\varphi')$, namely Φ. A little reflection shows that $|P'(U)|$ is bounded by $\sum_{\Phi \in 2^{sub(\varphi')}} |Eq_\Phi|$: there is an obvious injection from $P'(U)$ into the disjoint union $\biguplus_{\Phi \in 2^{sub(\varphi')}} Eq_\Phi$ given by $[w] \mapsto (\Phi_{[w]}, [w])$. Thus if we can show that $|Eq_\Phi|$ is bounded by $t_n^{num(n)}$ for any Φ then we can conclude that $|P'(U)|$ is bounded by $2^{|sub(\varphi')|} \times t_n^{num(n)}$.

Notice that equivalence classes in Eq_Φ agree on formulas in $sub(\varphi')$ and by construction satisfy $\overline{\langle \emptyset, U\rangle}\top$, thus two equivalence classes $[w]$ and $[v]$ in Eq_Φ can be distinct only if there is $\overline{\langle U', U\rangle}\psi \in back(sub(\varphi'))$ such that $[w] \vDash \overline{\langle U', U\rangle}\psi$ and $[v] \nvDash \overline{\langle U', U\rangle}\psi$. In other words only if there is $U' \subset U$ such that $P'(U' \subset U)([w]) \neq P'(U' \subset U)([v])$. Without loss of generality we can take U' to have height n.

Every equivalence class $[w]$ in Eq_Φ can be mapped to the subset containing all the images of $[w]$ 'at height n': $[w] \mapsto \{P'(U' \subseteq U)([w])|U' \subseteq U, h(U') = n\}$. The latter is a subset of $\bigcup\{P'(U')|h(U') = n\}$. From what we argued in the previous paragraph, two distinct equivalence classes $[w]$ and $[v]$ in Eq_Φ generate two distinct subsets of $\biguplus\{P'(U')|h(U') = n\}$, because they must diverge on at least one image. We thus have an injection of Eq_Φ into the powerset $\wp(\bigcup\{P'(U')|h(U') = n\})$.

Furthermore, each image of the mapping $[w] \mapsto \{P'(U' \subseteq U)([w])|U' \subseteq U, h(U') = n\}$ has the same size: there will a different image of $[w]$ for each subset of U of height n. Suppose there are k such subsets of U, we thus have an injection of Eq_Φ into the set of subsets of $\bigcup\{P'(U')|h(U') = n\}$ of size k. Consequently $|Eq_\Phi|$ has upper bound $C(t_n, k) \leq t_n^k \leq t_n^{num(n)}$, where the latter inequality is due to $k \leq num(n)$ and $C(t_n, k)$ is the number of subsets of t_n of size k.

Such upper bound on $|Eq_\Phi|$ gives us the desired upper bound on $|P'(U)|$ and thus the correct upper bound on t_{n+1}. This concludes our induction. This argument shows that the bound on the number of equivalence classes at height n is given recursively by a function g defined as follows: $g(0) := 2^{|sub(\varphi')|}$ and $g(n+1) = 2^{|sub(\varphi')|} \times g(n)^{num(n)} \times num(n+1)$. Calling m be the maximum height of an open in the poset, we obtain the final upper bound for the whole model by summing over all the bounds for all the heights up to m: $\sum_{0 \leq n \leq m} g(n)$.

The important thing to notice in this calculation is that $num(n)$ and m are parameters that are fixed by the poset $Open(\mathbb{X})$: they are the same for every formula in the hybrid path logic for $Open(\mathbb{X})$. Thus the only variable in computing this upper bound is φ'.[12] □

We can finally state and prove our decidability result.

3.6.7. THEOREM. *If the poset* $Open(\mathbb{X})$ *is finite then the logic* $\mathbf{HPL}_\omega(\mathbb{X}, N, At)$ *is decidable.*

Proof:
Given a formula $\phi \in \mathbf{HPL}_\omega(\mathbb{X}, N, At)$, let t be the bound on the size of the finite model for ϕ given by Proposition 3.6.5. Generate all the presheaf models up to size t and test for satisfiability. If the formula is satisfiable then we have a model satisfying it, if we do not find a model of size less or equal than t then by Proposition 3.6.5 we can conclude that ϕ is not satisfiable. □

The bounded model property does not hold for infinite poset categories, indeed not even the finite model property. Consider for example the poset category consisting of the natural numbers with arrows $n \to n-1$ and $0 \to 0$. Because of backward functionality, there is no finite presheaf model satisfying the formula $\langle Id_1 \rangle \top$: each relation $R_{n \to n-1}$ forces us to add another element, ad infinitum.

3.7 Conclusions

We addressed the issue of the expressivity of path logic, which constitutes an extension of the logic $\mathbf{LTTS^C}$ presented in the previous chapter. We first proved a characterization theorem for the finitary fragment and successively studied which

[12] This result can be improved in two directions, generalizing to generic finite poset categories and providing a slightly tighter bound on the size of the model. Since these added level of complications would not add any novel insight, we opted for this version of the proof.

interesting properties of presheaves can be encoded in the language. We covered some examples, among which properties of the sheaves of sections used in Topology and of the sheaf-theoretic approach to contextuality. From previous results we inferred that such formulas are invariant for path bisimulation on presheaves.

However, we showed that the two defining properties of a sheaf fall outside the scope of path logic. We devoted a section to the study of path bisimulation in the realm of sheaves. In this setting we investigated the relations between the different notions of behavioural equivalence, characterizing the existence of a span of open maps as well as the existence of a co-span. Finally, we suggested how path logic has to be enriched in order to capture the key features of sheaves. We demonstrated how the properties of the category can induce good behaviour of the corresponding logic by proving a decidability result for the logics of finite poset categories $Open(\mathbb{X})$.

The observations of this chapter seem to suggest that path logic might be the right logic to express the extension and restriction of contexts, where the contexts are arranged in a finite poset category. The next chapter moves the first steps in this direction, investigating how a logic of varying coalitions (qua contexts) can encode key concepts and results of Social Choice Theory.

Two more issues are prompted by the results of this chapter. The first one pertains to the bounded model property proved in the last section: we conjecture that the features of the category 'in the background' could be exploited to push down the size of the finite model and obtain better complexity results in special cases. The second one is more open ended and concerns the connection of path logic for topological spaces with the usual topological semantics for modal languages [93, 94] and the sheaf semantics for first-order modal logic [13].

Chapter 4

A modal logic for Social Choice Theory

4.1 Introduction

In previous chapters we studied path logic and its expressivity at an abstract level; we now showcase what well-chosen fragments can achieve in concrete areas. The case study of this chapter is the formalization of Social Choice Theory. We design a special-purpose logic to capture the key features of this particular setting and later explain how this logic can be translated back to path logic, showing the unifying potential of the latter formalism.

More precisely, we present a modal logic for social choice functions and describe how it can be used to prove three seminal impossibility results in this field. We discuss how this logic fares in comparison to other languages proposed for the same task and offer some remarks on the implementation of the logic by describing how to feed it to a SAT solver. The modality in this language encodes the capability of a coalition to enforce the truth of a certain formula, given that the people outside the coalition maintain their course of action. This suggests that the central aspect of this theory is the possibility to track what happens to the preferences expressed by a coalition of agents when said coalition is expanded or shrunk. Following this intuition we explicate how a social choice function can be understood as a presheaf model, where the base category is the poset of all possible coalitions.

The remainder of the chapter is organised as follows. Next Section sets the stage by introducing Social Choice Theory and the existing work on its formal foundations. Section 4.3 recalls the definition of a social choice function (SCF), and then introduces our logic of SCF's and establishes completeness for it. This is followed up in Section 4.4, where we show how various concepts of interest for social choice theory can be modelled in this logic. This includes a discussion of the universal domain assumption and encodings of desirable properties of SCF's, such as Pareto efficiency and monotonicity. The three theorems are encoded and then proved in Section 4.5. A translation into propositional logic, offering a means of implementation via a SAT solver, is presented in Section 4.6. Section 4.7 discusses related work and Section 4.8 explains the connections with path logic and presheaf models. Section 5.9 concludes.

4.2 Social Choice Theory

Social choice theory is the study of mechanisms for collective decision making [116]. This includes voting rules as mechanisms to collectively make political decisions, and consequently social choice theory is chiefly associated with the disciplines of political science and economics. But similar mechanisms can also be used to make decisions in multi-agent systems, to coordinate the actions of individual agents, to resolve conflicts between them, and to bundle their information and expertise [36]. Closely related applications of social choice theory in computer science furthermore include recommender systems [103], Internet search engines [7], and crowdsourcing [89].

This widening of the scope of social choice theory has renewed interest in the formal foundations of the field. As we are designing ever more specialised social choice mechanisms for novel types of tasks, better tools to analyse the formal properties of these mechanisms are needed. Specifically, there is now a growing literature on the formal verification of social choice mechanisms by means of logical modelling and the use of techniques from automated reasoning [1, 22, 37, 54, 57, 64, 97, 115, 118, 122]. We will review some of the contributions to this field in Section 4.7.

An obvious yardstick against which to measure different approaches to the formalisation of social choice frameworks is Arrow's Theorem [11], *the* seminal result in the field, which shows that it is impossible to design preference aggregation mechanisms for three or more alternatives that are Pareto efficient and for which the relative ranking of two alternatives is based only on the rankings for the same two alternatives submitted by the individual voters.

Recent work has modelled the Arrovian framework in propositional logic [115], first-order logic [64], higher-order logic [97, 122], and a tailor-made modal logic [1]. Some of this work has resulted in methods to prove Arrow's Theorem either automatically [115] or semi-automatically [97, 122], while other work has generated logical formalisations of the theorem that are easily accessible to humans and thus helpful in deepening our understanding of social choice [1, 64]. A shortcoming of the latter contributions, however, is that they have so far not resulted in a full proof of Arrow's Theorem or similar results *within* the chosen logical framework itself.[1] Rather, such work has proceeded by showing that a given logical system is complete w.r.t. an appropriate class of models of social choice theory, thereby proving that a rendering of Arrow's Theorem in the logical language in question must be a theorem of that logic. That is, such work has derived results about a

[1]However, in recent Perkov [104, 105] has given a syntactic proof of Arrow's Theorem in a natural deduction calculus for the modal logic of Ågotnes et al. [1]. (See also Section 4.7.)

given logic by means of reference to existing "semantic" proofs of Arrow's Theorem. The ultimate goal of such research, however, must be the opposite: to use the logic to derive proofs for Arrow's Theorem and similar results.

In this chapter, we close this gap by providing a Hilbert-style syntactic proof of Arrow's Theorem within a simple tailor-made modal logic that is shown to be complete. We have opted for a Hilbert calculus, rather than, say, an approach based on natural deduction, because Hilbert calculi are still the systems used most widely by modal logicians and thus facilitate comparison to proof systems for other logics, and because this choice allows for a particularly compact presentation of our assumptions. Having said this, other proof systems have other advantages (e.g., in view of readability of proofs or implementability) and thus certainly also have a place in the study of social choice theory.

Our logic of choice is a fragment of the *modal logic of social choice functions* proposed by Troquard et al. [118]. Troquard et al. have used their (full) logic to reason about the strategy-proofness of voting rules (but it has not previously been applied to Arrow's Theorem). This logic can be used to model a (resolute) *social choice function* (SCF), i.e., a function that maps any given profile of preference orders to a single winning alternative. While Arrow originally formulated his theorem for social welfare functions, i.e., functions that map any given profile of preference orders to a single social preference order [11], we will instead work with a standard variant of the theorem for SCF's [116]. Arguably, SCF's (returning a top alternative rather than a full ranking of all alternatives) are relevant to a wider range of applications. In any case, known techniques to prove either version of the theorem are very similar [54, 116]. Thus, our work also suggests how one might construct a similar syntactic proof of Arrow's Theorem for social welfare functions, using, for instance, a logic such as that of Ågotnes et al. [1].

Besides encoding and proving Arrow's Theorem, we also cover two further seminal impossibility results from social choice theory, namely Sen's Theorem [110] on the impossibility of a Paretian liberal and the Muller-Satterthwaite Theorem [96], thereby demonstrating the generality and flexibility of our approach. Both of these theorems have so far received only very little attention in the literature on logics for social choice, with the notable exception of the work of Tang and Lin [115]. Sen's Theorem shows that the Pareto principle, by which unanimously held preferences should be respected, and a very weak form of liberalism, by which there should be certain private issues that only concern a single agent and that therefore should be dictated by that agent, are incompatible. The Muller-Satterthwaite Theorem shows that the only SCF's that satisfy a particular - strong but intuitively appealing - form of monotonicity are the dictatorships and those social choice functions that bar certain alternatives from winning, even if they are preferred by all agents.

Arguably, these are three of the four most important classical impossibility results in social choice theory. The fourth, the Gibbard-Satterhwaite Theorem [61, 109] on the impossibility of devising a strategy-proof SCF, is outside the scope of this chapter as it requires us to model both declared preferences (as for the three theorems covered here) and actual preferences, so as to be able to distinguish truthful agents from agents engaging in strategic manipulation. The modal logic of SCF's we are working with can only model one type of preference. This is intended and appropriate for our purposes. However, the full original logic of Troquard et al. [118] can model these two layers of preferences - indeed, this is the main objective it had been designed for originally. Our work, together with the fact that the Gibbard-Satterthwaite Theorem may be considered a relatively simple corollary to the Muller-Satterthwaite Theorem requiring only a proof showing that strategy-proofness implies strong monotonicity [54], therefore strongly suggests that proving the Gibbard-Satterthwaite Theorem in the full logic of Troquard et al. using an extension of our approach is possible in principle.

Our proofs are presented as human-readable recipes for how to construct a fully formal derivation inside the modal logic of SCF's of the three impossibility theorems discussed. These recipes can be transformed into machine-readable proofs relatively easily, and it is therefore possible in principle to have the proofs verified automatically by a proof-checker for this logic. In this sense, our contribution narrows the gap between, on the one hand, work on logics for modelling social choice [1, 64, 118] and, on the other, work on automated reasoning for social choice [37, 57, 97, 115, 122]. Having said this, there currently is no work on automated theorem proving for the modal logic we are working with, so while narrowed, the aforementioned gap has not yet been fully closed. As a further step in this direction, we also discuss how to translate from modal logic into propositional logic. While this does result in a blow-up of the size of the representation of theorems (meaning that we lose readability for humans) it makes it possible for us to use standard tools, particularly SAT solvers, to automatically reason about these theorems. This perspective provides a close connection to the approach pioneered by Tang and Lin [115], and later refined by others [37,57], of automatically proving results in social choice theory using SAT solvers.

4.3 A modal logic of social choice functions

In this section, we recall the formal definition of a SCF and introduce the fragment of the logic put forward by Troquard et al. [118] required to define such a SCF, adapting some of their notation and terminology to our purposes. We then demonstrate that the known completeness theorem for the full logic extends to the fragment that is of interest to us here. Finally, we discuss the limitations of this logic in view of expressing properties of families of SCF's ranging over electorates of varying size, as well as how to overcome these limitations in practice.

4.3.1 Social choice functions

Let $N = \{1, \ldots, n\}$ be a finite set of *agents* (or *individuals*) and let X be a finite set of *alternatives* (or *candidates*). To vote, each agent $i \in N$ expresses her preferences by supplying a linear order $\succcurlyeq_i$ over X, i.e., a binary relation that is reflexive, antisymmetric, complete, and transitive.[2] Let $\mathcal{L}(X)$ denote the set of all such linear orders. We shall also refer to $\succcurlyeq_i$ as the *ballot* provided by agent i, to stress the fact that this is the preference declared by the agent, but not necessarily her true preference. A *profile* is an n-tuple $(\succcurlyeq_1, \ldots, \succcurlyeq_n) \in \mathcal{L}(X)^n$ of such ballots, one for each agent.

4.3.1. Definition. A *resolute social choice function* is a function $F : \mathcal{L}(X)^n \to X$ mapping any given profile of ballots to a single winning alternative.

Examples for resolute SCF's are well-known voting rules, such as the Borda rule or the plurality rule [116] - when combined with a suitable tie-breaking rule that ensures that there always is just a single winner. Under the Borda rule, for instance, an agent assigns as many points to a given alternative as she lists other alternatives below it (with the alternatives obtaining the most points winning). Ties may be broken, for instance, by using the ballot of the first agent.

4.3.2 Language

Troquard et al. [118] have introduced a modal logic, which they call $\Lambda^{\text{scf}}[N, X]$, to reason about resolute SCF's (mapping declared preferences to winners) as well as the agents' truthful preferences. This logic can be used to model strategic behaviour in voting. Here we are not specifically interested in this strategic component, but rather in the purely aggregative aspect of social choice, i.e., in the question of whether a given SCF fairly aggregates individual ballots into a social decision. We shall refer to the relevant fragment of the logic of Troquard et al. as $L[N, X]$, the *logic of SCF's* parametrised by N and X. Next, we define the language, i.e., the set of well-formed formulas, of this logic.

This language is built on top of two types of atomic propositions. First, for every $i \in N$ and $x, y \in X$, $p^i_{x \succcurlyeq y}$ is an atomic proposition (with the intuitive meaning that agent i prefers x to y). $\mathit{Pref}[N, X] := \{p^i_{x \succcurlyeq y} \mid i \in N \text{ and } x, y \in X\}$ is the set of all such propositions. Second, by a slight abuse of notation, every alternative $x \in X$ is also an atomic proposition (with the intuitive meaning that x wins). Besides the usual propositional connectives, we have a modal operator $\Diamond_C$ for every coalition of agents $C \subseteq N$ (with the intuitive meaning that C can

[2] The strict part $\succ_i$ of $\succcurlyeq_i$ is a strict linear order, a relation that is irreflexive, complete, and transitive. While most work in voting theory tends to take such strict linear orders as primitive, we instead follow Troquard et al. [118] and work with non-strict linear orders. Ultimately, both approaches are equivalent: $\succcurlyeq_i$ uniquely determines $\succ_i$, and vice versa.

ensure the truth of a given formula, provided the others do not alter their ballots). The following definition summarises how the language is constructed.

4.3.2. Definition. The set of well-formed formulas φ in the language of $L[N, X]$ is generated by the following Backus-Naur Form :

$$\varphi \quad ::= \quad p \mid x \mid \top \mid \neg\varphi \mid \varphi \vee \varphi \mid \Diamond_C \varphi$$

where $p \in \mathit{Pref}[N, X]$, $x \in X$ and $C \subseteq N$.

Additional propositional connectives and a dual modal operator are defined in the usual manner: $\varphi \wedge \psi$ is short for $\neg(\neg\varphi \vee \neg\psi)$, $\varphi \to \psi$ is short for $\neg\varphi \vee \psi$, $\varphi \leftrightarrow \psi$ is short for $(\varphi \to \psi) \wedge (\psi \to \varphi)$, $\bot$ is short for $\neg\top$, and $\Box_C\varphi$ is short for $\neg\Diamond_C\neg\varphi$. For $i \in N$, we write $\Diamond_i$ as a shorthand for $\Diamond_{\{i\}}$ and $\Box_i$ as a shorthand for $\Box_{\{i\}}$.

The full logic of Troquard et al. [118] includes an additional pair of modal operators to speak about true preferences.

4.3.3 Semantics

The semantics of the logic is a standard possible-worlds semantics for modal logics, defined in terms of a set of possible worlds, a family of accessibility relations, and a valuation function [32]. We first give a short high-level description intended for readers familiar with such semantics, and then provide complete formal definitions.

First, the set of possible worlds is the set of all possible profiles - which is fully determined by N and X. The semantics of atomic propositions of the form $p^i_{x \succcurlyeq y}$ will be defined solely in terms of this set of possible worlds: $p^i_{x \succcurlyeq y}$ is true at a given world/profile w, if agent i prefers x to y in w. Only to model the truth of atomic propositions of the form x will we require a valuation function. Valuation functions here are SCF's: x is true at world/profile w if the SCF in question maps profile w to the winning alternative x. Finally, for every coalition $C \subseteq N$, there is an accessibility relation between worlds/profiles: w is connected to w' if they differ only w.r.t. the preferences of agents in C. These accessibility relations will be used to define the semantics of modal formulas of the form $\Diamond_C\varphi$ in the usual manner.

4.3.3. Definition. A model is a triple $M = \langle N, X, F \rangle$, consisting of a finite set of agents N with $n = |N|$, a finite set of alternatives X, and a SCF $F : \mathcal{L}(X)^n \to X$.

For fixed sets N and X, we sometimes write M_F for the model $M = \langle N, X, F \rangle$ based on the SCF F. From now on we shall use the terms *'world'* and *'profile'* interchangeably. We are now ready to define what it means for a formula φ to be true at a world $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$ in a given model M.

4.3.4. Definition. Let $M = \langle N, X, F \rangle$ be a model. We write $M, w \models \varphi$ to express that the formula φ is true at the world $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n) \in \mathcal{L}(X)^n$ in M. The satisfaction relation $\models$ is defined inductively:

- $M, w \models p^i_{x \succcurlyeq y}$ iff $x \succcurlyeq_i y$
- $M, w \models x$ iff $F(\succcurlyeq_1, \ldots, \succcurlyeq_n) = x$
- $M, w \models \neg\varphi$ if $M, w \not\models \varphi$
- $M, w \models \varphi \vee \psi$ iff $M, w \models \varphi$ or $M, w \models \psi$
- $M, w \models \Diamond_C \varphi$ iff $M, w' \models \varphi$ for some world $w' = (\succcurlyeq'_1, \ldots, \succcurlyeq'_n) \in \mathcal{L}(X)^n$ with $\succcurlyeq_i = \succcurlyeq'_i$ for all agents $i \in N \setminus C$.

That is, $\Diamond_C \varphi$ is true at w, if the agents in C can make φ true by changing their own ballots, assuming none of the other agents change as well. Thus, $\Box_C \varphi$ is true at w if φ holds at every world that is reachable from w by only the agents in C changing their ballots. Notice that the semantics of this operator can be easily seen as a standard relational semantics for a relation R_C defined as $(w, w') \in R_C$ iff the ballots of the agents $i \in N \setminus C$ coincide in w and w'. From this point of view what we are doing is essentially looking at one fixed frame, where the worlds are the profiles and the relations are themselves derived by the profiles: the different models on this one frame are given by different SCF's, qua valuation.

In some sense, the truth of every formula of the form $p^i_{x \succcurlyeq y}$ is under the control of agent i. Because of this feature, this kind of logic is sometimes classified as a *logic of propositional control*. The motivation underlying such logics is essentially game-theoretic: every individual is conceived as having "control" over a set of atomic propositions. The choice of a particular truth value for these atomic propositions can be seen as an action of the individual, and therefore a valuation of all the atomic propositions of this sort corresponds to a strategy profile. For more details and motivations on logics of propositional control we refer to the work of van der Hoek and Wooldridge [120], Gerbrandy [58], Balbiani et al. [14] and Troquard et al. [118], amongst others. We also note that these logics are closely related to Pauly's coalition logic [101], Boolean games [35, 68], and the Ceteris Paribus Logic of Grossi et al. [65].

Let φ be a formula in the language based on N and X. Then φ is called *satisfiable*, if there exist a SCF F and a world $w \in \mathcal{L}(X)^n$ such that $M_F, w \models \varphi$. It is called *true in the model* M, denoted $M \models \varphi$, if $M, w \models \varphi$ for every world $w \in \mathcal{L}(X)^n$. Finally, it is called *valid*, denoted $\models \varphi$, if $M \models \varphi$ for every model M based on N and X.

The logic of Troquard et al. [118] is known to be decidable and this result immediately extends to the fragment of their logic discussed here:

4.3.5. Proposition. *Determining whether a formula in the language of $L[N, X]$ is valid is a decidable problem.*

Proof:
Since N and X are fixed, we can enumerate all models and check for each of them whether our formula is true at every world in the model. □

4.3.4 Axiomatisation and completeness

Next, we review the axiomatisation due to Troquard et al. [118], restricted to the fragment $L[N,X]$ discussed here; we then adapt their completeness result to this fragment. The first few axioms ensure that the propositions of the form $p^i_{x \succcurlyeq y}$ really encode linear orders.

(1) $p^i_{x \succcurlyeq x}$ (reflexivity)

(2) $p^i_{x \succcurlyeq y} \leftrightarrow \neg p^i_{y \succcurlyeq x}$ for $x \neq y$ (antisymmetry and completeness)

(3) $p^i_{x \succcurlyeq y} \wedge p^i_{y \succcurlyeq z} \rightarrow p^i_{x \succcurlyeq z}$ (transitivity)

Here x, y and z range over atomic propositions in X, and i ranges over agents. Before we continue with the axiomatisation, let us first introduce a couple of additional language constructs to refer to ballots and profiles within the logical language. Consider a profile $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n) \in \mathcal{L}(X)^n$. For a given agent $i \in N$, let $x_1, x_2, \ldots, x_m$ be a permutation of the elements of X such that $x_1 \succcurlyeq_i x_2 \succcurlyeq_i \cdots \succcurlyeq_i x_m$. Then $ballot_i(w)$ is defined as the following formula:

$$ballot_i(w) \;\; := \;\; p^i_{x_1 \succcurlyeq x_2} \wedge p^i_{x_2 \succcurlyeq x_3} \wedge \cdots \wedge p^i_{x_{m-1} \succcurlyeq x_m}$$

Thus, $ballot_i(w)$ is true at world w' if and only if w and w' agree as far as the ballot of agent i is concerned. Note that $ballot_i(w)$ is a purely syntactic representation of a semantic notion (namely, agent i's preference order $\succcurlyeq_i$). Similarly, we define *profile*(w) as the following formula:

$$profile(w) \;\; := \;\; ballot_1(w) \wedge ballot_2(w) \wedge \cdots \wedge ballot_n(w)$$

Hence, the formula *profile*(w) is true at world w, and only there. This shows that *nominals*, i.e., formulas uniquely identifying worlds [32], are definable within this logic. Furthermore, due to the finiteness of X and N, there can be only finitely many formulas of type *profile*(w) that are consistent with the axioms.

Let $N^w_{x \succcurlyeq y} := \{i \in N \mid x \succcurlyeq_i y\}$ denote the set of agents that prefer x over y in profile $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$. By a slight abuse of notation, we use the same expression as a construct of our language:

$$N^w_{x \succcurlyeq y} \;\; := \;\; \bigwedge \{p^i_{x \succcurlyeq y} \mid x \succcurlyeq_i y \text{ in } w\}$$

We write $N^w_{x \succcurlyeq y}$ to denote both the set of agents and the formula; the context will disambiguate the intended meaning. Note that $\bigwedge_{x,y \in X} N^w_{x \succcurlyeq y}$ is logically equivalent to *profile*(w): this reflects the fact that a profile can either be presented by specifying the preferences of each individual or by specifying the sets of agents preferring one alternative over another, for all pairs of alternatives.

For any two alternatives $x, y \in X$, we define *profile*$(w)(x,y)$ as the formula fixing the relative ordering of x and y for all agents as in profile w:

$$\mathit{profile}(w)(x,y) \;\; := \;\; N^w_{x \succcurlyeq y} \wedge N^w_{y \succcurlyeq x}$$

This formula will be used to express the fact that two profiles 'agree' on the preferences concerning the alternatives x and y.

We now state the remaining axioms defining the logic $L[N, X]$:

(4) all propositional tautologies

(5) $\Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$ (K(i))

(6) $\Box_i\varphi \to \varphi$ (T(i))

(7) $\varphi \to \Box_i\Diamond_i\varphi$ (B(i))

(8) $\Diamond_i\Box_j\varphi \leftrightarrow \Box_j\Diamond_i\varphi$ (confluence)

(9) $\Box_{C_1}\Box_{C_2}\varphi \leftrightarrow \Box_{C_1 \cup C_2}\varphi$ (union)

(10) $\Box_\emptyset\varphi \leftrightarrow \varphi$ (empty coalition)

(11) $(\Diamond_i p \wedge \Diamond_i \neg p) \to (\Box_j p \vee \Box_j \neg p)$, where $i \neq j$ (exclusiveness)

(12) $\Diamond_i \mathit{ballot}_i(w)$ (ballot)

(13) $\Diamond_{C_1}\delta_1 \wedge \Diamond_{C_2}\delta_2 \to \Diamond_{C_1 \cup C_2}(\delta_1 \wedge \delta_2)$ (cooperation)

(14) $\bigvee_{x \in X}(x \wedge \bigwedge_{y \in X \setminus \{x\}} \neg y)$ (resoluteness)

(15) $(\mathit{profile}(w) \wedge \varphi) \to \Box_N(\mathit{profile}(w) \to \varphi)$ (functionality)

Here φ and ψ range over arbitrary formulas, x over atomic propositions in X, i and j over agents, C_1 and C_2 over coalitions, and w over profiles. In axiom (11), p is ranging only over atomic propositions in the set $\mathit{Pref}[N, X]$, and in axiom (13) δ_1 and δ_2 do not contain any common atoms.

Axioms (4)–(8) describe well-known properties of normal modal logics [32]. Axiom (9) describes the capability of a coalition to enforce a certain formula in terms of the capabilities of its sub-coalitions. Axiom (10) states that the empty coalition cannot enforce any formula. Axiom (11) enforces a division among the atomic propositions of the shape $p^i_{x \succcurlyeq y}$: if an atom is controlled by an agent i, then other agents cannot change its value. Axiom (12) ensures that every agent can express every possible preference. Due to axiom (13), if two formulas δ_1 and δ_2 do not contain a common atom and two coalitions C_1 and C_2 can each enforce one of the formulas, then the joint coalition can enforce the conjunction $\delta_1 \wedge \delta_1$. Axiom (14) expresses that any outcome associated with a profile must be a single winning alternative. Thus, this axioms encodes the resoluteness of the SCF in question. Finally, axiom (15) ensures that every profile is associated with a single outcome, i.e., it encodes the fact that the SCF being modelled must be a function.

The inference rules of the logic are *modus ponens* and *necessitation* w.r.t. all modalities of the form $\Box_i$ [32]:

- (MP) from $\varphi \to \psi$ and φ, infer ψ
- (Nec$_i$) from $\vdash \varphi$, infer $\vdash \Box_i\varphi$

Here we write $\vdash \varphi$ to express that a well-formed formula φ in the language parametrised by N and X is a *theorem* of the logic $L[N, X]$, in the sense that it can be derived from axioms (1)–(15), together with the above inference rules. The $\vdash \varphi$ appearing in the second rule thus indicates that the rule can only be applied to theorems. We define a set of formulas Γ to be *consistent* if we cannot derive a contradiction from it. The theorems of $L[N, X]$ coincide with the valid formulas:

4.3.6. Theorem (Completeness). *The logic $L[N, X]$ is sound and complete w.r.t. the class of models of SCF's.*

Proof:
Since our logic is a fragment of $\Lambda^{\text{scf}}[N, X]$, the soundness result due to Troquard et al. [118] applies directly. The same is not true for completeness. However, as we shall outline next, the proof of Troquard et al. [118] for the richer logic can be adapted to our fragment, *mutatis mutandis*.

The strategy of the proof is a canonical model construction, with a little variation over the standard proof. As we mentioned, the models of Definition 4.3.3 can be repackaged as particular Kripke models. The latter structures are tuples $\langle W, (R_C)_{C\subseteq N}\rangle$ where W is the set of profiles and $R_C \subseteq W \times W$ are relations defined as

$$wR_Cw' \quad \text{iff} \quad w \restriction N \setminus C = w' \restriction N \setminus C,$$

where $w \restriction N \setminus C$ is the profile w restricted to only the individuals outside of C. Intuitively, wR_Cw' holds if all the agents in $N \setminus C$ express the same preferences in w and w'. We proceed with a canonical model argument to establish completeness.

Given a consistent formula φ, we build a maximally consistent set Γ_φ containing it using the usual Lindenbaum construction. Nevertheless, the set of all MCS is not by itself a model of $L[N, X]$: we have different MCSs containing the same formula *profile*(w) but containing different alternatives, so we cannot associate a SCF to this set of states. However, Γ_φ does contain a full specification of a SCF, in the shape of formulas $\Diamond_N(\textit{profile}(w) \wedge x)$, so we can select the MCSs that 'agree' with Γ_φ with respect to the specification of the SCF.

Define $Cluster(\Gamma_\varphi)$ to be the set of maximally consistent sets that describe the same SCF:

$$\begin{aligned} Cluster(\Gamma_\varphi) \quad := \quad & \{\Gamma \mid \forall w \in \mathcal{L}(X)^n, \forall x \in X : \\ & \Diamond_N(\mathit{profile}(w) \wedge x) \in \Gamma \quad \mathit{iff} \\ & \Diamond_N(\mathit{profile}(w) \wedge x) \in \Gamma_\varphi\} \end{aligned}$$

We then consider the submodel of the canonical model generated by $Cluster(\Gamma_\varphi)$. Let us call this submodel M_φ. It is then easy to check that:

- the Truth Lemma holds for M_φ;
- there is a bijection between profiles and states of M_φ;
- M_φ is one of the aforementioned particular Kripke models corresponding to the models of our logic.

The first item is shown in the customary way. One direction of second item holds because, due to the axioms, each MCS contains only one formula of shape *profile*(w). Moreover, for any profile w', the set

$$\{\mathit{profile}(w')\} \cup \{\Diamond_N(\mathit{profile}(w) \wedge x) | \Diamond_N(\mathit{profile}(w) \wedge x) \in \Gamma_\varphi, x \in X, w \in\}$$

is consistent and can be extended to a MCS, therefore $Cluster(\Gamma_\varphi)$ contains exactly one MCS for each profile. In light of this last observation the third item follows straightforwardly. □

4.3.5 Representing families of social choice functions

To complete the outline of the expressive capabilities of $L[N, X]$, we illustrate how it is possible to encode a SCF as a formula. Given a SCF F, its representation will be:

$$\rho^F \quad = \quad \bigwedge \{\mathit{profile}(w) \to x \mid w \in \mathcal{L}(X)^n \text{ and } F(w) = x\}$$

That is, ρ^F is simply the conjunction, over all profiles w, of implications between a formula describing w and a formula identifying the winning alternative for profile w under F. In other words, we need to have the full graph of the function, that is, the full set of input-output pairs, to be able to encode F in the language. This is indeed possible, because, strictly speaking, ρ^F represents the function only for a fixed number of alternatives and a fixed number of agents. Moreover, since we are able to encode any set of input-output pairs, we can represent any SCF in the language.

Unfortunately, for the very same reason, ρ^F cannot be taken as a proper representative of a SCF, because it only tells us what the output of the function is in a very limited case: when the alternatives are exactly those in X and when the agents are exactly those in N. In practice, however, we are interested in *families* of SCF's. If, say, F is the Borda rule and X and N both have cardinality 3, then ρ^F will only express the workings of the Borda rule for 3 alternatives and 3 agents. A full representation of the Borda rule (which formally is a family of SCF's in the sense of Definition 4.3.1), however, should contain the information necessary to compute the output from *any* given profile. It should be a conjunction of all the formulas ρ^F for all possible choices of X and N. But even assuming that we had all such sets of pairs, there are countably many ρ's of this kind, and our logical language does not contain countable conjunctions. Given that the language is not powerful enough to encode an algorithmic specification, there is no hope that our logic, or a similar logic, will do better than using ρ^F in representing SCF's. Indeed, this restriction to specific sets of alternatives and agents is a recognised limitation of most existing logic-based approaches to modelling frameworks of social choice [54].

Interestingly, however, this problem affects the representations of the properties of SCF's only partially. Since most of the properties do not directly refer to the specific number of alternatives and agents, we can formulate the properties leaving X and N as parameters. The same can be done when proving the relative dependencies between properties. This means that, to prove that property P_1 entails P_2, we prove that, for fixed choices of X and N, there is a proof in the logic from the formula encoding P_1 to the formula encoding P_2 (both these formulas are instantiated to X and N themselves). This is the approach we shall take here.

4.4 Modelling features of Social Choice Theory

In this section, we show how to model several important concepts of social choice theory in our logic. We start by proving the Universal Domain Lemma, which demonstrates that there exists a formula in our language that expresses that for every possible preference profile there exists a world where it is realised, and that is a theorem of our logic. This simple but important result will be used throughout the chapter. We then pause to introduce and encode a notion that will feature in several properties and proofs in the next (sub)sections, the concept of a decisive coalition. Finally, we formalise the main properties featuring in the classical impossibility theorems we want to prove, particularly Pareto efficiency, independence of irrelevant alternatives, strong monotonicity, and liberalism. For each property we suggest an encoding in the logic and prove that it indeed captures the corresponding semantic notion.

Throughout, we exploit freely the finiteness of the language, using big conjunctions and disjunctions to quantify over individuals, alternatives, and profiles.

4.4.1 The Universal Domain Lemma

The following lemma states that all the possible profiles are also possible worlds in the semantics. This fact, which is implicit in our definition of a SCF, is called the *universal domain* condition in Arrow's original work [11].

4.4.1. LEMMA (UNIVERSAL DOMAIN LEMMA). *For every possible profile $w \in \mathcal{L}(X)^n$, we have that $\vdash \Diamond_N profile(w)$.*

Proof:
Take any profile w. Then $ballot_1(w)$ encodes the preferences of the first agent. We have, by axiom (12), that $\Diamond_1 ballot_1(w)$, and similarly for the second agent we get $\Diamond_2 ballot_2(w)$. Because $ballot_1(w)$ and $ballot_2(w)$ contain different atoms (the former only atoms with superscript 1, the latter only atoms with superscript 2), we can apply axiom (13) and obtain $\Diamond_{\{1,2\}}(ballot_1(w) \wedge ballot_2(w))$. We can repeat this reasoning for all the finitely many agents in N to prove $\Diamond_N profile(w)$. □
Even though a theorem of this shape is somewhat surprising from a Modal logic point of view, the reader should recall that on the semantic side we are dealing with one fixed frame, whose worlds are the profiles. The Universal Domain Lemma therefore encodes in the syntax the fact that all profiles are available as worlds.

4.4.2 Decisive coalitions

We will call a coalition of agents $C \subseteq N$ *decisive* over a pair of alternatives $(x, y) \in X^2$ if the members of C preferring x to y is a sufficient condition for preventing y from winning. We use the following formula to encode decisiveness of C over (x, y):

$$C dec(x, y) \quad := \quad \left(\bigwedge_{i \in C} p^i_{x \succcurlyeq y} \right) \rightarrow \neg y$$

If C is decisive on every pair, we will simply write $C dec$. Along the same lines, we define a *weakly decisive* coalition C for (x, y) as a coalition that can bar y from winning if *exactly* the agents in C prefer x to y. We encode weak decisiveness of C over (x, y) as follows:

$$C wdec(x, y) \quad := \quad \left(\bigwedge_{i \in C} p^i_{x \succcurlyeq y} \wedge \bigwedge_{i \notin C} p^i_{y \succcurlyeq x} \right) \rightarrow \neg y$$

The reader can easily check that these syntactic notions match the semantic ones; for example, in the case of decisiveness we have that $C dec(x, y)$ is true in the model M_F if and only if the coalition C is decisive over that pair of alternatives for the corresponding SCF F.

4.4.3 Pareto efficiency

We introduce several properties that one might reasonably want to require a SCF to satisfy. The first is *Pareto efficiency*, expressing the desideratum that, if all the agents rank an alternative x above another alternative y, then y should not win.

4.4.2. Definition. *A SCF F is Pareto efficient if, for every profile $w \in \mathcal{L}(X)^n$ and every pair of distinct alternatives $x, y \in X$ with $N^w_{x \succcurlyeq y} = N$, we obtain $F(w) \neq y$.*

This is formalised as follows:

$$Par \ := \ \bigwedge_{x \in X} \bigwedge_{y \in X \setminus \{x\}} \left[\left(\bigwedge_{i \in N} p^i_{x \succcurlyeq y} \right) \to \neg y \right]$$

Observe that Par is equivalent to $N dec$, i.e., to saying that the grand coalition N is decisive on every pair.

4.4.3. Lemma. *For every SCF F, $M_F \models Par$ if and only if F is Pareto efficient.*

Proof:
Straightforward. □

4.4.4 Independence of irrelevant alternatives

Our next property of interest is *independence of irrelevant alternatives* (IIA). It expresses the intuitively desirable property of a SCF F that, for every two profiles and for every two alternatives x and y, if the outcome of F in the first profile is x and the two profiles are identical as far as the preferences of the agents over x and y are concerned, then the outcome of F in the second profile should not be y. The original formulation of IIA given by Arrow [11] was applied to social welfare functions rather than SCF's. Our definition is the most natural adaptation of Arrow's idea to SCF's. It has also been used by Taylor [116], amongst others.

4.4.4. Definition. *A SCF F satisfies IIA if, for every pair of profiles $w, w' \in \mathcal{L}(X)^n$ and every pair of distinct alternatives $x, y \in X$ with $N^w_{x \succcurlyeq y} = N^{w'}_{x \succcurlyeq y}$, it is the case that $F(w) = x$ implies $F(w') \neq y$.*

We formalise this property in our logic as follows:

$$IIA \ := \ \bigwedge_{w \in \mathcal{L}(X)^n} \bigwedge_{x \in X} \bigwedge_{y \in X \setminus \{x\}} [\Diamond_N(\mathit{profile}(w) \wedge x) \to (\mathit{profile}(w)(x,y) \to \neg y)]$$

That is, if in some world (reachable via the $\Diamond_N$-modality) we observe profile w with alternative x winning, then in the present world, if it agrees with w as far as the relative ranking of x and y is concerned, y cannot be the winner.

The following lemma formally establishes the correspondence between the syntactic formulation of IIA and its semantic definition.

4.4.5. LEMMA. *For every SCF F, $M_F \vDash IIA$ if and only if F satisfies the property of independence of irrelevant alternatives.*

Proof:
From right to left, assume F satisfies IIA. We want to prove every conjunct of the formula *IIA*. So let w' be a world such that $M_F, w' \vDash \Diamond_N(\mathit{profile}(w) \wedge x)$. We want to show that $M_F, w' \vDash (\mathit{profile}(w)(x,y) \rightarrow \neg y)$. So suppose $M_F, w' \vDash \mathit{profile}(w)(x,y)$, which entails $N^w_{x \succcurlyeq y} = N^{w'}_{x \succcurlyeq y}$. By the semantics of $\Diamond_N$, there is a world w'' such that $M_F, w'' \vDash \mathit{profile}(w) \wedge x$, which entails $N^w_{x \succcurlyeq y} = N^{w''}_{x \succcurlyeq y}$. Thus, also $N^{w'}_{x \succcurlyeq y} = N^{w''}_{x \succcurlyeq y}$. From $M_F, w'' \vDash x$ we can infer $F(w'') = x$. Now we can apply IIA to w'' and w' and obtain $F(w') = x$ and thus $F(w') \neq y$. Again by the semantics, this is tantamount to $M_F, w' \vDash \neg y$.

From left to right, assume $M_F \vDash IIA$. Consider two profiles w, w' and two alternatives x, y with $N^w_{x \succcurlyeq y} = N^{w'}_{x \succcurlyeq y}$. Now assume $F(w) = x$. We thus have $M_F, w \vDash \mathit{profile}(w) \wedge x$ and, by the semantics of $\Diamond_N$, also $M_F, w' \vDash \Diamond_N(\mathit{profile}(w) \wedge x)$. By modus ponens and *IIA* we get $M_F, w' \vDash (\mathit{profile}(w)(x,y) \rightarrow \neg y)$. But we assumed $N^w_{x \succcurlyeq y} = N^{w'}_{x \succcurlyeq y}$, hence $M_F, w' \vDash \mathit{profile}(w)(x,y)$ and thus $M_F, w' \vDash \neg y$, which by the semantics entails $F(w') \neq y$. □

4.4.5 Strong monotonicity

Next is a monotonicity property known as *Maskin monotonicity* or *strong monotonicity*. It requires that, whenever alternative x wins in a given profile and we (weakly) improve the standing of x *vis-à-vis* all other alternatives, then x should still win in the new profile - even if the relative rankings of other alternatives change in the profile as well. While its formal definition is similar to that of IIA, there are subtle differences: we are now quantifying over all other alternatives y rather than considering one specific such alternative.

4.4.6. DEFINITION. *A SCF F is strongly monotonic if, for every pair of profiles $w, w' \in \mathcal{L}(X)^n$ and every alternative $x \in X$, it is the case that $F(w) = x$ and $N^w_{x \succcurlyeq y} \subseteq N^{w'}_{x \succcurlyeq y}$ for all $y \in X \setminus \{x\}$ together imply $F(w') = x$.*

This property can be encoded as follows:

$$SM \;:=\; \bigwedge_{w \in \mathcal{L}(X)^n} \bigwedge_{x \in X} \left[\Diamond_N(\mathit{profile}(w) \wedge x) \;\wedge\; \left(\bigwedge_{y \in X \setminus \{x\}} N^w_{x \succcurlyeq y} \right) \;\rightarrow\; x \right]$$

4.4.7. LEMMA. *For every SCF F, $M_F \vDash SM$ if and only if F is strongly monotonic.*

Proof:
From left to right, suppose $M_F \vDash SM$ is the case. Let w and w' be two profiles, assume that $F(w) = x$ and $N^w_{x \succcurlyeq y} \subseteq N^{w'}_{x \succcurlyeq y}$ for all $y \in X \setminus \{x\}$. Due to $F(w) = x$, we have $M_F, w \vDash \mathit{profile}(w) \wedge x$ and, by the semantics of $\Diamond_N$, also $M_F, w' \vDash \Diamond_N(\mathit{profile}(w) \wedge x)$. By the second assumption, namely $N^w_{x \succcurlyeq y} \subseteq N^{w'}_{x \succcurlyeq y}$, we obtain that the second conjunct of *SM*, namely $\bigwedge_{y \in X \setminus \{x\}} N^w_{x \succcurlyeq y}$, is also true at w'. From the validity of *SM* we can conclude $M_F, w' \vDash x$ and hence $F(w') = x$.

From right to left, suppose F is strongly monotonic. Let w, w' be profiles and x an alternative. Finally, assume $M_F, w' \vDash \Diamond_N[x \wedge \mathit{profile}(w)] \wedge \bigwedge_{y \in X \setminus \{x\}} N^w_{x \succcurlyeq y}$. Due to the first conjunct we know that $F(w) = x$, while in light of the second we can conclude that $N^w_{x \succcurlyeq y} \subseteq N^{w'}_{x \succcurlyeq y}$ for all $y \in X \setminus \{x\}$, because by the semantics all the supporters of x over y in w still support x over y in w'. By strong monotonicity we get $F(w') = x$ and $M_F, w' \vDash x$. Since w' was generic we can conclude that *SM* is a validity in M_F. □

4.4.6 Surjectivity

The most basic property we consider is *surjectivity*. It expresses the desideratum that every alternative should be the winner for at least one profile.

4.4.8. Definition. *A SCF F is surjective if, for every alternative $x \in X$ there exists a profile $w \in \mathcal{L}(X)^n$ such that $F(w) = x$.*

We can encode surjectivity as follows:

$$Sur \;:=\; \bigwedge_{x \in X} \bigvee_{w \in \mathcal{L}(X)^n} \Diamond_N(\mathit{profile}(w) \wedge x)$$

4.4.9. Lemma. *For every SCF F, $M_F \vDash Sur$ if and only if F is surjective.*

Proof:
Straightforward. □

4.4.7 Liberalism

The idea that a form of *liberalism* can be modelled as a property of SCF's is due to Sen [110]. He postulated that every agent should have the power to determine the relative ranking of at least two alternatives x and y. For example, x might be the state of the world in which Barack Obama is president of the United States of America and you paint the walls of your bedroom in pink, and y might be the state of the world where Barack Obama is president of the United States of America and you paint the walls of your bedroom in white. Then you should have the power of excluding one of x and y from being the collectively chosen alternative (which of course does not mean that the other one of the two necessarily needs to be chosen). In this case, we say that you are (two-way) decisive on x and y.

4.4.10. Definition. *A SCF F satisfies the property of liberalism if, for every individual $i \in N$ there exist two distinct alternatives $x, y \in X$ for which i is two-ways decisive.*

The property of liberalism can be encoded as follows:

$$Lib \;:=\; \bigwedge_{i \in N} \bigvee_{x \in X} \bigvee_{y \in X \setminus \{x\}} (\{i\}\mathit{dec}(x,y) \wedge \{i\}\mathit{dec}(y,x))$$

4.4.11. Lemma. *For every SCF F, $M_F \vDash Lib$ if and only if F satisfies liberalism.*

Proof:
From left to right, suppose $M_F \vDash Lib$. Suppose for the sake of contradiction that F does not satisfy liberalism. If there is an individual i that is not two-ways decisive on any pairs then for every pair there is a profile w such that the outcome $F(w)$ is in conflict with the preferences of i (say, $x \succcurlyeq_i y$ and $F(w) = y$). This means that $\{i\}\mathit{dec}(x,y) \wedge \{i\}\mathit{dec}(y,x)$ cannot be a validity in the model M_F, and the same holds for all the pairs, so $\bigvee_x \bigvee_{y \neq x}(\{i\}\mathit{dec}(x,y) \wedge \{i\}\mathit{dec}(y,x))$ cannot be a validity either, for our fixed i. This in turn entails that $M_F \vDash Lib$ is not the case, contradiction.

From right to left, say that F satisfies liberalism. For an agent i, it is easy to check that, calling x, y the alternatives for which i is decisive, we must have $\{i\}\mathit{dec}(x,y) \wedge \{i\}\mathit{dec}(y,x)$ as a validity on the model M_F. Thus, also $\bigvee_x \bigvee_{y \neq x}(\{i\}\mathit{dec}(x,y) \wedge \{i\}\mathit{dec}(y,x))$ is a validity, and the same holds for every i, so we get the validity of Lib. □

4.4.8 Dictatorships

Finally, we will require one *un*desirable property of SCF's. A *dictatorship* is a SCF for which one individual, the dictator, can enforce their top alternative as the outcome. Denote with top_i^w that alternative $x \in X$ for which $x \succcurlyeq_i y$ for all other alternatives $y \in X$ in profile $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$.

4.4.12. Definition. *A SCF F is a dictatorship if there exists an agent $i \in N$ (the dictator) such that, for every profile $w \in \mathcal{L}(X)^n$, we obtain $F(w) = top_i^w$.*

The property of being a dictatorship is encoded by the following formula:

$$Dic \;:=\; \bigvee_{i \in N} \bigwedge_{x \in X} \bigwedge_{y \in X \setminus \{x\}} (p^i_{x \succcurlyeq y} \to \neg y)$$

Observe that Dic is equivalent to $\bigvee_{i \in N}\{i\}\mathit{dec}$, i.e., a SCF is dictatorial if and only if there exists an individual that is decisive on every pair.

4.4.13. Lemma. *For every SCF F, $M_F \vDash Dic$ if and only if F is a dictatorship.*

Proof:
From right to left, suppose F is a dictatorship, and call the dictator i. Let $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$ be a profile. We want to show that the disjunct corresponding to i is true at w. Thus, for any two distinct alternative x, y we want to show that $p^i_{x \succcurlyeq y} \to \neg y$ is true at w. First, if $x \succcurlyeq_i y$, then $top^w_i \neq y$ and thus, due to F being a dictatorship of i, we have $F(w) \neq y$. By the semantics, this entails $M_F, w \vDash \neg y$ and thus $M_F, w \vDash p^i_{x \succcurlyeq y} \to \neg y$. Second, if $x \not\succcurlyeq_i y$, then $M_F, w \nvDash p^i_{x \succcurlyeq y}$, and the implication holds vacuously.

From left to right, suppose $M_F \vDash Dic$. Then one disjunct must be valid, say for agent i. Let $x = top^w_i$ under profile w. Then $M_F, w \vDash \bigwedge_{y \in X \setminus \{x\}} p^i_{x \succcurlyeq y}$. Since (the disjunct referring to i in) the condition Dic is true at w, we obtain $M_F, w \vDash \bigwedge_{y \in X \setminus \{x\}} \neg y$. By resoluteness we get $M_F, w \vDash x$ and thus $F(w) = x$. □

Note that, in the presence of axiom (14), encoding resoluteness, the disjunction in the formula Dic is actually an exclusive one, i.e., not only must there be some dictator, but there must be exactly one dictator.[3]

4.5 Impossibility theorems

We are now ready to state the three major impossibility theorems we are interested in as formulas in the language of our modal logic of SCF's. For each of them, we then demonstrate how to construct a full proof of the theorem within the axiomatic system we have seen to be complete for our logic (cf. Theorem 4.3.6). We start with Arrow's Theorem and then prove the Muller-Satterthwaite Theorem as a corollary. The third theorem, Sen's Theorem on the impossibility of a Paretian liberal, is mathematically much simpler and also admits a relatively short proof in our logic.

Before we begin, we need to make one important remark concerning the expressivity of our logic. Given that the language of $L[N, X]$ is parametrised by the set of individuals N and the set of alternatives X, strictly speaking the aforementioned theorems, which all apply to scenarios with arbitrary numbers of individuals and alternatives (provided those numbers are sufficiently large), cannot be stated or proven within the logic. To prove each of these impossibility theorems in their full generality we have to resort to a meta-argument, using a proof schema, to show that, for each choice of N and X, it is possible to prove a version of the theorem in the logic instantiated to those two parameters. The same *proviso* also holds for the properties of SCF's featuring in the previous section:

[3]The reader can prove this using the Universal Domain Lemma, formula *Dic*, and axiom (14). The gist of the proof is to take a profile where two dictators disagree and to show that this leads to a contradiction.

rather than being formulas in the logic, they are schemas of the representations of the properties in the logic.

4.5.1 Encoding Arrow's Theorem

First published in 1951, Arrow's Theorem is widely regarded as *the* seminal contributions to social choice theory [11]. The original theorem concerns *social welfare functions*, i.e., functions mapping profiles of (weak) preference orders (permitting indifference between alternatives) to single collective preference orders. The version we present here is adapted for preference orders that do not permit indifferences between alternatives and to SCF's (which return a single winning alternative rather than a collective order). We refer to Taylor [116] for an extensive discussion of this variant of the theorem. From a mathematical point of view, both variants are essentially equivalent and can be proven using the same methods [54, 116]. We focus on linear orders (not permitting indifferences), because most standard voting rules impose this requirement on ballots [116]. We furthermore focus on SCF's, because the problem of choosing a single best alternative is more pervasive in applications than that of choosing a full ranking over alternatives.

Arrow showed that, rather surprisingly, any SCF for three or more alternatives that is Pareto efficient and that satisfies the property of independence of irrelevant alternatives must be dictatorial.

4.5.1. Theorem (Arrow's Theorem). *Any SCF for at least three alternatives that satisfies IIA and the Pareto condition is a dictatorship.*

We now proceed to code a proof of Arrow's Theorem in our logic. We will follow the guideline of an existing proof [54, 111], based on the concept of decisive coalitions (as defined in Section 4.4.2). What is novel about our approach is that we show that this technique can be fully embedded into a formal derivation of the axiomatic system for $L[N, X]$ presented earlier. We offer an outline on the main steps of the proof, from which a complete formal derivation can be recovered.

The proof is based on two lemmas. The first lemma shows that, under certain conditions, a coalition being weakly decisive over a specific pair of alternatives implies that the same coalition is (not only weakly) decisive over all pairs.

4.5.2. Lemma. *Consider a language parametrised by X such that $|X| \geqslant 3$. Then for any coalition $C \subseteq N$ and any two distinct alternatives $x, y \in X$, we have that:*

$$\vdash \mathit{Par} \wedge \mathit{IIA} \wedge C\mathit{wdec}(x, y) \rightarrow C\mathit{dec}$$

Proof:
Suppose x, y, x' and y' are distinct alternatives.[4] To prove $C\mathit{dec}$ we need to prove

[4]With three alternatives the argument is analogous but simplified, since two of the alternatives coincide.

each of the conjuncts in the following formula:

$$\bigwedge_{x\in X}\bigwedge_{y\in X\setminus\{x\}}\left[\left(\bigwedge_{i\in C} p^i_{x\succcurlyeq y}\right)\to\neg y\right]$$

Now consider the following derivation:

(1) By finiteness of agents and alternatives and the theorems $p^i_{x'\succcurlyeq y'}\vee p^i_{y'\succcurlyeq x'}$ for all $i\in N$ we can, rearranging conjunctions and disjunctions, prove the consequent of the following formula; the implication follows.

$$(\bigwedge_{i\in C} p^i_{x'\succcurlyeq y'})\to[(\bigwedge_{i\in C} p^i_{x'\succcurlyeq y'})\wedge\bigvee_{C'\subseteq N\setminus C}((\bigwedge_{i\in C'} p^i_{x'\succcurlyeq y'})\wedge(\bigwedge_{i\notin C'\cup C} p^i_{y'\succcurlyeq x'}))]$$

(2) By applying distributivity to (1).

$$(\bigwedge_{i\in C} p^i_{x'\succcurlyeq y'})\to\bigvee_{C'\subseteq N\setminus C}[(\bigwedge_{i\in C} p^i_{x'\succcurlyeq y'})\wedge(\bigwedge_{i\in C'} p^i_{x'\succcurlyeq y'})\wedge(\bigwedge_{i\notin C'\cup C} p^i_{y'\succcurlyeq x'})]$$

(3) We will present the derivation of the following formula below.

$$\begin{aligned}&\bigwedge_{C\subseteq N}(\mathit{Par}\wedge\mathit{IIA}\wedge\mathit{Cwdec}(x,y)\to\\&[(\bigwedge_{i\in C} p^i_{x'\succcurlyeq y'})\wedge(\bigwedge_{i\in C'} p^i_{x'\succcurlyeq y'})\wedge(\bigwedge_{i\notin C'\cup C} p^i_{y'\succcurlyeq x'})\to\neg y'])\end{aligned}$$

(4) By propositional reasoning from (3).

$$\begin{aligned}&\mathit{Par}\wedge\mathit{IIA}\wedge\mathit{Cwdec}(x,y)\to\\&\bigvee_{C'\subseteq N\setminus C}[(\bigwedge_{i\in C} p^i_{x'\succcurlyeq y'})\wedge(\bigwedge_{i\in C'} p^i_{x'\succcurlyeq y'})\wedge(\bigwedge_{i\notin C'\cup C} p^i_{y'\succcurlyeq x'})\to\neg y']\end{aligned}$$

(5) By propositional reasoning from (2) and (4).

$$\mathit{Par}\wedge\mathit{IIA}\wedge\mathit{Cwdec}(x,y)\to[(\bigwedge_{i\in C} p^i_{x'\succcurlyeq y'})\to\neg y']$$

We still need to show (all the finitely many instances of) step (3). We prove each of them in the following way. Consider a specific profile $w=(\succcurlyeq_1,\ldots,\succcurlyeq_n)$ for which we can rearrange the conjuncts in the formula *profile*(w) as follows:

$$\begin{aligned}\mathit{profile}(w)\;&=\;(\bigwedge_{i\in C} p^i_{x\succcurlyeq y})\wedge(\bigwedge_{i\in N}(p^i_{x'\succcurlyeq x}\wedge p^i_{y\succcurlyeq y'}))\wedge\\&\quad(\bigwedge_{i\in C\cup C'} p^i_{x'\succcurlyeq y'})\wedge(\bigwedge_{i\notin C} p^i_{y\succcurlyeq x})\wedge(\bigwedge_{i\notin C\cup C'} p^i_{y'\succcurlyeq x'})\wedge\alpha\end{aligned}$$

Here α is the formula expressing the fact that all the other alternatives (if any) are ranked by all agents below x, y, x', y'. We are now ready to present a derivation for a specific conjunct of (3):

(a) For any $z \in X \setminus \{x, y, x', y'\}$:
$Par \wedge profile(w) \to \neg x \wedge \neg y' \wedge \neg z$
from formula *Par*, the second part of *profile*(w), and α

(b) $Cwdec(x, y) \wedge profile(w) \to \neg y$
by definition of $Cwdec(x, y)$

(c) $Par \wedge Cwdec(x, y) \to (profile(w) \to x')$
by axiom (14), encoding resoluteness, with (a) and (b)

(d) $\Diamond_N profile(w)$
by the Universal Domain Lemma

(e) $Par \wedge Cwdec(x, y) \to \Diamond_N(profile(w) \wedge x')$
by standard modal reasoning from (c) and (d)

(f) $Par \wedge IIA \wedge Cwdec(x, y) \to \Diamond_N(profile(w) \wedge x')$
by propositional reasoning from (e)

(g) $Par \wedge IIA \wedge Cwdec(x, y) \to [(profile(w)(x', y') \to \neg y')]$
from (f) and formula *IIA* w.r.t. x' and y'

But $profile(w)(x', y')$ consists of the following conjuncts:

$$(\bigwedge_{i \in C} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \in C'} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \notin C' \cup C} p^i_{y' \succcurlyeq x'})$$

Hence, we may infer that this latter formula entails $\neg y'$. Repeating this line of reasoning for all conjuncts we obtain (3); this concludes the proof. □

The next lemma establishes a syntactic counterpart of what is known as the *Contraction Lemma* in the literature [111]. It says that, under certain conditions, for any way of splitting a decisive coalition of two or more agents into two sub-coalitions, one of those sub-coalitions must also be decisive.

4.5.3. Lemma (Contraction Lemma). *Consider a language parametrised by X such that $|X| \geqslant 3$. Then for any coalition $C \subseteq N$ and any two coalitions C_1 and C_2 that form a partition of C, we have that:*

$$\vdash Par \wedge IIA \wedge C\,dec \to (C_1\,dec \vee C_2\,dec)$$

Proof:
Consider C, C_1 and C_2 as in the statement of the lemma (i.e. $C = C_1 \cup C_2$ and $C_1 \cap C_2 = \emptyset$) and let x, y, z be three distinct alternatives. Now consider any profile w for which *profile*(w) has the following form:

$$\begin{aligned} \mathit{profile}(w) \quad = \quad & (\bigwedge_{i \notin C_2} p^i_{x \succcurlyeq y}) \wedge (\bigwedge_{i \in C_1} p^i_{x \succcurlyeq z}) \wedge (\bigwedge_{i \in C_1 \cup C_2} p^i_{y \succcurlyeq z}) \wedge \\ & (\bigwedge_{i \in C_2} p^i_{y \succcurlyeq x}) \wedge (\bigwedge_{i \notin C_1} p^i_{z \succcurlyeq x}) \wedge (\bigwedge_{i \notin C_1 \cup C_2} p^i_{z \succcurlyeq y}) \wedge \alpha \end{aligned}$$

Here α encodes the fact that all other alternatives (if any) are ranked by all agents below x, y, z. By propositional reasoning and the fact that in profile w all agents in C prefer y over z we can derive:

$$C\,dec \rightarrow (\mathit{profile}(w) \rightarrow \neg z) \tag{4.1}$$

For any other alternative k different from x or y, we can derive:

$$Par \rightarrow (\mathit{profile}(w) \rightarrow \neg k) \tag{4.2}$$

This is so because α in *profile*(w) encodes the fact that all other alternatives are ranked by all agents below x, y, z. Formulas (4.1) and (4.2), together with axiom (14), encoding resoluteness, enforce that x or y must be the outcome:

$$Par \wedge C\,dec \rightarrow ((\mathit{profile}(w) \rightarrow x) \vee (\mathit{profile}(w) \rightarrow y)) \tag{4.3}$$

As an aside, we note that we know (again from resoluteness) that this disjunction must be exclusive. By the Universal Domain Lemma, we have that $\Diamond_N \mathit{profile}(w)$ is a theorem, and thus, using standard modal reasoning on formula (4.3), we obtain:

$$Par \wedge C\,dec \rightarrow (\Diamond_N(\mathit{profile}(w) \wedge x) \vee \Diamond_N(\mathit{profile}(w) \wedge y)) \tag{4.4}$$

Now propositional reasoning together with *IIA*, first w.r.t. the pair (x, z) and then w.r.t. the pair (y, x), allows us to derive from formula (4.4) the following formula:

$$Par \wedge IIA \wedge C\,dec \rightarrow ((\mathit{profile}(w)(x, z) \rightarrow \neg z) \vee (\mathit{profile}(w)(y, x) \rightarrow \neg x))$$

Recall that in *profile*(w) the agents in C_1 are the only ones supporting x over z. Hence, $(\mathit{profile}(w)(x, z) \rightarrow \neg z)$ means that C_1 is weakly decisive for the pair (x, z). Likewise, the agents in C_2 are the only ones supporting y over x; thus $(\mathit{profile}(w)(y, x) \rightarrow \neg x)$ means that C_2 is weakly decisive for the pair (y, x). In this fashion we can conclude that:

$$Par \wedge IIA \wedge C\,dec \rightarrow (C_1 wdec(x, z) \vee C_2 wdec(y, x)) \tag{4.5}$$

We can now use Lemma 4.5.2 and propositional reasoning on formula (4.5) to derive:

$$Par \wedge IIA \wedge C\,dec \rightarrow (C_1\,dec \vee C_2\,dec)$$

We have thus shown that $Par \wedge IIA \wedge C\,dec \rightarrow (C_1\,dec \vee C_2\,dec)$ must be a theorem of the logic. Note that the disjunction is still exclusive. □

We can now state and prove a syntactic counterpart of Arrow's Theorem:

4.5.4. Theorem (Arrow). *Let $L[N, X]$ be a logic with a language parametrised by X such that $|X| \geqslant 3$. Then we have:*

$$\vdash Par \wedge IIA \rightarrow Dic$$

Proof:
As mentioned earlier, Par is equivalent to $N\,dec$. Exploiting IIA, we can apply the Contraction Lemma and prove that one of two disjoint subsets of N is decisive. Repeating the process finitely many times (we have finitely many agents), we can show that one of the singletons that form N is decisive. But this is tantamount to saying that there exist a decisive agent, i.e., a dictator, so the formula $\bigvee_{i \in N} \{i\}\,dec$, which is equivalent to Dic. Hence, the formula $Par \wedge IIA \rightarrow Dic$ can be derived as a theorem of the logic $L[N, X]$ for any set X with $|X| \geqslant 3$ as claimed. □

Note that throughout the proof we have made implicit use of the condition $|X| \geqslant 3$ when assuming the availability of three distinct alternatives (in fact, in the proof of Lemma 4.5.2 we have only gone through the most interesting case, requiring at least four alternatives).

As we already mentioned, the proof provided here is not, strictly speaking, a full syntactic proof of Arrow's Theorem *within* the logic, because the language is parametric in the set of agents N and the set of alternatives X. Nevertheless, apart from the *proviso* on the number of alternatives stated in Theorem 4.5.4, our proof is independent of the choice of N and X; that is to say, this proof can be used as a *template* to prove the appropriate instance of Arrow's Theorem in *any* logic $L[N, X]$ for N and X such that $|X| \geqslant 3$.

Due to Theorem 4.3.6 establishing completeness of the logic and Lemmas 4.4.3, 4.4.5, and 4.4.13 establishing the correctness of our representation of the Arrovian conditions within the logic, Theorem 4.5.4 is equivalent to the usual, semantic, rendering of Arrow's Theorem for SCF's stated as Theorem 4.5.1. Thus, our purely syntactic proof constitutes an independent proof of the theorem. This shows that the logic $L[N, X]$ is a useful tool for reasoning about nontrivial concepts in social choice. In the remainder of this section we offer further support for this assertion, by proving two additional results.

4.5.2 Encoding the Muller-Satterthwaite Theorem

The Muller-Satterthwaite Theorem [96] establishes that, when there are at least three alternatives, the only SCF's that are strongly monotonic - and that do not rule out some of the alternatives as potential winners to begin with (by failing surjectivity) - are the dictatorships. Like Arrow's Theorem, this result shows that certain intuitively appealing properties of SCF's cannot be realised in general. We directly give a syntactic formulation of this important result in our logic.

4.5.5. Theorem (Muller-Satterthwaite). *Let $L[N, X]$ be a logic with a language parametrised by X such that $|X| \geqslant 3$. Then we have:*

$$\vdash SM \wedge Sur \rightarrow Dic$$

Proof:
We adopt the standard strategy, see e.g. [54], namely show that $\vdash SM \wedge Sur \rightarrow Par \wedge IIA$. Then, by the syntactic derivation of Arrow's Theorem given earlier, we obtain *Dic*. We begin by showing that the two premises entail IIA. If we can show that

$$SM \rightarrow [\Diamond_N(\mathit{profile}(w) \wedge x) \rightarrow (\mathit{profile}(w)(x, y) \rightarrow \neg y)]$$

for any w and any distinct x, y then we have that the two premises prove the conjunction of all such consequents, which is IIA.

The general strategy is the following: first we construct a profile w'' which ranks the alternatives x, y above all others and preserves the ordering of w encoded in $\mathit{profile}(w)(x, y)$; second, by *SM*, we conclude that x must be the outcome in this profile w''; third we show that, for any profile w' that agrees with $\mathit{profile}(w)(x, y)$, if the outcome at w' is y then, again by *SM*, the outcome at w'' is also y; this last passage contradicts the fact that x is the outcome at w'', hence y cannot be the outcome at any such w'. We proceed to encode this reasoning.

First, construct a formula representing w'':

$$\mathit{profile}(w'') := \mathit{profile}(w)(x, y) \wedge \bigwedge_{i \in N} \bigwedge_{z \neq x, y} (p^i_{x \succcurlyeq z} \wedge p^i_{y \succcurlyeq z})$$

By construction, we clearly have that:

$$\mathit{profile}(w'') \rightarrow \bigwedge_{k \in X \setminus \{x\}} N^w_{x \succcurlyeq k}$$

Together with *SM*, this latter formula readily entails the following implication:

$$SM \wedge \Diamond_N(\mathit{profile}(w) \wedge x) \wedge \mathit{profile}(w'') \rightarrow x$$

By the Universal Domain Lemma, we know that we have $\Diamond_N \mathit{profile}(w'')$. Thus:

$$SM \wedge \Diamond_N(\mathit{profile}(w) \wedge x) \rightarrow \Diamond_N(\mathit{profile}(w'') \wedge x) \tag{4.6}$$

This concludes the first two parts, showing that x must be the outcome for the profile w''. We now reason by contradiction, assuming

$$SM \wedge \Diamond_N(\mathit{profile}(w) \wedge x) \wedge \mathit{profile}(w)(x,y) \wedge y$$

and deriving $\Diamond_N(\mathit{profile}(w'') \wedge y)$, in contradiction with formula (4.6), thereby forcing us to conclude that the following holds:

$$SM \wedge \Diamond_N(\mathit{profile}(w) \wedge x) \wedge \mathit{profile}(w)(x,y) \rightarrow \neg y$$

This is then enough to infer one of the conjuncts of IIA. We can derive from the axioms that:

$$\mathit{profile}(w)(x,y) \wedge y \rightarrow \bigvee_{w'}[\mathit{profile}(w') \wedge \mathit{profile}(w)(x,y) \wedge y]$$

That is to say, there is a profile w' containing the preferences in $\mathit{profile}(w)(x,y)$ for which the outcome is y. By the Universal Domain Lemma, we can put a diamond in front of the profile formula. Hence, after some rearrangement we obtain:

$$\mathit{profile}(w)(x,y) \wedge y \rightarrow \bigvee_{w'} \Diamond_N[\mathit{profile}(w') \wedge y] \tag{4.7}$$

Notice now that the part inside the disjunction looks like the first formula in the antecedent of *SM*, formulated for variable y. Upon inspection we can also check that for all such w', we get:

$$\mathit{profile}(w'') \rightarrow \bigwedge_{k \neq x} N^{w'}_{x \succcurlyeq k}$$

This is the case because in w'' any k different from x, y is ranked below these two alternatives by all agents and moreover $\mathit{profile}(w)(x,y) = \mathit{profile}(w')(x,y) = \mathit{profile}(w'')(x,y)$. Thus we know that by applying *SM* we obtain:

$$SM \wedge \Diamond_N[\mathit{profile}(w') \wedge y] \wedge \mathit{profile}(w'') \rightarrow y$$

Now we can push *SM* inside the disjunction in formula (4.7), use the Universal Domain Lemma to get $\Diamond_N \mathit{profile}(w'')$, and apply the latter formula to conclude that each of the disjuncts entails $\Diamond_N(\mathit{profile}(w'') \wedge y)$. But then the whole disjunction entails it and we can derive:

$$\mathit{profile}(w)(x,y) \wedge y \rightarrow \Diamond_N(\mathit{profile}(w'') \wedge y)$$

This contradicts formula (4.6), since only one alternative can be the outcome and $x \neq y$. Hence, we have derived *IIA*.

Now for the derivation of *Par*. It is enough to show that $SM \wedge Sur$ entails each conjunct of the following form:

$$(\bigwedge_{i\in N} p^i_{x\succcurlyeq y}) \to \neg y \tag{4.8}$$

From *Sur* we know that $\bigvee_w \Diamond_N(\mathit{profile}(w) \wedge x)$. For each w we can construct a profile w'' which is the same as w but with the difference that x has been ranked over y by all agents:

$$\mathit{profile}(w'') := \bigwedge_{i\in N} p^i_{x\succcurlyeq y} \wedge \bigwedge_{z,k\neq x} \mathit{profile}(w)(z,k) \wedge \bigwedge_{y>_i z} p^i_{x\succcurlyeq z} \wedge \bigwedge_{z>_i y} \mathit{profile}(w)(x,z)$$

where $y >_i z$ in the subscript is just notation to mean that y is ranked over z by i in w and similarly for $z >_i y$. Clearly by this formula we have that if $z >_i y$ in w then in w'' their ranking in unchanged, while for $z = y$ and $y >_i z$ now x is ranked above z; thus

$$\mathit{profile}(w'') \to \bigwedge_{y\in X\setminus\{x\}} N^w_{x\succcurlyeq y}$$

Hence by *SM* we have that $\Diamond_N(\mathit{profile}(w) \wedge x) \wedge \mathit{profile}(w'') \to x$, that is, x is still the outcome in w''. Hence, every disjunct in *Sur* entails $\Diamond_N(\mathit{profile}(w'') \wedge x)$. Note that such profile w'' might be different for different disjuncts. Now notice that the antecedent in the formula (4.8), namely $(\bigwedge_{i\in N} p^i_{x\succcurlyeq y})$, by construction is just $\mathit{profile}(w'')(x,y)$ (for all w'' constructed in such fashion). So pushing the latter into the disjunction we obtain that each disjunct entails:

$$\Diamond_N(\mathit{profile}(w'') \wedge x) \wedge \mathit{profile}(w'')(x,y)$$

But this is the antecedent of *IIA*, hence each disjunct entails $\neg y$. Therefore the whole disjunction entails $\neg y$, and we have proved the desired implication (4.8). □

4.5.3 Encoding Sen's approach to rights

Sen's Theorem [110] shows that it is impossible to satisfy both the property of Pareto efficiency and the property of liberalism. Unlike the other impossibility theorems discussed, this result does not depend on any assumptions regarding the number of alternatives. We again give directly a syntactic formulation.

4.5.6. Theorem (Sen). *Consider any logic $L[N,X]$. Then we have:*

$$\vdash \neg(Par \wedge Lib)$$

Proof:
Our derivation will mirror the standard proof of the theorem [54, 110]. It is sufficient to show that ($Par \wedge Lib$) entails a contradiction. To make the notation lighter we will use the following abbreviation, meaning that an agent i is two-way decisive over the pair (x, y):

$$Lib^i(x, y) := \{i\}dec(x, y) \wedge \{i\}dec(y, x)$$

Consider only two of the conjuncts of Lib, say for agents i_1 and i_2. If we can prove that these two conjuncts together with Par entail a contradiction then we are done. Begin by rearranging the conjunction of disjunctions in the definition of Lib into a disjunction of conjunctions. For two agents this will look like this:

$$\bigvee_{x_1,x_2,y_1,y_2} (Lib^{i_1}(x_1, y_1) \wedge Lib^{i_2}(x_2, y_2)) \tag{4.9}$$

This formula essentially says that there are two pairs of elements on which the two agents are respectively two-way decisive. If we can prove that each of the disjuncts entails a contradiction, then by the laws of disjunction we can infer that the whole formula entails a contradiction. Note that we can push Par inside such a conjunction. Therefore, the task at hand is to show that formulas of the following shape entail a contradiction for every choice of the four alternatives:

$$Par \wedge (Lib^{i_1}(x_1, y_1) \wedge Lib^{i_2}(x_2, y_2))$$

We focus on the cases, i.e., the disjuncts, in which these are all distinct alternatives; the cases of two or three alternatives follow via a similar argument with some alternatives being identified. For each choice of $x_1, x_2, y_1, y_2 \in X$ build the profile $w_{x_{1,2},y_{1,2}}$ with the following properties:

- Individual i_1 ranks x_1 above y_1.
- Individual i_2 ranks x_2 above y_2.
- All individuals rank y_1 above x_2 and also y_2 above x_1.
- All individuals rank x_1, x_2, y_1, y_2 above all other alternatives.

These properties correspond to the following formulas:

- $p^{i_1}_{x_1 \succcurlyeq y_1}$
- $p^{i_2}_{x_2 \succcurlyeq y_2}$
- $\bigwedge_{i \in N}(p^i_{y_1 \succcurlyeq x_2} \wedge p^i_{y_2 \succcurlyeq x_1})$
- $\bigwedge_{i \in N}(p^i_{x_1 \succcurlyeq z} \wedge p^i_{x_2 \succcurlyeq z} \wedge p^i_{y_1 \succcurlyeq z} \wedge p^i_{y_2 \succcurlyeq z})$ for all other alternatives $z \in X$

Therefore, they will be part of a big conjunction forming $\mathit{profile}(w_{x_{1,2},y_{1,2}})$. By combining the latter two of the above formulas with the formula representing the Pareto condition, we can derive the following two formulas:

- $(Par \wedge \bigwedge_{i\in N}(p^i_{y_1 \succcurlyeq x_2} \wedge p^i_{y_2 \succcurlyeq x_1})) \rightarrow (\neg x_2 \wedge \neg x_1)$
- $(Par \wedge \bigwedge_{i\in N}(p^i_{y_1 \succcurlyeq z} \wedge p^i_{y_2 \succcurlyeq z} \wedge p^i_{x_1 \succcurlyeq z} \wedge p^i_{x_2 \succcurlyeq z})) \rightarrow \neg z$
  for all other alternatives $z \in X$

Thus, we can derive:

$$(Par \wedge \mathit{profile}(w_{x_{1,2},y_{1,2}})) \rightarrow (\neg x_2 \wedge \neg x_1 \wedge \neg z)$$

It is also easy to prove that the following two formulas hold:

- $Lib^{i_1}(x_1, y_1) \wedge p^{i_1}_{x_1 \succcurlyeq y_1} \rightarrow \neg y_1$
- $Lib^{i_2}(x_2, y_2) \wedge p^{i_2}_{x_2 \succcurlyeq y_2} \rightarrow \neg y_2$

Recall that the formulas $p^{i_1}_{x_1 \succcurlyeq y_1}$ and $p^{i_2}_{x_2 \succcurlyeq y_2}$ are also contained in $\mathit{profile}(w_{x_{1,2},y_{1,2}})$. Hence, summing up what we have seen so far, we obtain:

$$[Par \wedge Lib^{i_1}(x_1, y_1) \wedge Lib^{i_2}(x_2, y_2) \wedge \mathit{profile}(w_{x_{1,2},y_{1,2}})] \rightarrow [\neg x_1 \wedge \neg x_2 \wedge \neg y_1 \wedge \neg y_2 \wedge \overline{Z}]$$

where we use $\overline{Z}$ as a shorthand for the conjunction $\bigwedge_{z\in X\setminus\{x_1,x_2,y_1,y_2\}} \neg z$. The consequent of the implication above is a negation of all the alternatives in X, a formula that is inconsistent with the first part of axiom (14), the axiom encoding resoluteness of the SCF. Hence, we obtain:

$$[Par \wedge Lib^{i_1}(x_1, y_1) \wedge Lib^{i_2}(x_2, y_2) \wedge \mathit{profile}(w_{x_{1,2},y_{1,2}})] \rightarrow \bot \tag{4.10}$$

Thanks to the Universal Domain Lemma we know that the theorems of the logic include the formula $\Diamond_N \mathit{profile}(w_{x_{1,2},y_{1,2}})$. So if we are given $Par \wedge Lib^{i_1}(x_1, y_1) \wedge Lib^{i_2}(x_2, y_2)$, we can certainly deduce:

$$Par \wedge Lib^{i_1}(x_1, y_1) \wedge Lib^{i_2}(x_2, y_2) \wedge \Diamond_N \mathit{profile}(w_{x_{1,2},y_{1,2}})$$

By this formula, formula (4.10), and modal reasoning we can conclude:

$$[Par \wedge Lib^{i_1}(x_1, y_1) \wedge Lib^{i_2}(x_2, y_2) \wedge \Diamond_N \mathit{profile}(w_{x_{1,2},y_{1,2}})] \rightarrow \Diamond_N \bot$$

Since $\Diamond_N \bot \rightarrow \bot$ is a theorem of normal modal logic we get:

$$[Par \wedge Lib^{i_1}(x_1, y_1) \wedge Lib^{i_2}(x_2, y_2)] \rightarrow \bot$$

Thus, we have shown that one of the disjuncts of formula (4.9) implies a contradiction. Repeating the same proof for every permutation of the four alternatives, we can thus prove that the whole disjunction entails a contradiction. Therefore $(Par \wedge Lib)$ entails a contradiction and we are done. □

4.6 Implementing the logic

In this section we expand on the possibility of implementing the logic. As we will see, it is possible to translate the language of $L[N,X]$ into classical propositional logic, and more specifically into the propositional language used by Tang and Lin [115]. This paves the way for the application of SAT solvers to check the validity of formulas in our logic, thereby allowing for a fully automated check of the validity of the theorems formulated in this chapter.

The language for modelling social choice functions used by Tang and Lin consists of two predicates: $p(i,x,y,w)$, expressing that in profile w agent i prefers x over y, and $s(x,w)$, expressing that alternative x is the winner in profile w. In full generality, these predicates belong to a multi-sorted first order logic with variables for agents, alternatives, and profiles. However, when the number of agents and alternatives is fixed, we can translate the quantified formulas into propositional formulas substituting for the variables all the finitely many constants; this is how Tang and Lin obtain a propositional language that can be fed into a SAT solver. Formulas in the resulting propositional language are also evaluated on the models given in Definition 4.3.4:

- $M \models p(i,x,y,w)$ iff $x \succcurlyeq_i y$ in profile w
- $M \models s(x,w)$ iff $F(w) = x$

We show here how to adapt the so-called Standard Translation [32] from modal logic into first-order logic to a translation from our modal language into the multi-sorted first-order logic with predicates $p(i,x,y,w)$ and $s(x,w)$. Once this is done, the formulas of the latter language can be turned into propositional clauses and checked following the approach of Tang and Lin [115]. Consider the following translation of the language of $L[N,X]$ into the language with predicates $p(i,x,y,w)$ and $s(x,w)$. The translation is parametric in w, a variable ranging over profiles:

$$\begin{array}{rcl} t_w(p^i_{x\succcurlyeq y}) & \mapsto & p(i,x,y,w) \\ t_w(\neg\varphi) & \mapsto & \neg t_w(\varphi) \\ t_w(\varphi \wedge \psi) & \mapsto & t_w(\varphi) \wedge t_w(\psi) \\ t_w(x) & \mapsto & s(x,w) \\ t_w(\Diamond_C \varphi) & \mapsto & \exists w' (\bigwedge_{i \in N \setminus C} \bigwedge_{x \neq y \in X} [t_w(p^i_{x\succcurlyeq y}) \leftrightarrow t_{w'}(p^i_{x\succcurlyeq y})] \wedge t_{w'}(\varphi)) \end{array}$$

The other propositional connectives are handled accordingly. The next proposition establishes the connection between the model checking for $L[N,X]$ and the model checking for the propositional language associated to the same parameters.

4.6.1. Proposition. *For every formula φ in the language of $L[N,X]$, profile w, and SCF F, φ is satisfiable at M_F, w if and only if $t_w(\varphi)$ is satisfiable at M_F.*

Proof:
The proof proceeds by induction on the complexity of φ. The base cases are immediate by the translation and the semantics; we expand only on the case of the modality.

First, suppose $\Diamond_C\varphi$ is satisfiable. Then there are a SCF F and a profile w such that $M_F, w \models \Diamond_C\varphi$, which in turn entails that there is another profile w' with $\succcurlyeq_i = \succcurlyeq'_i$ for all $i \in N \setminus C$ such that $M_F, w' \models \varphi$. By the induction hypothesis, $t_{w'}(\varphi)$ is satisfiable at M_F. Since $\succcurlyeq_i = \succcurlyeq'_i$ is the case, we will have that $\bigwedge_{i\in N\setminus C}\bigwedge_{x,y\in X}[t_w(p^i_{x\succcurlyeq y}) \leftrightarrow t_{w'}(p^i_{x\succcurlyeq y})]$ is true at M_F. So we can conclude that $t_w(\varphi)$ is satisfiable at M_F when w' is the witness of the existential quantifier.

For the other direction, suppose $t_w(\Diamond_C\varphi)$ is satisfiable at M_F. Then there exists a profile w' such that $\bigwedge_{i\in N\setminus C}\bigwedge_{x,y\in X}[t_w(p^i_{x\succcurlyeq y}) \leftrightarrow t_{w'}(p^i_{x\succcurlyeq y})]$ and $t_{w'}(\varphi)$ are true. From the second formula and induction hypothesis we get that $M_F, w' \models \varphi$, while from the first we can conclude that $\succcurlyeq_i = \succcurlyeq'_i$ for all $i \in N \setminus C$. Thus, $M_F, w \models \Diamond_C\varphi$. □

To check for the satisfiability of a formula φ in $L[N, X]$ we can translate it into the propositional language, check the satisfiability of the resulting formula and use the last proposition to infer the satisfiability of φ. For the details of how to implement the propositional language in order to make it amenable for a SAT solver see [115].
The reader may now wonder: why are we using modal logic at all, if we can collapse everything to propositional logic? The key here is *size:* the readability of the formulas of $L[N, X]$, and therefore its usefulness as a tool for formalisation, is lost in the translation into propositional logic.

To make this point precise, we inductively define a function *size* assigning a size to each formula in a modal propositional language: the size of propositional atoms is 1, and the size of any other formula is the sum of the sizes of its immediate subformulas plus 1. For example, the size of $p \wedge \neg q$ is $1 + 1 + (1 + 1) = 4$. It is easy to see that the size contributed by the propositional atoms and the boolean connectives remains constant during the translation: $size(t_w(p^i_{x\succcurlyeq y})) = size(p(i, x, y, w)) = 1 = size(p^i_{x\succcurlyeq y})$, and similarly for the other cases. For the modality, however, we have a significant difference. The formula $\Diamond_C\varphi$ has size $size(\varphi) + 1$, while its translation $t_w(\Diamond_C\varphi)$ has size

$$|\mathcal{L}(X)|^n \times [s(t_w(\varphi)) + 1 + (|N \setminus C| \times (|X^2| - |X|) \times 4) - 1]$$

This formula comes from the definition of the translation. First of all, after X and N have been fixed, we have to transform the existential quantifier into a big disjunction over all possible profiles; this explains the multiplication with the factor $|\mathcal{L}(X)|^n$. Within the square brackets, we have to add the size of the translation of φ to the size of the formula $\bigwedge_{i\in N\setminus C}\bigwedge_{x,y\in X}[t_k(p^i_{x\succcurlyeq y}) \leftrightarrow t_{k'}(p^i_{x\succcurlyeq y})]$, plus 1 because of the conjunction. Now let us look at the latter formula. If we take the bi-implication between atomic propositions as primitive, the inner formula has size 3 (otherwise it would be even greater). This needs to be multiplied with the size of the complement of C and the size of $|X^2|$ minus all the pairs in the diagonal (we consider $x \neq y$). Counting the conjunction associated with each of the instances of the bi-implication and subtracting 1 for the additional conjunction that we are considering, we arrive at the formula above.

The reader can get a feel of the blow-up by considering the following example. Let us analyse the simple case in which there are 3 alternatives and 2 agents, and where C is a singleton. We take $\varphi = x$, an atomic proposition, so that $size(\varphi) = 1$ and $size(\Diamond_C \varphi) = 2$. On the other hand, the size of the translation into propositional logic is $size(t_k(\Diamond_C \varphi)) = (3!)^2 \times [1 + 1 \times (9 - 3) \times 4] = 36 \times 25 = 900$. Clearly, formulas of such size are unwieldy for humans; their best use is for automated reasoning.

Thus, the logic $L[N, X]$ can fulfill two roles in the study of social choice theory. First, as demonstrated in the main part of this chapter, it is a convenient formalism in which to cast proofs of theorems regarding the characterisation of SCF's in terms of basic properties. Second, as demonstrated in this section, it can serve as a convenient interface between social choice theory and propositional logic, with $L[N, X]$ ensuring readability and the propositional counterpart allowing for the use of standard computational tools, particularly SAT solvers, to automatically reason about the SCF's.

4.7 Related work

The idea of using formal methods to subject social procedures to the same kind of formal analysis routinely applied to algorithms and software systems can be traced back to, at least, the work of Parikh [98, 99]. The two main arguments motivating this kind of enterprise are obvious and well known: formal analysis will deepen our understanding of social procedures; and formal analysis can increase our confidence in the correctness of social procedures. Pauly [102] has suggested a third argument that is specific to the use of logic in social choice theory: the expressive power of a logical language required to express a choice-theoretic property (such a IIA) is a relevant criterion in judging the interestingness of a characterisation result making use of such a property. A fourth argument fueling this line of research is that it has the potential to uncover entirely new characterisation and impossibility results [37, 57, 115] - results that are of independent interest to economists [41].

Successful applications of logic and automated reasoning to social choice theory have included the automated verification of the correctness of practical algorithms for implementing voting rules [22] and the automated search for new impossibility theorems in the domain of ranking sets of objects [57]. However, most work to date has focussed on the Arrovian framework of preference aggregation and the challenges of representing Arrow's Theorem in a variety of logical frameworks [1,64], of verifying the correctness of existing proofs for the theorem [97, 122], and of finding new such proofs [115]. Indeed, Arrow's Theorem is arguably the best yardstick against which to measure new formal methods for reasoning about problems of social choice. The work of Lange et al. [79] on the use of automated reasoning in different areas of economic theory, such as auctions and cooperative games, demonstrates that the basic concepts and techniques developed for the seemingly narrow domain of Arrovian preference aggregation can have a ripple-on effect on the use of formal methods in economics more widely.

Regarding Arrow's Theorem, starting at the top as far as the expressive power of the logical systems employed is concerned, Nipkow [97] and Wiedijk [122] have shown how to verify existing proofs for the theorem in higher-order logic proof assistants. Grandi and Endriss [64] have shown that classical first-order logic is sufficiently expressive to model all aspects of Arrow's Theorem, with the sole exception being the requirement that the set of agents be finite (the theorem is not valid for infinite electorates; cf. the use of induction in the proof of Theorem 4.5.4). In particular, modelling IIA does not require second-order quantification. At the most extreme end of the spectrum, Tang and Lin [115] have shown that the theorem can be embedded into classical propositional logic, albeit only for a fixed set of agents and a fixed set of alternatives. This embedding itself ceases to be useful for deepening our understanding of social choice (as it involves thousands of clauses, even for the simplest case of $|N| = 2$ and $|X| = 3$). Instead, the significance of the work of Tang and Lin derives from the fact that they have been able to provide a fully automated proof of the theorem based on this embedding. The work of Ågotnes et al. [1], like our own work, is orthogonal to these other contributions, in that they design a new tailor-made logic for social choice theory, rather than encoding those concepts into already existing logics. Note that Troquard et al. [118], the originators of the logic $\Lambda^{\text{scf}}[N, X]$ we have used here, have themselves not attempted to model Arrow's Theorem.

Examples for work in this vein addressing results other than Arrow's Theorem are still rare. Tang and Lin [115] have extended their approach to proving Arrow's Theorem also to the Muller-Satterthwaite Theorem and to Sen's Theorem. Nipkow [97], besides treating Arrow's Theorem, also has verified a proof of the Gibbard-Satterthwaite Theorem using a higher-order logic proof assistant. Grandi and Endriss [64] also formalise Sen's Theorem.

To date, the approaches to modelling Arrow's Theorem in logical frameworks in Hilbert-style calculi, namely the contributions of Ågotnes et al. [1] and of Grandi and Endriss [64], have not yet yielded a complete proof of the theorem *within* that same logical framework, although Ågotnes et al. [1] do succeed in providing a syntactic proof of a relevant lemma. In recently published work, Perkov [104, 105] has outlined a natural deduction proof of Arrow's Theorem using the language of Ågotnes et al. [1]. There currently are no results of this kind available for either the Muller-Satterthwaite Theorem or Sen's Theorem.

A recent survey on logic and social choice theory [54] has identified three critical points in existing work on logics for modelling concepts in social choice: (1) whether the approach does not require us to fix the sets of agents and alternatives upfront, (2) whether the universal domain assumption can be expressed in an elegant manner, and (3) whether the approach facilitates automation. Regarding point (1), as discussed in Section 4.3.5, our logic is indeed subject to the common limitation of requiring us to fix the cardinalities of N and X before even the notion of a well-formed formula can be defined, but we have also demonstrated that in practice this limitation can be overcome by working with schemas parametrised by N and X. Point (2) is convincingly taken care of by Lemma 4.4.1, the Universal Domain Lemma. Point (3), finally, is addressed in Section 4.6, where we show how to reduce the satisfiability problem of the logic $L[N, X]$ to the satisfiability problem for propositional logic. Of course, to directly develop automated reasoning tools for $L[N, X]$, thereby foregoing the need for translation and the associated blow-up in the problem size, is still of some interest. Evidence for the claim that also this direction is feasible and promising is given by Troquard [117], who has initiated a study of algorithms for model checking for the full logic $\Lambda^{\text{scf}}[N, X]$, including a prototype implementation.

4.8 The link to path logic

Finally, we explain how all this relates to the previous chapters. The modality $\Diamond_C$ enables us to analyze what happens when a portion of the agents is allowed to change their ballots, while the actions of the others are kept fixed. This phenomenon is also expressible in path logic, when we are in the right setting.

We begin by showing how the frame for the models of $L[N, X]$ can be seen as a presheaf. Consider the poset category $\wp(N)$ consisting of coalitions of agents and inclusions between them. Define the presheaf $Prof : \wp(N)^{op} \to \mathbf{Set}$ (where $Prof$ stands for 'profiles'):

$$C \subseteq N \mapsto \mathcal{L}(X)^C$$
$$C \hookrightarrow C' \mapsto Prof(C \hookrightarrow C') : \mathcal{L}(X)^{C'} \to \mathcal{L}(X)^C$$

This presheaf assigns to each coalition C the set of all possible profiles over C, while the function $Prof(C \hookrightarrow C') : \mathcal{L}(X)^{C'} \to \mathcal{L}(X)^C$ sends a profile over C' to a profile over C by discarding the ballots of the agents in $C' \setminus C$.

Consider now the relational structure obtained from $Prof$ with the procedure outlined in Chapter 2, call it M_{Prof}. The carrier of this structure is

$$\biguplus\{\mathcal{L}(X)^C | C \subseteq N\}$$

that is, the states are all the possible profiles over all the possible coalitions in $\wp(N)$. Note that this includes all the profiles over N itself, namely the worlds of the models for the logic $L[N, X]$. In fact M_{Prof} is an equivalent presentation of the set of profiles $\mathcal{L}(X)^N$, which constitutes the frame of the models for $L[N,X]$: on one hand M_{Prof} contains the set $\mathcal{L}(X)^N$, on the other hand given $\mathcal{L}(X)^N$ we can canonically reconstruct M_{Prof} by considering all the possible restrictions of the profiles in $\mathcal{L}(X)^N$ to smaller coalitions.

A pair of states (v, v') is in the relation $R_{C \hookrightarrow C'}$ (recall that relations in M_{PA} are indexed by the arrows of the category, inclusions in this case) if

- v is a profile over the coalition C,
- v' is a profile over the coalition C',
- v is obtained by v' by discarding the ballots of the agents in $C' \setminus C$.

The next step is to explicate how to encode $L[N,X]$ into a path logic for the frame M_{Prof}. For the propositional variables take $At := \{p^i_{x \succcurlyeq y} \mid i \in N \text{ and } x, y \in X\} \cup X$. We are not interested in all models over M_{Prof}, but only in those where the interpretation of the atomic propositions in X is given by a SCF F; call a *social choice model over* M_{Prof} a pair $\langle M_{Prof}, F\rangle$ where the latter is a SCF F. On such models we can interpret the atomic propositions as before:

- $\langle M_{Prof}, F\rangle, w \models p^i_{x \succcurlyeq y}$ iff $w = \langle \succcurlyeq_1, \ldots, \succcurlyeq_k\rangle$, $1 \leqslant i \leqslant k$ and $x \succcurlyeq_i y$
- $\langle M_{Prof}, F\rangle, w \models x$ iff $w = \langle \succcurlyeq_1, \ldots, \succcurlyeq_n\rangle \in Prof(N)$ and $F(w) = x$

Therefore $p^i_{x \succcurlyeq y}$ is true if w is a profile over a coalition containing i and $x \succcurlyeq_i y$ is the case in w, while x is true at the worlds that are 'full' profiles, that is over the coalition N, and where $F(w) = x$.

This allows for a translation from $L[N,X]$ into the path logic $\mathbf{PL}_\omega(\wp(N), At)$, namely the finitary path logic for the base category $\wp(N)$, where At is defined as described above.

4.8.1. Definition. Define the translation $t : L[N,X] \to \mathbf{PL}_\omega(\wp(N), At)$ as follows:

- $t(p^i_{x \succcurlyeq y}) = p^i_{x \succcurlyeq y}$
- $t(x) = x$
- $t(\neg\varphi) = \neg t(\varphi)$
- $t(\psi \wedge \varphi) = t(\psi) \wedge t(\varphi)$
- $t(\Diamond_C \varphi) = \overline{\langle N \backslash C \hookrightarrow N\rangle}\langle N \backslash C \hookrightarrow N\rangle t(\varphi)$

This translation establishes a connection between the truth of formulas in the two kinds of models.

4.8.2. PROPOSITION. *For any φ in $L[N,X]$, any $w \in \mathcal{L}(X)^N$ and for any SCF F we have*

$$M_F, w \models \varphi \textit{ iff } \langle M_{Prof}, F\rangle, w \models t(\varphi)$$

Proof:
By induction on φ. The base case is immediate by the semantics, while the propositional cases are given by IH. For the case of the modality, suppose $M_F, w \models \Diamond_C\varphi$ is the case: by the semantics there is $w' = (\succcurlyeq'_1, \ldots, \succcurlyeq'_n) \in \mathcal{L}(X)^n$ such that $M_F, w' \vDash \varphi$ and $\succcurlyeq_i = \succcurlyeq'_i$ for all agents $i \in N\backslash C$. By construction w' is also a world in M_{Prof}, so by IH we can conclude that $\langle M_{Prof}, F\rangle, w' \vDash t(\varphi)$. Since $\succcurlyeq_i = \succcurlyeq'_i$ for all agents $i \in N\backslash C$, w and w' are mapped to the same restriction by the function $Prof(N\backslash C \hookrightarrow N)$; call v this profile over $N\backslash C$. Thus $\langle M_{Prof}, F\rangle, v \vDash \langle N\backslash C \hookrightarrow N\rangle t(\varphi)$, namely from v we can extend the profile to a profile over N and reach w' with the relation $R_{N\backslash C\hookrightarrow N}$. Finally, since w is mapped to v by the inverse of $R_{N\backslash C\hookrightarrow N}$, we have $\langle M_{Prof}, F\rangle, v \vDash \overline{\langle N\backslash C \hookrightarrow N\rangle}\langle N\backslash C \hookrightarrow N\rangle t(\varphi)$. The latter formula is $t(\Diamond_C\varphi)$, so we are done. The converse is proved following the same line of reasoning. □

We can then exploit this to transport the theorems of $L[N,X]$ into validities of $\mathbf{PL}_\omega(\wp(N), At)$ for the class of social choice models over M_{Prof}.

4.8.3. COROLLARY. *For any φ in $L[N,X]$, if $\vdash_{L[N,X]} \varphi$ then for any SCF F we have $\langle M_{Prof}, F\rangle \models \langle Id_N\rangle\top \to t(\varphi)$*

Proof:
If φ is a theorem of the logic then by completeness it must be a validity of the class of models for $L[N,X]$. By the previous Proposition, $\langle M_{Prof}, F\rangle, w \models t(\varphi)$ will be the case for any $w \in \mathcal{L}(X)^N$ and for any SCF F. So let v be a profile in M_{Prof} and F be a SCF: if v satisfies $\langle Id_N\rangle\top$ then it is a profile in $\mathcal{L}(X)^N$ and thus $\langle M_{Prof}, F\rangle, v \models t(\varphi)$, hence $\langle M_{Prof}, F\rangle \models \langle Id_N\rangle\top \to t(\varphi)$. □

The last corollary in particular applies to the impossibility theorems proved in the previous sections.

Before presenting our conclusions, we note that a categorical reading of Arrow's Theorem was also offered in [2]; interestingly, in this paper the author focuses on presheaves over the base category $\wp(X)$, namely the powerset of the set of alternatives, stressing how Independence of Irrelevant Alternatives corresponds to a particular naturality condition. The path logic corresponding to these presheaf models would contain modalities to extend and restrict the set of alternatives; it remains to be investigated whether this logic can be used to formalize notions and results from social choice theory.

4.9 Conclusions

In this chapters we demonstrated what well-chosen fragments of path logic can achieve in concrete areas, taking the formalization of Social Choice Theory as our case study. We proposed a simple modal logic for speaking about basic concepts of preference aggregation and showed how to encode in the logic some known proofs of pivotal results such as Arrow's Theorem, the Muller-Satterthwaite Theorem and Sen's Theorem. The logic in question is a fragment of a logic introduced by Troquard et al., which we have shown to be complete by adapting their original completeness proof. Inspired by the work of Tang and Lin, we furthermore have suggested a pragmatic approach to implementing automated reasoning tools for the logic via a translation into propositional logic. As opposed to the formalism of these authors, our logic is not only computationally tractable but also human readable, as witnessed by the aforementioned encodings. Finally, we explained how such modal logic falls under the scope of path logic.

The last observation provided an analysis of the modality $\Diamond_C$ in terms of a logic of varying coalitions, that is, the path logic for the poset category $\wp(N)$. The prominence of the coalitional aspect in Social Choice, highlighted by the encoding of the proofs presented in this chapter, indicates two possible directions for future research. The first is to what extent other classical results of Social Choice, such as May's Theorem on the characterisation of the simple majority rule [92], can be encoded in the logic of varying coalitions. The second is the study of which other multi-agent scenarios can be encoded in presheaves over $\wp(N)$ and studied with the corresponding path logic. The pattern restrict-extend described in Section 4.8, when applied to coalitions, exactly encodes the idea of allowing a set of agents to modify their action while the others' actions are kept fixed. This idea is at the heart of other multi-agent notions, e.g. in the definition of Nash equilibrium.

Chapter 5

Bisimulation for conditional modalities

5.1 Introduction

In the previous chapters we observed how different models can be seen as examples of presheaves over suitable categories and how a multi-purpose logic for presheaves can be used in different settings. The perspective of that line of research puts the semantics in the foreground, in the sense that the language arises from the choice of the base category. If for example the base category is the powerset of a set of agents, as in the last chapter, one obtains a logic of varying coalitions.

However, working in Modal Logic one is often interested in a specific language over a specific class of structures. In this case the language is fixed and has its independent motivations, thus the approach described in the previous chapters may not be of immediate help. Nevertheless, also in this case a category-theoretic mindset provides a uniform approach in defining basic notions and raising basic questions. Insofar as the language has an associated notion of (bi)simulation, identified via a Hennessy-Milner-like result, we can regard said class of models as a category where the (bi)simulations play the role of arrows.

Since arrows are first-class citizens from the perspective of Category Theory, this stance highlights the importance of some specific issues and casts new light on some known problems. Two questions that become prominent are:

(1) What is the right notion of bisimulation for a given modal operator?

(2) Is such notion closed under composition?

It turns out that there is a group of well-known modalities for which these questions are not settled, namely conditional modalities. This chapter provides a structured answer to these two issues for this class of operators. This work paves the way for the next chapter, where we analyze the different categories arising from several different languages.

5.2 Conditional modalities

The Modal Logic literature offers a number of examples of conditional modalities, developed for a variety of reasons: conditionals from conditional logic, conditional belief, relativized common knowledge, to name a few. Yet there has been little work so far in developing model-theoretic tools to study such operators, which have been used mainly for the purpose of modelling our intuitions. The notable exception is conditional belief. The problem of finding the right notion of bisimulation for conditional belief has been the focal point of some recent publications in the field of formal epistemology [8–10, 48, 49].

In this chapter we attempt to understand what is *conditional* about conditional modalities, proposing a framework that covers all the aforementioned operators. The cornerstone of our approach is a general notion of bisimulation for conditional modalities, where the latter are interpreted on selection functions. Conditional logics, together with selection functions, have a long history and tradition in philosophical logic [42, 83, 113, 121]; they have been used in various applications such as non-monotonic inference, belief change and the analysis of intentions.

To ensure that the notion of bisimulation is a good fit for the logic, the key result that one would like to obtain is the classical theorem establishing the correspondence between bisimilarity and modal equivalence, usually on some restricted class of models, echoing the analogous theorem for basic modal logic.[1] In other words, one wants to characterize exactly when two models are indistinguishable by means of a conditional modality.

Such result is however not the end of the story, a well behaved notion of bisimulation should also satisfy the following list of desiderata:

(1) The bisimulation should be **structural**, that is, it should not make reference to formulas of the modal language besides the atomic propositions featuring in the basic condition "if w and w' are bisimilar then for every p we have $w \in V(p)$ iff $w' \in V(p)$".[2]

(2) Ideally such bisimulation should be **closed under unions and relational composition**. The former ensures the existence of a largest bisimulation, while the latter guarantees that the related notion of bisimilarity is transitive.

(3) The definition of such bisimulation should be **modular**, that is, independent from additional parts of the structure that do not appear in the semantics of the conditional modality: two states should be indistinguishable only if they behave in the same way with respect to the features that the conditional modality can "detect".

[1] See [32].

[2] For example, a non-structural notion of bisimulation for conditional belief on epistemic plausibility models was given in [48], but was regarded as problematic by the author himself for the same reason.

(4) When the unconditional modality is amenable to different semantics, the bisimulation for the conditional version should **generalize** the bisimulation for the **un-conditional modality uniformly across semantics**.

We use this list as a benchmark to assess the quality of a notion of bisimulation. In this chapter we provide a notion of bisimulation for conditional modalities that complies with the list and prove the correspondence between bisimilarity and modal equivalence for the semantics on selection functions.

In the next section we prove the main results at the general level of conditional models, while in the rest of the chapter we showcase the versatility of our framework along three directions of applications. First, in Section 5.4 we discuss the case of conditional belief on plausibility models, deriving some undefinability observations along the way. In Section we conduct a similar analysis for conditional belief on evidence models, showing how we can handle the same operator interpreted on different semantics (as for point 4 in our list). Second, Section proves that our approach covers more than just conditional belief by applying it to the operator of relativized common knowledge. Finally, in Section we explain how the central definition and results are amenable for a multi-agent generalization. We discuss related work in Section and conclude in Section .

5.3 Bisimulation for conditional modalities

Consider the language $\mathcal{L}_{\rightsquigarrow}$ of conditional logic

$$\varphi ::= \quad p \,|\, \neg\varphi \,|\, \psi \wedge \varphi \,|\, \psi \rightsquigarrow \varphi$$

where $p \in At$, a set of atomic propositions. The formulas $\psi \rightsquigarrow \varphi$ are supposed to encode statements such as "φ is the case, conditional on ψ". The other connectives are defined in the usual way.

As a semantics, we consider *selection functions* of type $W \times \wp(W) \to \wp(W)$, along the lines of [83]. Similar considerations can be cast in the more general framework proposed by Chellas in [42], but the generality of neighborhood selection functions is not really needed here, neither to prove our results nor to encompass the examples we mentioned; we thus limit ourselves to Lewis' original proposal.

5.3.1. Definition. A *conditional model* is a tuple $\mathcal{M} = \langle W, f, V\rangle$ with W a non-empty set of worlds, a function $f : W \times \wp(W) \to \wp(W)$ called *selection function* and $V : W \to \wp(At)$ a valuation function. The selection function is required to satisfy two conditions:

(1) for all $w \in W$ we have $f(w, X) \subseteq X$;

(2) if $X \subseteq Y$ then, for all $w \in W$, if $f(w, Y) \subseteq X$ then $f(w, Y) = f(w, X)$.

The intuition behind the selection function is that $f(w, X)$ selects the worlds in X that are 'relevant' at w.[3] For a given model $\mathcal{M}$, the semantics of the language is defined recursively via an interpretation function $[\![-]\!]_{\mathcal{M}} : \mathcal{L}_{\rightsquigarrow} \to \wp(W)$, where for the propositional part of the language the clauses are the usual ones and for conditionals we have the Stalnaker-Lewis semantics:

$$w \in [\![\psi \rightsquigarrow \varphi]\!]_{\mathcal{M}} \quad \text{iff} \quad f(w, [\![\psi]\!]_{\mathcal{M}}) \subseteq [\![\varphi]\!]_{\mathcal{M}}$$

This encodes the idea that "φ is the case, conditional on ψ" in a world w iff all the ψ worlds that are relevant at w according to f are worlds that satisfy φ. As customary, via the interpretation function $[\![-]\!]_{\mathcal{M}}$ we can define a satisfaction relation $\vDash \subseteq W \times \mathcal{L}_{\rightsquigarrow}$ putting $\mathcal{M}, w \vDash \psi$ iff $w \in [\![\psi]\!]_{\mathcal{M}}$; we will freely switch between the two notations.

To motivate our semantic clauses above, let us first recall that Gabbay [56] argues that our most general intuitions about non-monotonic derivations are captured by consequence relations $\vdash_{NM}$ satisfying the following three conditions, that he calls Reflexivity, Cut and Cautious Monotonicity:

- $\varphi \vdash_{NM} \varphi$
- $\varphi \vdash_{NM} \psi$ and $(\varphi \wedge \psi) \vdash_{NM} \theta$ entail $\varphi \vdash_{NM} \theta$
- $\varphi \vdash_{NM} \psi$ and $\varphi \vdash_{NM} \theta$ entail $(\varphi \wedge \psi) \vdash_{NM} \theta$

The Cut condition is obviously only a very special case of Gentzen's Cut rule, and it is sometimes called Cautious Transitivity. We'll adopt this last terminology, in order to avoid any confusions with the standard Cut rule. In terms of our conditional language, these requirements amount to claiming the validity of the following schemas:

- $\varphi \rightsquigarrow \varphi$ (Reflexivity)
- $((\varphi \rightsquigarrow \psi) \wedge ((\varphi \wedge \psi) \rightsquigarrow \theta)) \to (\varphi \rightsquigarrow \theta)$ (Cautious Transitivity)
- $((\varphi \rightsquigarrow \psi) \wedge (\varphi \rightsquigarrow \theta)) \to ((\varphi \wedge \psi) \rightsquigarrow \theta)$ (Cautious Monotonicity)

In terms of selection functions, the semantic clauses corresponding to these validities are:

- $f(w, X) \subseteq X$ (Reflexivity)
- $f(w, Y) \subseteq X$ and $f(w, X \cap Y) \subseteq X'$ entail $f(w, Y) \subseteq X'$ (Cautious Transitivity)

[3] Where the vague term 'relevant' may assume different interpretations depending on the context: 'similar' in sphere models, 'plausible' in doxastic logic, 'normal' in default reasoning, and so on.

- $f(w, Y) \subseteq X$ and $f(w, Y) \subseteq X'$ entail $f(w, X \cap Y) \subseteq X'$
  (Cautious Monotonicity)

Indeed, it is easy to see that these clauses are exactly what is needed to validate the above three schemas. Moreover, they are more general than most other settings for conditional logic, conditional beliefs etc.[4] Such clauses are in fact equivalent to our requirements on conditional models, which constitute a more compact presentation.

5.3.2. PROPOSITION. *Conditional models are exactly those satisfying Gabbay's requirements, when formulated in terms of selection functions.*

It is clear that Reflexivity is exactly our clause (1); the following two lemmas show that, in the presence of Reflexivity, Cautious Transitivity and Cautious Monotonicity correspond to the two inclusions in our clause (2).

5.3.3. LEMMA. *Cautious Transitivity entails the left-to-right inclusion in the second condition on selection functions. In presence of Reflexivity, the latter condition entails Cautious Transitivity.*

Proof:
Suppose $X \subseteq Y$ and $f(w, Y) \subseteq X$. Substitute X' with $f(w, X)$ in the definition of Cautious Transitivity: the premises are now $f(w, Y) \subseteq X$, which we have by assumption, and $f(w, X \cap Y) = f(w, X) \subseteq f(w, X)$, which is trivially the case. By Cautious Transitivity we can then conclude $f(w, Y) \subseteq f(w, X)$, as desired.

For the other direction, assume $f(w, Y) \subseteq X$ and $f(w, X \cap Y) \subseteq X'$. To conclude $f(w, Y) \subseteq X'$ it is enough to derive $f(w, Y) \subseteq f(w, X \cap Y)$. Notice now that Y and $X \cap Y$ satisfy the antecedent of the second condition: on one hand $X \cap Y \subseteq Y$ by definition, on the other hand $f(w, Y) \subseteq X \cap Y$ follows from our assumption $f(w, Y) \subseteq X$ and Reflexivity $f(w, Y) \subseteq Y$. Thus applying the second condition we obtain $f(w, Y) \subseteq f(w, X \cap Y)$ and we are done. □

5.3.4. LEMMA. *Cautious Monotonicity entails the right-to-left inclusion in the second condition on selection functions. In presence of Reflexivity, the converse also holds.*

[4] In particular, one can show that Lewis' 'sphere models' are an example of conditional models. The later modification due to Grove [66], in order to model belief revision, is also a special case; interestingly, the appropriate selection function is suggested by Grove himself in [66] p. 159. As we will show, our clauses are weaker than the semantic requirements of conditional doxastic logic. A further example are the models for non-monotonic logics. Our conditions are more general than the models of, for example, the non-monotonic system P of Kraus, Lehmann and Magidor [77] or the conditional logic introduced by Halpern in [67].

Proof:
Suppose $X \subseteq Y$ and $f(w, Y) \subseteq X$. Replacing X' with $f(w, Y)$ in the definition of Cautious Monotonicity we obtain $f(w, Y) \subseteq X$ and $f(w, Y) \subseteq f(w, Y)$. The former is given by assumption and the latter is a tautology, so applying Cautious Monotonicity we can conclude $f(w, X) = f(w, X \cap Y) \subseteq X' = f(w, Y)$.

For the converse, assume $f(w, Y) \subseteq X$ and $f(w, Y) \subseteq X'$. To obtain $f(w, X \cap Y) \subseteq X'$ it is enough to show $f(w, X \cap Y) \subseteq f(w, Y)$. Notice that we have $f(w, Y) \subseteq X \cap Y$, by assumption $f(w, Y) \subseteq X$ and Reflexivity $f(w, Y) \subseteq Y$. Coupled with $X \cap Y \subseteq Y$, we are in position to use the right-to-left inclusion in the second condition, thus obtaining $f(w, X \cap Y) \subseteq f(w, Y)$. □

We now turn to the definition of bisimulation for conditional modalities, the notion that is supposed to capture when two models are indistinguishable from the perspective of our conditional language. First we lay out some notation: given a relation $R \subseteq W \times W'$, $X \subseteq W$ and $X' \subseteq W'$ define

- $R[X] := \{y \in W' | \exists x \in X, (x, y) \in R\}$
- $R^{-1}[X'] := \{x \in W | \exists y \in X', (x, y) \in R\}$

5.3.5. Definition. [Bisimulation] Given two conditional models $\mathcal{M}_1$ and $\mathcal{M}_2$, a *conditional bisimulation* is a non-empty relation $Z \subseteq W_1 \times W_2$ such that if $(w, w') \in Z$ then

- $V(w) = V(w')$,
- for all $X \subseteq W_1$ and $X' \subseteq W_2$ such that $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$ we have that for every $x \in f^1(w, X)$ there exists a $y \in f^2(w', X')$ (where f^2 is the selection function in $\mathcal{M}_2$) such that $(x, y) \in Z$, and vice versa.

The non-standard part of this definition, namely the quantification over subsets X and X' together with the additional requirement $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$, is meant to handle the precondition ψ in the conditional $\psi \rightsquigarrow \varphi$. One would want the sets X and X' in the definition to be modally definable. However, to ensure that those sets are modally definable we would have to quantify over the formulas in the language and this would clash with the desideratum of having a structural bisimulation. Our solution is to replace "modally definable" with a structural condition that is close enough.[5]

The relation of *(conditional) bisimilarity* is defined as the existence of a conditional bisimulation: two states w and w' are bisimilar iff there exists a conditional bisimulation Z such that $(w, w') \in Z$. In other words, the relation

[5]As a consequence of this quantification over subsets, the time needed to check for a bisimulation can be exponential on the size of the input models; this is however not surprising, since the bisimulation intends to capture an operator with preconditions.

of bisimilarity between models $\mathcal{M}_1$ and $\mathcal{M}_2$ is the union of all the bisimulation relations between these models. The next result implies that *bisimilarity is itself a bisimulation*, and hence it is the *largest bisimulation* between two given models.

5.3.6. PROPOSITION. *Conditional bisimulations are closed under unions.*

Proof:
Given a family of conditional bisimulations $\{Z_i \subseteq W_1 \times W_2\}_{i\in I}$, consider their union $\bigcup_{i\in I} Z_i$. Suppose $(w, w') \in \bigcup_{i\in I} Z_i$ holds and two sets $X \subseteq W_1$ and $X' \subseteq W_2$ are such that $\bigcup_{i\in I} Z_i[X] \subseteq X'$ and $(\bigcup_{i\in I} Z_i)^{-1}[X'] \subseteq X$.

To establish that $\bigcup_{i\in I} Z_i$ is a conditional bisimulation we need to show that for every $x \in f^1(w, X)$ there is $y \in f^2(w', X')$ such that $(x, y) \in \bigcup_{i\in I} Z_i$. Notice that from $(w, w') \in \bigcup_{i\in I} Z_i$ we can deduce that there is an index i for which $(w, w') \in Z_i$. We also know that

(1) $\{y|\exists x \in X(x, y) \in Z_i\} = Z_i[X] \subseteq \bigcup_{i\in I} Z_i[X] \subseteq X'$,

(2) $\{x|\exists y \in X'(x, y) \in Z_i\} = Z_i^{-1}[X'] \subseteq \bigcup_{i\in I} Z_i^{-1}[X'] \subseteq X$.

Therefore X and X' also satisfy the preconditions for the relation Z_i: applying the property of conditional bisimulation we obtain that for every $x \in f^1(w, X)$ there is $y \in f^2(w', X')$ such that $(x, y) \in Z_i$. But the latter fact entails $(x, y) \in \bigcup_{i\in I} Z_i$, we are done. The converse direction is proved symmetrically. $\Box$

The last proposition secures only half of our second desideratum for a notion of bisimulation (see list in Section 5.2). We postpone the matter or relational composition to the sext subsection. The next thing to check is that our definition is suited to our conditional language: *bisimilar states satisfy the same conditional formulas.*

5.3.7. DEFINITION. [$\mathcal{L}_{\rightsquigarrow}$-equivalence] We say that two worlds w, w' in conditional models $\mathcal{M}, \mathcal{M}'$ are $\mathcal{L}_{\rightsquigarrow}$-*equivalent* iff they satisfy the same formulas in $\mathcal{L}_{\rightsquigarrow}$: i.e. for every $\varphi \in \mathcal{L}_{\rightsquigarrow}$, $\mathcal{M}, w \vDash \varphi$ holds iff $\mathcal{M}', w' \vDash \varphi$ holds.

5.3.8. THEOREM (BISIMILARITY ENTAILS $\mathcal{L}_{\rightsquigarrow}$-EQUIVALENCE). *Given two conditional models $\mathcal{M}_1$ and $\mathcal{M}_2$, if $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is a conditional bisimulation, then w and w' are $\mathcal{L}_{\rightsquigarrow}$-equivalent.*

Proof:
The proof is by induction on the structure of formulas; the case of $p, \neg, \wedge$ are treated as usual, we only show the case of the conditional modality. Suppose Z is a conditional bisimulation, $(w, w') \in Z$ and $\mathcal{M}_1, w \vDash \psi \rightsquigarrow \varphi$. Note that by induction hypothesis on ψ we have that $[\![\psi]\!]_{\mathcal{M}_1}$ and $[\![\psi]\!]_{\mathcal{M}_2}$ satisfy the right requirements and therefore can act as X and X' in the preconditions of the bisimulation property. Because of $w \vDash \psi \rightsquigarrow \varphi$ we have $f^1(w, [\![\psi]\!]_{\mathcal{M}_1}) \subseteq [\![\varphi]\!]_{\mathcal{M}_1}$.

Now consider $v' \in f^2(w', [\![\psi]\!]_{\mathcal{M}_2})$. By vice versa of the bisimulation property we know that there exists a $v \in f^1(w, [\![\psi]\!]_{\mathcal{M}_1})$ such that $(v, v') \in Z$. By assumption and induction hypothesis on φ we get $\mathcal{M}_2, v' \vDash \varphi$. Since v' was generic we can conclude that $f^2(w', [\![\psi]\!]_{\mathcal{M}_2}) \subseteq [\![\varphi]\!]_{\mathcal{M}_2}$, thus $\mathcal{M}_2, w' \vDash \psi \rightsquigarrow \varphi$. For the converse use the other direction of the bisimulation property. □

Our next theorem is the key result of this paper, providing a partial converse to the previous result. This is an analogue of the Hennessy-Milner theorem from modal logic, stating that on finite models bisimilarity completely captures $\mathcal{L}_{\rightsquigarrow}$-equivalence. We first notice that, on finite models, if two sets X and X' satisfy the condition $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$ for the relation of $\mathcal{L}_{\rightsquigarrow}$-equivalence then we can build a formula α that approximates the role of X and X' as preconditions.

5.3.9. Lemma. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two finite conditional models and X and X' be two subsets of the first and second model respectively. Let Z be the relation of $\mathcal{L}_{\rightsquigarrow}$-equivalence. Call A the set of elements of $\mathcal{M}_1$ that have a $\mathcal{L}_{\rightsquigarrow}$-equivalent counterpart in X and denote with A' the corresponding set for X'. If $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$ then there is a formula α such that $[\![\alpha]\!]_{\mathcal{M}_1} = X \cup A$ and $[\![\alpha]\!]_{\mathcal{M}_2} = X' \cup A'$.*

Proof:
Suppose X and X' are two sets satisfying $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$. Notice that we can divide the domain of $\mathcal{M}_1$ into three disjoint parts

- X
- A, the set of elements having some $\mathcal{L}_{\rightsquigarrow}$-equivalent counterparts in X
- $W_1 \backslash (X \cup A)$

Notice how the conditions on X and X' ensure that the elements in A do not have any counterpart in $\mathcal{M}_2$: $a \in A$ cannot have a $\mathcal{L}_{\rightsquigarrow}$-equivalent counterpart in X', or otherwise a would be already in X; on the other hand a cannot have an $\mathcal{L}_{\rightsquigarrow}$-equivalent counterpart in $W_2 \backslash X'$ or X itself would violate the first precondition. A symmetric partition can be defined on the model $\mathcal{M}_2$, switching the roles of X and X'; we will indicate with A' the corresponding region in $\mathcal{M}_2$.

Since the image of X under Z lies within X', we know that the elements in X are not $\mathcal{L}_{\rightsquigarrow}$-equivalent to the elements outside X', thus the elements in $X \cup A$ are also not $\mathcal{L}_{\rightsquigarrow}$-equivalent to the elements outside X'. As we are dealing with finite models we can enumerate the elements in $X \cup A$, call them $x_1, \ldots, x_n$. Similarly, we can put the elements of $W_2 \backslash X'$ and $W_1 \backslash (X \cup A)$ all together in a finite list $y_1, \ldots, y_m$. By our assumptions and definition of the partition we know that every element in $X \cup A$ is not $\mathcal{L}_{\rightsquigarrow}$-equivalent to any element in $W_2 \backslash X'$ or $W_1 \backslash (X \cup A)$.

So for each i and j, with $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$, there is a formula ψ_{ij} such that $x_i \vDash \psi_{ij}$ and $y_j \nvDash \psi_{ij}$. We can thus construct a formula

$$\gamma := \bigvee_{1 \leqslant i \leqslant n} \bigwedge_{1 \leqslant j \leqslant m} \psi_{ij}$$

that is true at each x_i in $X \cup A$ and false at each y_j in $(W_2 \backslash X') \cup (W_1 \backslash (X \cup A))$. Symmetrically, there must be a formula γ' that is true at $X' \cup A'$ and false at $W_1 \backslash X \cup (W_2 \backslash (X' \cup A'))$. Now consider the formula

$$\alpha := \gamma \vee \gamma'$$

Let us have a closer look at the extension $[\![\alpha]\!]_{\mathcal{M}_1}$ of α in $\mathcal{M}_1$. We have that γ' is false outside X, hence its extension lies within X. As for γ, we know it is true at $X \cup A$ and false in $W_1 \backslash (X \cup A)$. Thus the extension of $\gamma \vee \gamma'$, and therefore of the formula α itself, is $X \cup A$. We can provide an analogous argument to show that the interpretation of α in $\mathcal{M}_2$ is $X' \cup A'$. □

5.3.10. Theorem ($\mathcal{L}_{\rightsquigarrow}$-equivalence entails bisimilarity). *Given two finite conditional models $\mathcal{M}_1$ and $\mathcal{M}_2$, if w and w' are $\mathcal{L}_{\rightsquigarrow}$-equivalent then they are bisimilar.*

Proof:
We show that the relation Z of $\mathcal{L}_{\rightsquigarrow}$-equivalence is a (conditional) bisimulation. Let $(w, w') \in Z$ and suppose Z does not satisfy the bisimulation property for sets X and X': this means that there is an $x \in f^1(w, X)$ such that for all $y \in f^2(w', X')$ we have $(x, y) \notin Z$ (assumption $\star$). By Lemma 5.3.9 we can build a formula α such that $[\![\alpha]\!]_{\mathcal{M}_1} = X \cup A$ and $[\![\alpha]\!]_{\mathcal{M}_2} = X' \cup A'$. We can now exploit the properties of selection functions to derive the desired contradiction.

Consider a generic element x' in $f^1(w, [\![\alpha]\!]_{\mathcal{M}_1})$. Since $f^1(w, [\![\alpha]\!]_{\mathcal{M}_1}) \subseteq [\![\alpha]\!]_{\mathcal{M}_1}$ by the first property of selection functions, we know that x' must be either in X or in A. If there is an element $x' \in f^1(w, [\![\alpha]\!]_{\mathcal{M}_1})$ in A, since we know that elements in A are not $\mathcal{L}_{\rightsquigarrow}$-equivalent to any element in W_2, we can build a formula β that is false at x' and true everywhere in W_2, thus a fortiori in $f^2(w', [\![\alpha]\!]_{\mathcal{M}_2})$. This gives us the contradiction that we want: $w \vDash \neg(\alpha \rightsquigarrow \beta)$ and $w' \vDash \alpha \rightsquigarrow \beta$. We can thus assume that $f^1(w, [\![\alpha]\!]_{\mathcal{M}_1}) \subseteq X$. This is enough to apply the second property of selection functions and conclude that $f^1(w, [\![\alpha]\!]_{\mathcal{M}_1}) = f^1(w, X)$.

This ensures that the element $x \in f(w, X)$ given by assumption $\star$ is indeed also in $f^1(w, [\![\alpha]\!]_{\mathcal{M}_1})$. If we now look at the set $f^2(w', [\![\alpha]\!]_{\mathcal{M}_2})$, repeating a reasoning similar to the one just outlined we can conclude that $f^2(w', [\![\alpha]\!]_{\mathcal{M}_2}) = f^2(w', X')$. By assumption we have that x is not $\mathcal{L}_{\rightsquigarrow}$-equivalent to any $y \in f^2(w', X')$. We can thus build a formula β that is false at x and true everywhere in $f^2(w', [\![\alpha]\!]_{\mathcal{M}_2})$; this gives us the contradiction $w \vDash \neg(\alpha \rightsquigarrow \beta)$ and $w \vDash \alpha \rightsquigarrow \beta$. □

5.3.1 Closure under composition

Closure under relational composition turns out to be more tricky: we need bisimulation to 'transfer' preconditions in a coherent manner. In this subsection we propose a sufficient condition to obtain closure under relational composition.

5.3.11. Definition. Given a conditional model $\mathcal{M} = \langle W, f, V \rangle$, define the *relevant worlds for* w to be $W^w = \bigcup_{Y \subseteq W} f(w, Y)$. The model $\mathcal{M}$ is *grounded* if, for any $X \subseteq W$, $X \cap W^w \neq \emptyset$ entails $f(w, X) \neq \emptyset$.

If $f(w, X)$ selects the worlds in X that are 'relevant' at w, the set W^w is the collection of all the relevant worlds for w, taking into account all possible preconditions.[6] A conditional model is grounded when, given a precondition X that is consistent with the collection of all worlds relevant for w, the selection function returns a non-empty set of relevant worlds for w in X. The idea that conditioning with sets that are consistent with the current information should yield consistent results is widespread in Formal Epistemology, see for example Lewis in [83] and Board in [34]. The following equivalent definition of grounded models will be useful in later sections.

5.3.12. Lemma. *A model $\mathcal{M}$ is* grounded *iff, for any $x \in W$, $x \in W^w$ entails $f(w, \{x\}) \neq \emptyset$.*

Proof:
The new condition is a special case of the main definition when instantiated to singletons, so one direction is given. For the right-to-left direction, suppose by contradiction that $X \cap W^w \neq \emptyset$ and $f(w, X) = \emptyset$. Let $x \in X \cap W^w$: we have $x \in W^w$ and thus $f(w, \{x\}) \neq \emptyset$. However, $\{x\} \subseteq X$ and $f(w, X) = \emptyset \subseteq \{x\}$ trigger the second condition on conditional models, which states that $f(w, X) = f(w, \{x\})$, contradiction. □

5.3.13. Definition. A conditional bisimulation $Z \subseteq W_1 \times W_2$ is *diffuse* if for every $x \in W_1$ there are $w \in W_1$ and $w' \in W_2$ such that $(w, w') \in Z$ and $x \in W_1^w$, and vice versa.

The idea of diffuse bisimulations is that every element in both models must be relevant for at least one world that is in the bisimulation relation.

5.3.14. Definition. A relation $R \subseteq X \times Y$ is *two-ways surjective* if for every $x \in X$ there is a $y \in Y$ such that $(x, y) \in R$ and for every $y \in Y$ there is an $x \in X$ such that $(x, y) \in R$.

[6] The notation is borrowed by Board, see [34] p.56.

5.3.15. Lemma. *Any diffuse conditional bisimulation between grounded models is two-ways surjective.*

Proof:
Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be such models and suppose $Z \subseteq W_1 \times W_2$ is a conditional bisimulation. Suppose moreover that Z is not two-ways surjective, say because there is an $x \in W_1$ with no counterpart in W_2. Take $\{x\}$ and $\emptyset$ and notice that they fulfill the preconditions of the property of conditional bisimulation: $Z[\{x\}] = \emptyset \subseteq \emptyset$ and $Z^{-1}[\emptyset] = \emptyset \subseteq \{x\}$.

Since the bisimulation is diffuse we know that there are $w \in W_1$ and $w' \in W_2$ such that $(w, w') \in Z$ and $x \in W_1^w$. From the latter fact we infer that $\{x\} \cap W_1^w \neq \emptyset$, thus by the fact that $\mathcal{M}_1$ is grounded we conclude that $f^1(w, \{x\}) \neq \emptyset$. Thanks to the assumption $f^1(w, \{x\}) \subseteq \{x\}$ on conditional models we can infer $f^1(w, \{x\}) = \{x\}$. Since $(w, w') \in Z$, we must conclude that for every $z \in f^1(w, \{x\})$ there is a $y \in f^2(w', \emptyset)$ such that $(z, y) \in Z$. However, by the first condition on selection function we have $f^2(w', \emptyset) \subseteq \emptyset$, so there can be no counterpart for x, contradiction. The other direction is proved analogously. □

5.3.16. Proposition. *Restricted to any class of grounded models, the notion of diffuse conditional bisimulation is closed under relational composition.*

Proof:
Suppose $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ are three grounded models and $Z_1 \subseteq W_1 \times W_2$ and $Z_2 \subseteq W_2 \times W_3$ are two diffuse conditional bisimulations connecting them. To show that their relational composition $Z_1; Z_2$ is also a diffuse conditional bisimulation we first need to show that it is not empty. By Z_1 being not empty we know that there is $(w, w') \in Z_1$. By the previous Lemma we know that Z_1 and Z_2 are two-ways surjective. The latter fact ensures that there is some w'' such that $(w', w'') \in Z_2$, thus $(w, w'') \in Z_1; Z_2$.

For the property of conditional bisimulation, suppose $(w, w'') \in Z_1; Z_2$. By definition it means that there is a w' such that $(w, w') \in Z_1$ and $(w', w'') \in Z_2$. Now consider two sets $X \subseteq W_1$ and $X'' \subseteq W_3$ such that $Z_1; Z_2[X] \subseteq X''$ and $(Z_1; Z_2)^{-1}[X''] \subseteq X$.

What we need to show is that for every $x \in f^1(w, X)$ there is a $z \in f^3(w'', X'')$ such that $(x, z) \in Z_1; Z_2$. The idea is to define a set X' that works as intermediate precondition and allows us to apply the property of conditional bisimulation of Z_1 and Z_2. Define

$$X' := \{y \in W_2 | \exists x \in X,\, (x, y) \in Z_1 \quad \text{or} \quad \exists z \in X'',\, (y, z) \in Z_2\}$$

We check that

- $Z_1[X] \subseteq X'$,

- $Z_1^{-1}[X'] \subseteq X$.

The first item holds by definition of X'. For the second item suppose $(x, y) \in Z_1$ and $y \in X'$. By two-ways surjectivity of Z_2 we know that there is a z such that $(y, z) \in Z_2$, hence $(x, z) \in Z_1; Z_2$. By definition of X' we can now make a case distinction. In the first case there is an element $x' \in X$ such that $(x', y) \in Z_1$. We can then conclude that $(x', z) \in Z_1; Z_2$ and thus by assumption $Z_1; Z_2[X] \subseteq X''$ we have $z \in X''$. But then by the latter fact and $(x, z) \in Z_1; Z_2$, coupled with $(Z_1; Z_2)^{-1}[X''] \subseteq X$, we can infer that $x \in X$. In the second case we have that there is a $z' \in X''$ such that $(y, z') \in Z_2$. This gives us immediately that $(x, z') \in Z_1; Z_2$ and thus by assumption $(Z_1; Z_2)^{-1}[X''] \subseteq X$ we can conclude $x \in X$.

Since X and X' fulfill the preconditions of the property of conditional bisimulation for Z_1, we can deduce that for every $x \in f^1(w, X)$ there is $y \in f^2(w', X')$ such that $(x, y) \in Z_1$. We can now repeat the same proof strategy for X' and X'' and apply the property of Z_2 to obtain that for every $y \in f^2(w', X')$ there is $z \in f^3(w'', X'')$ such that $(y, z) \in Z_2$. Concatenating this with the previous result we get the desired conclusion: for every $x \in f^1(w, X)$ there is a $z \in f^3(w'', X'')$ such that $(x, z) \in Z_1; Z_2$. The converse is proved symmetrically.

It remains to show that $Z_1; Z_2$ is diffuse. Let $x \in W_1$, we need to find $w \in W_1$ and $w'' \in W_3$ such that $(w, w'') \in Z_1; Z_2$ and $x \in W_1^w$. Since Z_1 is diffuse, we know there are $w \in W_1$ and $w' \in W_2$ such that $(w, w') \in Z_1$ and $x \in W_1^w$. By Z_2 being two-ways surjective we know there is w'' such that $(w', w'') \in Z_2$, thus $(w, w'') \in Z_1; Z_2$. The converse is proved symmetrically. □

5.3.17. Proposition. *Restricted to grounded models and diffuse conditional bisimulations, the relation of bisimilarity is an equivalence relation.*

Proof:
We need to show that the relation of bisimilarity is reflexive, symmetric and transitive. For reflexivity, it is immediate to see that the identity relation is a diffuse conditional bisimulation. The definition of diffuse conditional bisimulation is itself symmetric, hence the converse of a diffuse conditional bisimulation is always a diffuse conditional bisimulation; the symmetry for bisimilarity follows. As for transitivity, Proposition 5.3.16 ensures that if there are two diffuse conditional bisimulations Z_1 and Z_2 such that $(w, w') \in Z_1$ and $(w', w'') \in Z_2$ then there is a diffuse conditional bisimulation containing the pair (w, w''), namely the relational composition $Z_1; Z_2$. □

We will see that in the next two sections these restrictions vanish, because in those particular settings all models are grounded and all bisimulations are diffuse. In later sections we will encounter examples where the restriction does limit the scope of our results; we then characterize grounded models and diffuse bisimulations in those particular contexts.

5.4 Plausibility models

We now turn to applications, discussing our first example of conditional modality: *conditional belief* interpreted on plausibility models. Plausibility models are widely used in formal epistemology [20,25]; their introduction can be traced back at least to [83]. They consist of a carrier, to be understood as a collection of possible worlds, and a preorder for each world, representing how an agent ranks the possible scenarios in terms of plausibility, from the perspective of the current world.

5.4.1. Definition. A *plausibility model* is a tuple $\mathcal{M} = \langle W, \{\leqslant_w\}_{w\in W}, V\rangle$ with W a non-empty set of worlds, a family of reflexive and transitive relations $\leqslant_w \subseteq W \times W$ and $V : W \to \wp(At)$ a valuation function.

The strict relation $<_w$ is defined as usual from $\leqslant_w$. Given a set $X \subseteq W$, let

$$Min_w(X) = \{v \in X | \neg\exists v' \in X \quad \text{s.t.} \quad v' <_w v\}$$

We can think of $Min_w(X)$ as the set of most plausible worlds in X with respect to w.[7] When we want to specify the ordering we write $Min_{\leqslant_w}(X)$.

Among the variety of operators that are studied in the setting of plausibility models, a prominent part is played by the operator of conditional belief, usually written as $B^\psi\varphi$. The standard belief operator can be defined via the conditional one as $B^\top\varphi$. On plausibility models the semantic clauses for belief and conditional belief are:

- $\mathcal{M}, w \vDash B\varphi$ iff for all $v \in Min_w(W)$ we have $\mathcal{M}, v \vDash \varphi$
- $\mathcal{M}, w \vDash B^\psi\varphi$ iff for all $v \in Min_w([\![\psi]\!]_\mathcal{M})$ we have $\mathcal{M}, v \vDash \varphi$

The notion of bisimulation for the standard belief operator on plausibility models, together with the corresponding Theorem, are both folklore.

5.4.2. Definition. Given two plausibility models $\mathcal{M}_1$ and $\mathcal{M}_2$, a *plausibility B-bisimulation* is a non-empty relation $Z \subseteq W_1 \times W_2$ such that if $(w, w') \in Z$ then

- $V(w) = V(w')$;
- for every $x \in Min_w W_1$ there is $y \in Min_{w'} W_2$ such that $(x, y) \in Z$, and vice versa.

5.4.3. Theorem. *Bisimilarity with respect to plausibility B-bisimulation entails modal equivalence with respect to the language with only the belief operator. On models having finitely many minimal elements, modal equivalence with respect to the latter language entails bisimilarity for plausibility B-bisimulation.*

[7] We sometimes omit the parenthesis in $Min_w(X)$ in what follows.

5.4.1 Plausibility CB-bisimulation

To obtain a bisimulation for conditional belief on plausibility models we show how the latter are an instance of conditional models; this move will indicate a systematic way to specialize the results of Section 5.3 to this particular context.

5.4.4. Definition. A plausibility model $\mathcal{M}$ is *well-founded* if it contains no infinite descending chains for any of the relations $\leqslant_w$.[8]

5.4.5. Proposition. *Well-founded plausibility models are conditional models, where* $f(w, X) = Min_w X$.

Proof:
We need to check that the newly defined f fulfills the prerequisites of selection functions in Definition 5.3.1. The first condition on selection functions is fulfilled by the very definition of Min_w. For the second one, suppose $X \subseteq Y$, $Min_w Y \subseteq X$ and take $x' \in Min_w Y$. Since $X \subseteq Y$, if there is no element below x' in Y then a fortiori there is no element below it in the subset X, thus in this circumstance $x' \in Min_w X$. For the other inclusion take $x' \in Min_w X$; we show x' is also minimal for Y. By contradiction, suppose there is $z \in Y \backslash X$ such that $z <_w x'$. Since we are in a well-founded model there must be a minimal element $z' \in Min_w Y$ such that $z' \leqslant_w z$; but by assumption $Min_w Y \subseteq X$, hence $z' \in X$ and $z' < x'$, contradicting the fact that x' is minimal in X. □

Notice that, setting $f(w, X) = Min_w X$, the definition of the satisfaction relation for conditional belief becomes an instance of the satisfaction relation for conditional modalities given in Section 5.3. If we now replace the new f in Definition 5.3.5, we obtain a new notion of bisimulation for conditional belief on plausibility models.

5.4.6. Definition. Given two plausibility models $\mathcal{M}_1$ and $\mathcal{M}_2$, a *plausibility CB-bisimulation* is a non-empty relation $Z \subseteq W_1 \times W_2$ s. t. if $(w, w') \in Z$ then

- $V(w) = V(w')$,
- for all $X \subseteq W_1$ and $X' \subseteq W_2$ such that $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$ we have that for every $x \in Min_w X$ then there exists a $y \in Min_{w'} X'$ such that $(x, y) \in Z$, and vice versa.

Since finite plausibility models are well-founded, we can now transfer the results of Section 5.3 on the correspondence between bisimilarity and modal equivalence. Throughout this section and the following one we use 'modal equivalence' meaning with respect to the language of conditional belief.

[8]Equivalently, assuming the axiom of choice, if every non empty subsets has minimal elements for all the relations $\leqslant_w$.

5.4.7. Theorem. *Given two plausibility models $\mathcal{M}_1$ and $\mathcal{M}_2$, if $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is a plausibility CB-bisimulation, then w and w' are modally equivalent. On finite plausibility models, if w and w' are modally equivalent then $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is a plausibility CB-bisimulation.*

We can also import the results concerning the closure under union and relational composition. Note that with the current definition of f the notation W^w trivializes: $W^w = \bigcup_{Y \subseteq W} f(w, Y) = \bigcup_{Y \subseteq W} Min_w(Y) = \bigcup_{\{x\} \subseteq W} Min_w(\{x\}) = W$. In other words, all the worlds in the model are relevant for every $w \in W$.

5.4.8. Lemma. *Every well-founded plausibility model is a grounded conditional model and every plausibility CB-bisimulation is diffuse.*

Proof:
Given a well-founded plausibility model $\mathcal{M}$ and $X \subseteq W$, if $X \cap W^w \neq \emptyset$ then $X \cap W \neq \emptyset$ therefore $X \neq \emptyset$. So by well-foundedness $f(w, X) = Min_w X \neq \emptyset$. This shows that the model is grounded. For the second part of the claim, let $Z \subseteq W_1 \times W_2$ be a plausibility CB-bisimulation and $x \in W_1$. Since the bisimulation is non-empty, there are $(w, w') \in Z$ and furthermore $x \in W_1 = W_1^w$, hence Z is diffuse. The same holds for every $y \in W_2$. □

5.4.9. Proposition. *On the class of well-founded plausibility models, the notion of plausibility CB-bisimulation is closed under arbitrary unions and relational composition.*

5.4.2 Undefinability

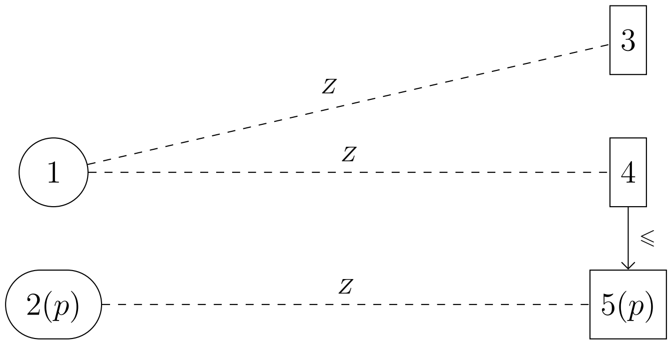
In this subsection we put the new notion of bisimulation to use, addressing the problem of inter-definability between conditional belief and other widely-used operators. For the rest of this section we employ plausibility models where $\leqslant_w$ is the same for all w, we thus remove the subscript. We begin with the operator of safe belief introduced in [21]:

$$\textit{Safe belief}: \mathcal{M}, w \vDash [\leqslant]\varphi \text{ iff for all } v \leqslant w \text{ we have } \mathcal{M}, v \vDash \varphi.$$

The dual operator is customarily defined as $\langle\leqslant\rangle\varphi := \neg[\leqslant]\neg\varphi$.

5.4.10. Proposition. *On plausibility models, safe belief is not definable in terms of the conditional belief operator.*

Proof:
Suppose $\langle\leqslant\rangle p$ is definable by a formula α in the language of conditional belief. Consider the two models depicted on the left and right side of the following picture (we omit reflexive arrows). We indicate within parenthesis the propositional atoms that are true at every world and with Z a CB-bisimulation between the models:



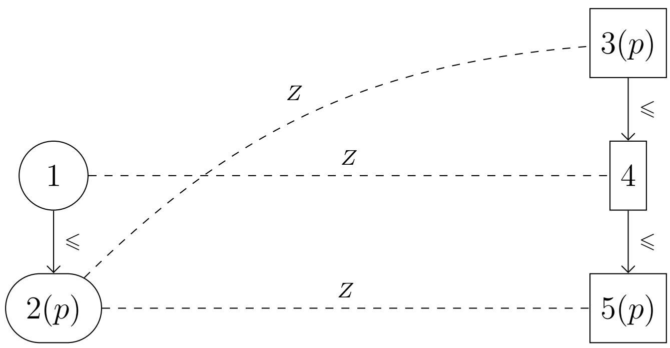
To check that Z is a CB-bisimulation, notice that only three pairs of sets fulfill the right precondition: $(\{1\},\{3,4\})$, $(\{2\},\{5\})$ and $(\{1,2\},\{3,4,5\})$. It is easy to see that the minimal elements of these pairs are connected by the bisimulation. Given that α is a formula in the language of conditional belief, it will be invariant between states that are bisimilar according to a CB-bisimulation. However, $\langle\leqslant\rangle p$ is true in the second model at 4 but false in the first model at 1; contradiction. □

The CB-bisimulation Z of this counterexample is not a bisimulation for safe belief, since it fails to satisfy the zig-zag condition: there are worlds 1, 5 and 4 such that $(1,4) \in Z$ and $5 \leqslant 4$ but no world w such that $w \leqslant 1$ and $(w,5) \in Z$. We now address the case of the strong belief operator, also introduced in [21].

Strong belief: $\mathcal{M}, w \vDash Sb\varphi$ iff there is $k \in W$ such that $\mathcal{M}, k \vDash \varphi$ and for all v, v' if $\mathcal{M}, v \vDash \varphi$ and $\mathcal{M}, v' \vDash \neg\varphi$ then $v \leqslant v'$.

5.4.11. Proposition. *On plausibility models, strong belief is not definable in terms of the conditional belief operator.*

Proof:
Suppose Sbp is definable by a formula α in the language of conditional belief. Consider the two models displayed below, where Z a CB-bisimulation and the propositional variables are attached to worlds as before:



The formula α in the language of conditional belief must be invariant between states that are bisimilar according to a CB-bisimulation; nevertheless, Sbp is true in the first model at 1 but false in the second model at 4, thus α will be true in one world and not in the other: contradiction. □

We now turn our attention to the definability of the conditional belief operator itself. We first warm up with a definition and two auxiliary observations.

5.4.12. DEFINITION. A *BSB-bisimulation*, a bisimulation for standard belief and safe belief, is a B-bisimulation satisfying an additional condition, namely the usual zig-zag condition for the $\leqslant$ relation: given two plausibility models $\mathcal{M}$ and $\mathcal{M}'$ and two worlds w and w' in the respective models, if $(w, w') \in Z$ then

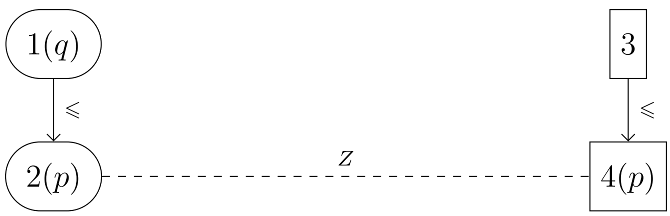
- for every $v \in W$ such that $v \leqslant w$ there is a $v' \in W'$ such that $(v, v') \in Z$ and $v' \leqslant w'$
- for every $v' \in W'$ such that $v' \leqslant w'$ there is a $v \in W$ such that $(v, v') \in Z$ and $v \leqslant w$

5.4.13. PROPOSITION. *On plausibility models, if two states w and w' are in a BSB-bisimulation then they are modally equivalent with respect to the language containing the belief and safe belief operators.*

Proof:
Straightforward induction on the complexity of the formula. □

5.4.14. PROPOSITION. *On plausibility models, conditional belief is not definable in terms of the language containing the operators of safe belief and standard belief.*

Proof:
Suppose $B^{\neg p}q$ is definable by a formula α in the language of belief and safe belief. Consider the two models displayed below, where Z a BSB-bisimulation and the propositional variables are attached to worlds as before:



Since 2 and 4 are in a BSB-bisimulation, by Proposition 5.4.13 they are modally equivalent in the language of belief and safe belief. Thus we can conclude $2 \vDash \alpha$ iff $4 \vDash \alpha$. But $2 \vDash B^{\neg p}q$ and $4 \nvDash B^{\neg p}q$, contradiction. □

Notice that the bisimulation used in this counterexample is not a plausibility CB-bisimulation.

5.5 Evidence models

We now change the semantics of the belief operator to evidence models, showing how the passage to conditional belief in this different setting follows the same pattern as in plausibility models; this allows us to conclude that the generalization from un-conditional to conditional modality works uniformly across semantics (see item 4 in our checklist in Section 5.2).

Introduced in [29], evidence models are structures capturing the evidence available to an agent in different possible worlds. The evidence available at a world w is represented via a family of sets of possible worlds: intuitively each set in the family constitutes a piece of evidence that the agent can use to draw conclusions at w. They constitute a generalization over plausibility models, but can be collapsed to plausibility models by considering the specialization preorder induced by the sets of evidence, however not without loss of information.[9]

5.5.1. Definition. An *evidence model* is a tuple $\mathcal{M} = \langle W, E, V\rangle$ with W a non-empty set of worlds, a function $E : W \to \wp(\wp(W))$ and $V : W \to \wp(At)$ a valuation function.

We indicate with $E(w)$ the set of subsets image of w. We furthermore assume $W \in E(w)$ and $\emptyset \notin E(w)$ for all $w \in W$.

The last requirement ensures that at every possible world the agents has trivial evidence, namely the whole set W, and does not have inconsistent evidence, i.e. the empty set.

5.5.2. Definition. A *w-scenario* is a maximal family $\mathcal{X} \subseteq E(w)$ having the finite intersection property (abbreviated in 'f.i.p.'), that is, for each finite subfamily $\{X_1, \dots, X_n\} \subseteq \mathcal{X}$ we have $\bigcap_{1\leqslant i\leqslant n} X_i \neq \emptyset$. Given a set $X \subseteq W$ and a collection $\mathcal{X} \subseteq E(w)$, the latter has the f.i.p. relative to X if for each finite subfamily $\{X_1, \dots, X_n\} \subseteq \mathcal{X}^X = \{Y \cap X | Y \in \mathcal{X}\}$ we have $\bigcap_{1\leqslant i\leqslant n} X_i \neq \emptyset$. We say that $\mathcal{X}$ is an *w-X-scenario* if it is a maximal family with the f.i.p. relative to X.

The semantics for belief and conditional belief on evidence models is:

- $\mathcal{M}, w \vDash B\varphi$ iff for every w-scenario $\mathcal{X}$ we have $\mathcal{M}, v \vDash \varphi$ for all $v \in \bigcap \mathcal{X}$
- $\mathcal{M}, w \vDash B^{\psi}\varphi$ iff for every w-$[\![\psi]\!]$-scenario $\mathcal{X}$ we have $\mathcal{M}, v \vDash \varphi$ for all $v \in \bigcap \mathcal{X}^{[\![\psi]\!]}$

The notion of bisimulation for the standard belief operator on evidence models establishes a connection between the scenarios of the two models:

[9]Evidence models contain information that is lost in the transition to plausibility models; such information is captured by operators such as the evidence modality. See [25, 26] for a discussion on the relationship between the two kinds of models. The sphere systems of [66] also constitute an example of neighborhood models with a close tie to relational structures.

5.5.3. Definition. Given two evidence models $\mathcal{M}_1$ and $\mathcal{M}_2$, an *evidence B-bisimulation* is a non-empty relation $Z \subseteq W_1 \times W_2$ s.t. if $(w, w') \in Z$ then

- $V(w) = V(w')$;
- for every w-scenario $\mathcal{X}$ and $x \in \bigcap \mathcal{X}$ there is a w'-scenario $\mathcal{Y}$ and $y \in \mathcal{Y}$ such that $(x, y) \in Z$, and vice versa.

The following result can be proven via the standard line of reasoning.

5.5.4. Theorem. *Bisimilarity with respect to evidence B-bisimulation entails modal equivalence with respect to the language with only the belief operator. On finite models, modal equivalence with respect to the latter language entails bisimilarity for evidence B-bisimulation.*

5.5.1 Evidence CB-bisimulation

We show that finite evidence models are an example of conditional models by means of two auxiliary lemmas.

5.5.5. Lemma. *On finite models, suppose $Y \supseteq X$. Then for every w-X-scenario $\mathcal{X}$ there is a w-Y-scenario $\mathcal{Y}$ such that $\mathcal{X} \subseteq \mathcal{Y}$. Conversely, for every w-Y-scenario $\mathcal{Y}$ there is a w-X-scenario $\mathcal{X}$ such that $\mathcal{X} \subseteq \mathcal{Y}$.*

Proof:
Let $\mathcal{X}$ be a w-X-scenario. Clearly $\mathcal{X}$ already has the f.i.p. relative to Y. Enumerate the sets K in $E(w)$ (there are finitely many), then proceed following the enumeration: if $K \in \mathcal{X}$ or $\mathcal{X} \cup \{K\}$ has the f.i.p. relative to Y then put K in $\mathcal{Y}$, otherwise not. Because of the first condition we get $\mathcal{X} \subseteq \mathcal{Y}$, while from the second one we obtain that $\mathcal{Y}$ is a w-Y-scenario.

For the second claim, enumerate the sets in $\mathcal{Y}$: $K_0, \ldots, K_m$. Construct $\mathcal{X}$ in stages beginning from $\mathcal{X}_0 = \emptyset$ and putting $\mathcal{X}_{n+1} = \mathcal{X}_n \cup \{K_n\}$ if $\bigcap \mathcal{X}_n^X \cap K_n \neq \emptyset$. Clearly $\mathcal{X} \subseteq \mathcal{Y}$. To see that $\mathcal{X}$ is maximal with the f.i.p. relative to X suppose that there is $K \notin \mathcal{X}$ such that $\bigcap \mathcal{X}^X \cap K \neq \emptyset$. By construction, if $\bigcap \mathcal{X}^X \cap K \neq \emptyset$ and $K \notin \mathcal{X}$ then $K \notin \mathcal{Y}$, hence by the maximality of $\mathcal{Y}$ it must be that $\bigcap \mathcal{Y}^Y \cap K = \emptyset$. Since $\bigcap \mathcal{X}^X \subseteq \bigcap \mathcal{Y}^Y$ by construction we get a contradiction. Therefore $\mathcal{X}$ is maximal with the f.i.p. relative to X. □

5.5.6. Lemma. *On finite models, if $Y \supseteq X$ then, for every w-X-scenario $\mathcal{X}$ and w-Y-scenario $\mathcal{Y}$ such that $\mathcal{X} \subseteq \mathcal{Y}$, if $y \in \bigcap \mathcal{Y}^Y$ then either $y \in \bigcap \mathcal{X}^X$ or $y \in Y \backslash X$. If no element $y \in \bigcap \mathcal{Y}^Y$ is in $Y \backslash X$ then $\bigcap \mathcal{X}^X = \bigcap \mathcal{Y}^Y$.*

Proof:
Let $y \in \bigcap \mathcal{Y}^Y$ and $y \notin Y \backslash X$. Then, since $y \in Y$, it must be that $y \in X$.

Since $y \in \bigcap \mathcal{Y}^Y$ we have that $y \in K$ for all $K \in \mathcal{Y}$, and hence $y \in K$ for all $K \in \mathcal{X}$. So $y \in \bigcap \mathcal{X}^X$. We can thus conclude that, if $y \notin Y \backslash X$ for all $y \in \bigcap \mathcal{Y}^Y$, $\bigcap \mathcal{X}^X \supseteq \bigcap \mathcal{Y}^Y$. For the other inclusion suppose $z \in \bigcap \mathcal{X}^X$ but not in $\bigcap \mathcal{Y}^Y$. Then there must be $K \in \mathcal{Y}$ such that $K \notin \mathcal{X}$ and $z \notin K$. By maximality of $\mathcal{X}$ it must be that K has empty intersection with $\bigcap \mathcal{X}^X$. Under the assumption that no element $y \in \bigcap \mathcal{Y}^Y$ is in $Y \backslash X$, the latter fact entails that $\bigcap \mathcal{Y}^Y$ must be empty, contradiction. Hence there can be no element z that is in $\bigcap \mathcal{X}^X$ but not in $\bigcap \mathcal{Y}^Y$, thus $\bigcap \mathcal{X}^X = \bigcap \mathcal{Y}^Y$. □

5.5.7. Proposition. *Finite evidence models are conditional models, where*

$$f(w, X) = \bigcup \{ \bigcap \mathcal{X}^X | \textit{for} \quad \mathcal{X} \quad w\text{-}X\textit{-scenario} \}$$

Proof:
The satisfaction of the first property of selection functions is ensured by the definition of $\mathcal{X}^X$: since each $\bigcap \mathcal{X}^X$ lies within X, the union will also be contained in X. For the second property suppose $Y \supseteq X$ and $f(w, Y) \subseteq X$. If $x \in f(w, Y)$ then there is a w-Y-scenario $\mathcal{Y}$ such that $x \in \bigcap \mathcal{Y}^Y$. By Lemma 5.5.5 we know there is a w-X-scenario $\mathcal{X}$ such that $\mathcal{X} \subseteq \mathcal{Y}$. By Lemma 5.5.6 either $x \in \bigcap \mathcal{X}^X$ or $x \in Y \backslash X$. But the latter cannot be because $x \in X$ by assumption, so $x \in \bigcap \mathcal{X}^X$. Then we can conclude that $x \in f(w, X)$.

Now for the other direction. If $x \in f(w, X)$ then there is a w-X-scenario $\mathcal{X}$ such that $x \in \bigcap \mathcal{X}^X$. By Lemma 5.5.5 there is a w-Y-scenario $\mathcal{Y}$ such that $\mathcal{X} \subseteq \mathcal{Y}$. Because $f(w, Y) \subseteq X$ we can infer that there is no element $y \in \bigcap \mathcal{Y}^Y$ that is in $Y \backslash X$ (that is, $\bigcap \mathcal{Y}^Y \subseteq X$), so by the second part of Lemma 5.5.6 we can conclude that $\bigcap \mathcal{X}^X = \bigcap \mathcal{Y}^Y$. This gives us $x \in \bigcap \mathcal{Y}^Y$ and thus $x \in f(w, Y)$. □

Notice that, setting $f(w, X) = \bigcup \{ \bigcap \mathcal{X}^X | \text{for} \quad \mathcal{X} \quad w\text{-}W\text{-scenario} \}$, the definition of the satisfaction relation for conditional belief on evidence models becomes an instance of the satisfaction relation for conditional modalities given in Section 5.3. Replacing the new f in Definition 5.3.5, we obtain a new notion of bisimulation for conditional belief on evidence models.

5.5.8. Definition. Given two evidence models $\mathcal{M}_1$ and $\mathcal{M}_2$, an *evidence CB-bisimulation* is a non-empty relation $Z \subseteq W_1 \times W_2$ s. t. if $(w, w') \in Z$ then

- $V(w) = V(w')$,
- for all $X \subseteq W_1$ and $X' \subseteq W_2$ such that $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$ we have that for every w-X-scenario $\mathcal{X}$ and $x \in \bigcap \mathcal{X}^X$ there is a w'-X'-scenario $\mathcal{Y}$ and $y \in \bigcap \mathcal{Y}^{X'}$ such that $(x, y) \in Z$, and vice versa.

We can now specialize the results of Section 5.3: bisimilarity in the latter sense corresponds to modal equivalence on finite evidence models.

5.5.9. THEOREM. *Given two evidence models $\mathcal{M}_1$ and $\mathcal{M}_2$ if $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is an evidence CB-bisimulation, then w and w' are modally equivalent. On finite evidence models, if w and w' are modally equivalent then $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is an evidence CB-bisimulation.*

As for plausibility models, we can infer the results concerning the closure under union and relational composition. Also in this context the definition of f renders the notation W^w trivial.

5.5.10. LEMMA. *For every evidence model $\mathcal{M}$ and $x \in W$, $f(w, \{x\}) = \{x\}$.*

Proof:
Given an evidence model $\mathcal{M}$, we check that $f(w, \{x\}) \neq \emptyset$ for all $w, x \in W$, for $f(w, X) = \bigcup\{\bigcap \mathcal{X}^X | \text{for} \quad \mathcal{X} \quad w\text{-}X\text{-scenario}\}$. The claim then follows from the condition $f(w, \{x\}) \subseteq \{x\}$. It is enough to show that there exist a w-$\{x\}$-scenario $\mathcal{X}$, then by the f.i.p. relative to $\{x\}$ we know that every element of $\mathcal{X}$ must contain x, thus $x \in \bigcap \mathcal{X}^{\{x\}}$ and $f(w, \{x\})$ is not empty. To find the desired w-$\{x\}$-scenario $\mathcal{X}$, take the family of all the sets in $E(w)$ containing x. This family is non-empty, since $W \in E(w)$ for every w in the domain of the model. Clearly this family is maximal with the f.i.p. relative to $\{x\}$ (not only, it is the only one). □
We can thus derive that, for this particular choice of f, $W^w = \bigcup_{Y \subseteq W} f(w, Y) = \bigcup_{\{x\} \subseteq W} f(w, \{x\}) = W$. In other words, all the worlds in the model are relevant for every $w \in W$.

5.5.11. LEMMA. *Every evidence model is a grounded conditional model and every evidence CB-bisimulation is diffuse.*

Proof:
Thanks to the previous Lemma we can appeal to Lemma 5.3.12 and conclude that evidence models are grounded. For the second part of the claim, let $Z \subseteq W_1 \times W_2$ be a evidence CB-bisimulation and $x \in W_1$. Since the bisimulation is non-empty, there are $(w, w') \in Z$ and furthermore $x \in W_1 = W_1^w$. The other direction is symmetric. □

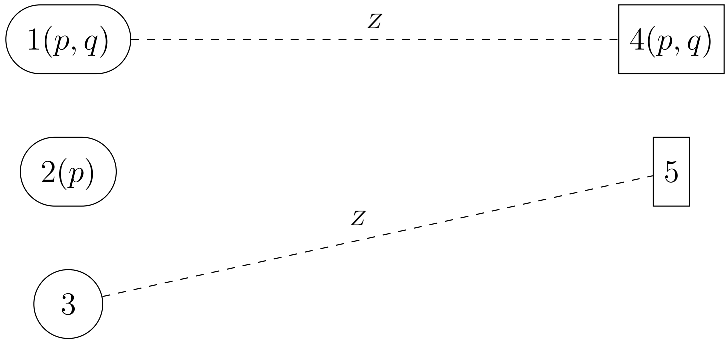
5.5.12. PROPOSITION. *The notion of evidence CB-bisimulation is closed under arbitrary unions and relational composition.*

5.5.2 Undefinability

Thanks to the now clearly defined bisimulation for conditional belief, we can give an argument for the undefinability of conditional belief in terms of plain belief.

5.5.13. PROPOSITION. *On evidence models, conditional belief is not definable in terms of the standard belief operator.*

Proof:
Suppose $B^p q$ is definable by a formula α in the language of standard belief. Consider the two models depicted on the left and right side of the following picture, where we indicate within parenthesis the propositional atoms that are true at every world and with Z an evidence B-bisimulation between the models:



The evidence available at each world is: $E(1) = \{\{1\}, \{3\}, \{2, 3\}, W_1\}$, $E(4) = \{\{4\}, \{5\}, W_2\}$, $E(2) = \{W_1\}$, $E(3) = \{\{3\}, W_1\}$, $E(5) = \{\{5\}, W_2\}$. The reader can check that the relation Z is an evidence B-bisimulation. Given that α is a formula in the language of belief, it will be invariant between states that are bisimilar according to a B-bisimulation. However, $B^p q$ is true in the second model at 4 but false in the first model at 1: there is a 1-$[\![p]\!]_{\mathcal{M}_1}$-scenario $\mathcal{X} = \{\{2, 3\}, W_1\}$ and $2 \in \bigcap \mathcal{X}^{[\![p]\!]_{\mathcal{M}_1}}$ such that $2 \not\models q$. Hence we obtain a contradiction. $\Box$

Note that the relation Z is not an evidence CB-bisimulation: the sets of worlds satisfying p in the two models satisfy the prerequisites, they are sent into each other by Z, but fail with respect to the main property, since there is a 1-$[\![p]\!]_{\mathcal{M}_1}$-scenario $\mathcal{X}$, and an element in $\bigcap \mathcal{X}^{[\![p]\!]_{\mathcal{M}_1}}$, namely 2, that has no bisimilar counterpart in the second model.

Another important operator to describe the features of evidence models is the so-called evidence modality [29].

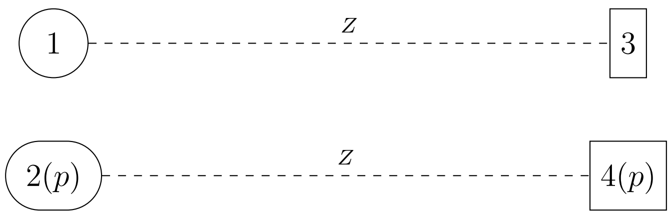
Evidence modality: $\mathcal{M}, w \models \Box\varphi$ iff there is $K \in E(w)$ such that, for all $v \in K$, $\mathcal{M}, v \models \varphi$.

It was shown in [29] that, on evidence models, standard belief cannot be defined in terms of the evidence modality. Since standard belief is definable in terms of conditional belief, we can conclude that also conditional belief is not definable via the evidence modality. Here we show that also the converse is the case.

5.5.14. PROPOSITION. *On evidence models, the evidence modality is not definable in terms of the conditional belief operator.*

Proof:
Suppose $\Box p$ is definable by a formula α in the language of conditional belief.

Consider the two models depicted on the left and right side of the following picture, where we indicate within parenthesis the propositional atoms that are true at every world and with Z a CB-bisimulation between the two models:



We take both models to be uniform, where $E_1 = \{\{1\}, \{2\}, W_1\}$ and $E_2 = \{W_2\}$. The reader can check that with this evidence the relation Z is a CB-bisimulation. Given that α is a formula in the language of normal belief, it will be invariant between states that are bisimilar according to a CB-bisimulation. Nevertheless, $\Box p$ is true in the first model at 1 but false in the second model at 3: in the first model there is an evidence set contained in the extension of p, namely $\{2\}$, while there is no such set in the second model; contradiction. $\square$

5.6 Relativized common knowledge

We now introduce a third example, the conditional modality known as relativized common knowledge, defined in [28]. Let $\mathcal{M} = \langle W, \{R_a\}_{a \in A}, V \rangle$ be a multi-agent Kripke model, where W is a non-empty set of worlds, each $R_a \subseteq W \times W$ is a relation and $V : W \to \wp(At)$ is a valuation function. Put $R := \bigcup_{a \in A} R_a$ and denote by R^+ its transitive closure. The operator of relativized common knowledge, denoted with $C(\varphi, \psi)$, is meant to capture the intuition that every R-path which consists exclusively of φ-worlds ends in a world satisfying ψ. Formally:

$$\mathcal{M}, w \vDash C(\varphi, \psi) \text{ iff } \mathcal{M}, v \vDash \psi \text{ for all } (w, v) \in (R \cap (W \times [\![\varphi]\!]))^+$$

5.6.1. Proposition. *Every Kripke model $\mathcal{M} = \langle W, \{R_a\}_{a \in A}, V \rangle$ can be converted into a conditional model, by taking $f(w, X) := \{v | (w, v) \in (R \cap (W \times X))^+\}$. Moreover, our semantics for conditionals for this f coincides with the above semantics for $C(\varphi, \psi)$.*

Proof:
Again we check the prerequisites of selection functions in Definition 5.3.1. Clearly all the worlds reachable with a path in X will also lie in X, hence the first condition on selection functions is given. For the second one, suppose $X \subseteq Y$, $f(w, Y) = \{v | (w, v) \in (R \cap (W \times Y))^+\} \subseteq X$ and take $x' \in f(w, Y)$. Hence there is a chain of Y-worlds leading to x'. We show $x' \in f(w, X)$ by induction on the length of the chain. The base case: if $(w, x') \in R$ then $x \in X$ by assumption

and therefore $(w, x') \in (R \cap (W \times X))^+ = f(w, X)$. Suppose now $x \in f(w, X)$ is the case for all $x \in f(w, Y)$ reachable with a chain of Y-worlds of length $\leqslant n$. Now say $x' \in f(w, Y)$ is reachable with a chain of Y-worlds of length $n + 1$. By $x' \in f(w, Y) \subseteq X$ we know that also $x' \in X$, thus the whole chain is in X and $x' \in f(w, X)$. For the other inclusion, it is straightforward to see that $X \subseteq Y$ immediately entails $f(w, X) \subseteq f(w, Y)$. □

Replacing the new f in Definition 5.3.5, we obtain a new notion of bisimulation for relativized common knowledge.

5.6.2. Definition. Given two Kripke models $\mathcal{M}_1$ and $\mathcal{M}_2$, a *bisimulation for relativized common knowledge* or *RCK-bisimulation* is a non-empty relation $Z \subseteq W_1 \times W_2$ such that if $(w, w') \in Z$ then

- $V(w) = V(w')$,
- for all $X \subseteq W_1$ and $X' \subseteq W_2$ such that $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$ we have that for every x such that $(w, x) \in (R_1 \cap (W_1 \times X))^+$ there exists a y such that $(w', y) \in (R_2 \cap (W_2 \times X'))^+$ such that $(x, y) \in Z$, and vice versa.

We can now derive our previous results for this specific setting. In this section we use 'modal equivalence' meaning with respect to the language containing only the usual propositional connectives and the relativized common knowledge operator.

5.6.3. Theorem. *Given two Kripke models $\mathcal{M}_1$ and $\mathcal{M}_2$, if $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is a RCK-bisimulation, then w and w' are modally equivalent. On finite models, if w and w' are modally equivalent then they are RCK-bisimilar.*

The closure under unions also follows. As for composition, note that the notion of relevant worlds for w, indicated with W^w, starts to play a significant part, limiting the scope of our general results. Putting together the definitions $W^w = \bigcup_{Y \subseteq W} f(w, Y)$ and $f(w, X) := \{v | (w, v) \in (R \cap (W \times X))^+\}$, W^w becomes the set of all the worlds reachable from w via an R-path (just substitute W for X in the definition of $f(w, X)$). Formally, $W^w = \{v | (w, v) \in R^+\}$. We can then characterize the grounded Kripke models.

5.6.4. Proposition. *A Kripke model $\mathcal{M}$ is grounded iff, for every $w, x \in W$, if $(w, x) \in R^+$ then there is an agent a such that $(w, x) \in R_a$.*

Proof:
Let $\mathcal{M}$ be grounded. By Lemma 5.3.12 if $x \in W^w = \{v | (w, v) \in R^+\}$ then $f(w, \{x\}) = \{v | (w, v) \in (R \cap (W \times \{x\}))^+\} \neq \emptyset$. This entails that there is an edge $(w, x) \in R$, thus there must be an agent a such that $(w, x) \in R_a$.

For the other direction, let $X \subseteq W$ and $w \in W$ and suppose $X \cap W^w \neq \emptyset$. Then there is $x \in X$ such that $(w, x) \in R^+$. By our assumption on the model we know there is an agent a such that $(w, x) \in R_a$. This is enough to conclude $(w, x) \in R$ and thus $x \in f(w, X) = \{v | (w, v) \in (R \cap (W \times X))^+\}$, therefore $f(w, X) \neq \emptyset$. $\Box$

In this context, a bisimulation $Z \subseteq W_1 \times W_2$ is diffuse if, for every $x \in W_1$, there are $w \in W_1$ and $w' \in W_2$ such that $(w, w') \in Z$ and x can be reached from w via an R-path (and vice versa).

5.6.5. Proposition. *On grounded Kripke models, diffuse bisimulations are closed under relational composition.*

5.7 Generalization to multi-agent models

We have seen how our framework covers different conditional modalities, even when the same operator is interpreted on different semantics. Now we address the question: can we extend the analysis of Section 5.4 to cover the multi-agent case? Given a set of agents A, consider the language

$$\varphi ::= \quad p \,|\, \neg\varphi \,|\, \psi \wedge \varphi \,|\, \psi \rightsquigarrow_a \varphi$$

where $\rightsquigarrow_a$ will denote the modality for agent a. This leads to an easy generalization of conditional models.

5.7.1. Definition. With the name *multi-agent conditional model* we indicate a tuple $\mathcal{M} = \langle W, A, \{f_a\}_{a \in A}, V \rangle$ with W a non-empty set of worlds, A a set of agents, $V : W \to \wp(At)$ a valuation function and for each agent $a \in A$ a selection function f_a satisfying the conditions listed in Definition 5.3.1.

The set of agents is nothing more than a set of labels for different selection functions, co-existing in the same models but essentially independent from each other. Instead of different agents, different labels could indicate different operators expressing distinct features of the models, depending on the interpretation. The semantics clause for the conditional modalities becomes:

$$\mathcal{M}, w \vDash \psi \rightsquigarrow_a \varphi \quad \text{iff} \quad f_a(w, [\![\psi]\!]_{\mathcal{M}}) \subseteq [\![\varphi]\!]_{\mathcal{M}}$$

for every $a \in A$. The bisimulation can also be relativized in the same fashion.

5.7.2. Definition. [Multi-agent Conditional Bisimulation] Given two multi-agent conditional models $\mathcal{M}_1$ and $\mathcal{M}_2$ based on the same set of agents, a *multi-agent conditional bisimulation* is a non-empty relation $Z \subseteq W_1 \times W_2$ such that if $(w, w') \in Z$ then

- $V(w) = V(w')$,

- for all $X \subseteq W_1$ and $X' \subseteq W_2$ such that $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$ we have that, for every $a \in A$, for every $x \in f_a^1(w, X)$ there exists a $y \in f_a^2(w', X')$ (where f^2's are the selection functions in $\mathcal{M}_2$) such that $(x, y) \in Z$, and vice versa.

The proofs of the following results are a straightforward generalization of the proofs of the analogous single-agent statements.

5.7.3. Theorem. *Given two multi-agent conditional models $\mathcal{M}_1$ and $\mathcal{M}_2$, if $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is a multi-agent conditional bisimulation, then w and w' are modally equivalent with respect to the logic of conditionals. On finite multi-agent conditional models, if w and w' are modally equivalent then $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is a multi-agent conditional bisimulation.*

5.7.4. Proposition. *Multi-agent conditional bisimulations are closed under arbitrary unions.*

The definitions of grounded models and diffuse bisimulation have to be generalized accordingly.

5.7.5. Definition. Define $W_a^w = \bigcup_{Y \subseteq W} f_a(w, Y)$. A multi-agent conditional model is *grounded* if, for any $X \subseteq W$ and $a \in A$, $X \cap W_a^w \neq \emptyset$ entails $f_a(w, X) \neq \emptyset$.

5.7.6. Definition. A multi-agent conditional bisimulation $Z \subseteq W_1 \times W_2$ is *diffuse* if for every $x \in W_1$ there are $a \in A$, $w \in W_1$ and $w' \in W_2$ such that $(w, w') \in Z$ and $x \in W_{1,a}^w$, and vice versa.

5.7.7. Proposition. *Restricted to any class of multi-agent grounded models, the notion of multi-agent diffuse conditional bisimulation is closed under relational composition.*

5.7.1 Multi-agent plausibility models

We now turn to our fourth and last example, meant to display how the general definitions unfold in the multi-agent case. Our structure of choice is multi-agent plausibility models, a popular device used to model the knowledge and beliefs of different agents [20].

5.7.8. Definition. A *multi-agent plausibility model* is a tuple $\mathcal{M} = \langle W, A, \{\leqslant_{a,w}, \sim_a\}_{a \in A, w \in W}, V\rangle$ with W a non-empty set of worlds, $\{\leqslant_{a,w}\}_{a \in A, w \in W}$ a family of reflexive and transitive relations $\leqslant_{a,w} \subseteq W \times W$ indexed by agents and worlds, $\{\sim_a\}_{a \in A}$ a family of "epistemic" equivalence relations $\sim_a \subseteq W \times W$ satisfying $\leqslant_{a,w} \subseteq \sim_a$ for every w, and $V : W \to \wp(At)$ a valuation function. A multi-agent plausibility model is *well-founded* if each relation $\leqslant_{a,w}$ is well-founded, i.e. $\forall X \subseteq W (X \neq \emptyset \Rightarrow Min_{\leqslant_{a,w}} X \neq \emptyset)$, where $Min_{\leqslant_{a,w}} X$ is the set of $\leqslant_{a,w}$-minimal elements of X. For every $w \in W$ we write $[w]_{\sim_a} := \{v \in W | w \sim_a v\}$ for the associated equivalence class.

The semantics of the multi-agent belief and conditional belief operators on (well-founded) multi-agent plausibility models is given by:

- $\mathcal{M}, w \vDash B_a\varphi$ iff for all $v \in Min_{\leqslant_{a,w}}([w]_{\sim_a})$ we have $\mathcal{M}, v \vDash \varphi$
- $\mathcal{M}, w \vDash B_a^{\psi}\varphi$ iff for all $v \in Min_{\leqslant_{a,w}}(\llbracket\psi\rrbracket \cap [w]_{\sim_a})$ we have $\mathcal{M}, v \vDash \varphi$

5.7.9. Proposition. *Well-founded multi-agent plausibility models are multi-agent conditional models, where* $f_a(w, X) = Min_{\leqslant_{a,w}}(X \cap [w]_{\sim_a})$.

Proof:
We want to ascertain that the newly defined f_a fulfills the prerequisites of selection functions in Definition 5.3.1. The first condition is again given by definition. For the second one, suppose $X \subseteq Y$, $Min_{\leqslant_{a,w}}(Y \cap [w]_{\sim_a}) \subseteq X$ and consider a generic element x' in $Min_{\leqslant_{a,w}}(Y \cap [w]_{\sim_a})$. Clearly from $X \subseteq Y$ we have $X \cap [w]_{\sim_a} \subseteq Y \cap [w]_{\sim_a}$. Since $x' \in X$ and there is no element below x' in $Y \cap [w]_{\sim_a}$, a fortiori there is no element below it in the subset $X \cap [w]_{\sim_a}$, hence $x' \in Min_{\leqslant_{a,w}}(X \cap [w]_{\sim_a})$. For the other inclusion take $x' \in Min_{\leqslant_{a,w}}(X \cap [w]_{\sim_a})$: we show x' is also minimal within $Y \cap [w]_{\sim_a}$. By contradiction suppose this is not the case: then there is $z \in Y \cap [w]_{\sim_a}$ such that $z <_{a,w} x'$. Since $\leqslant_{a,w}$ is well-founded, there must be a minimal element $z' \in Min_{\leqslant_{a,w}}(Y \cap [w]_{\sim_a})$ such that $z' \leqslant_{a,w} z$; but by assumption $Min_{\leqslant_{a,w}}(Y \cap [w]_{\sim_a}) \subseteq X$, hence $z' \in X$. This gives us a $z' \in X \cap [w]_{\sim_a}$ such that $z' <_{a,w} x'$, contradicting the fact that x' is minimal in $X \cap [w]_{\sim_a}$. □

Now that this step is secured, we can apply the general definition of multi-agent CB-bisimulation:

5.7.10. Definition. Given two multi-agent plausibility models $\mathcal{M}_1$ and $\mathcal{M}_2$, a *multi-agent plausibility CB-bisimulation* is a non-empty relation $Z \subseteq W_1 \times W_2$ such that if $(w, w') \in Z$ then

- $V(w) = V(w')$,
- for all $X \subseteq W_1$ and $X' \subseteq W_2$ such that $Z[X] \subseteq X'$ and $Z^{-1}[X'] \subseteq X$ we have, for all a, that for every $x \in Min_{\leqslant_{a,w}}(X \cap [w]_{\sim_a})$ there exists a $y \in Min_{\leqslant'_{a,w'}}(X' \cap [w']_{\sim'_a})$ such that $(x, y) \in Z$, and vice versa (where $\leqslant'_{a,w'}, \sim'_a$ are the relations associated to a in $\mathcal{M}_2$).

Therefore all our results on the correspondence between bisimilarity and modal equivalence and closure under union do carry over to this setting. For closure under composition, notice that now $W_a^w = \bigcup_{Y \subseteq W} Min_{\leqslant_{a,w}}(Y \cap [w]_{\sim_a}) = [w]_{\sim_a}$ (just replace all the singletons for Y).

5.7.11. Proposition. *Well-founded multi-agent plausibility models are grounded.*

Proof:
Let $X \cap W_a^w \neq \emptyset$; then $X \cap [w]_{\sim_a} \neq \emptyset$. Since each relation $\leqslant_{a,w}$ is well founded, there are be minimal elements in $X \cap [w]_{\sim_a}$, thus $f_a(w, X) = Min_{\leqslant_{a,w}}(X \cap [w]_{\sim_a}) \neq \emptyset$. □

In this setting a multi-agent plausibility CB-bisimulation $Z \subseteq W_1 \times W_2$ is diffuse if, for every $x \in W_1$, there are $w \in W_1$ and $w' \in W_2$ such that $(w, w') \in Z$ and x is in the information cell $[w]_{\sim_a}$ (and vice versa).

5.7.12. Proposition. *On well-founded multi-agent plausibility models, diffuse multi-agent plausibility CB-bisimulation are closed under relational composition.*

5.8 Related work

A different notion of bisimulation for conditional belief on multi-agent plausibility models was recently introduced in [10]. The authors prove the correspondence between bisimilarity and modal equivalence, respectively for the languages containing conditional belief and knowledge, safe belief and knowledge, degrees of belief and knowledge. But that analysis is confined to doxastic logic. Our approach has the following two distinctive features. First, the bisimulation for conditional belief stems from a general analysis of conditional modalities and it is not tailored to a specific application. This generality has the pleasant consequence that the key notions and proofs are relatively simple and transparent. Second, the notion of bisimulation for conditional belief offered here is modular, in the sense that it can be merged with other conditions when we consider languages with additional operators. In contrast, some results in [10] depend crucially on the existence of the knowledge operator.[10]

A notion of bisimulation containing a quantification over subsets has been proposed originally in [76], adapted in [51] to epistemic lottery models and later again reshaped to work in the context of epistemic neighborhood models in [50]. Such bisimulations were introduced to deal with probabilities and weights, not conditional modalities. The main difference with the present approach lies in the structure of the quantification. In our case the zig and zag conditions both share the same preconditions, a universal quantification over pairs of subsets satisfying certain prerequisites. In the aforementioned papers each direction has a $\forall\exists$ quantification, stating that for each subset in the first model (usually within the current information cell) there exists a subset in the second model fulfilling certain properties.

Finally, we touch on the connection with the standard Hennessy-Milner result. Such result holds for an *un*-conditional modality, namely the box operator on Kripke models. For un-conditional modalities the proof of 'modal equivalence

[10]Conversely, the undefinability result of Proposition 5.4.10 does not hold if we take knowledge into account, that is, we restrict the scope of belief to the current information cell.

entails bisimilarity' simplifies considerably: it carries through with the usual technique just by assuming the finiteness of $f(w, X)$ for all w. When $f(w, X) = \{v|wRv\}$, where R is the relation of the Kripke model, we obtain a conditional model for the box operator; in this circumstance the finiteness of $f(w, X)$ for all w corresponds to the condition known as 'finitely branching'.

5.9 Conclusions

In this chapter we proposed a general notion of bisimulation for conditional modalities interpreted on selection functions and proved some general results including a Hennessy-Milner theorem. This framework was applied to a series of examples. We showed how to derive a solid notion of bisimulation for conditional belief on both plausibility and evidence models. We exploited these notions to obtain some new undefinability results. A similar analysis was applied to the operator of relativized common knowledge. Finally, we described how to bring the whole approach into the realm of multi-agent models, generalizing the main framework and discussing its application to multi-agent plausibility models.

The attention towards such bisimulations and their closure under relational composition arose from the interest in seeing these classes of models as categories in which the bisimulations play the role of arrows. The desideratum of closure under composition is also aligned with the usual requirement of having a relation of bisimilarity that is an equivalence relation. This categorical perspective is further developed in the next chapter, where we investigate its bearing on other themes such as the existence of reduction laws for dynamic operators.

Along these lines, a further direction of research concerns the extension of our findings on the closure under relational composition. Our results could be strengthened on specific examples, where the selection functions may enjoy additional properties (e.g., the selection function for relativized common knowledge is fully monotonic). At the general level, one may want to characterize exactly those notions of bisimulations that are closed under relational compositions, for suitably large classes of models and operators.

A second open question regards infinite models: does modal equivalence entail bisimilarity on some natural class of infinite conditional models? We may furthermore ask how many 'classical' results of the model theory for basic modal logic we can obtain in the setting of conditional modalities. One natural example would be a version of the van Benthem characterization theorem.

Chapter 6

Tracking

6.1 Introduction

In this chapter we continue to study models of modal languages from a categorical standpoint, focusing on the case-study of plausibility and evidence models. We first explain how these classes of models can be arranged into different categories by means of different choices of bisimulations. Since notions of bisimulations are linked to certain modal languages, we can think of picking a notion of bisimulation as if selecting a language 'through which' we look at the models.

Regarding a class of models as a category whose arrows are bisimulations allows us to recast some known concepts and problems in categorical terms. An important notion in Modal Logic is that of update, namely a model-changing operation that occurs after the model is fed with new information. Requiring an update to be functorial, for these particular categories of models, means to ask (among other things) whether bisimilar models are mapped to bisimilar models. This suggests a link between functoriality of an update and the existence of reduction laws for the associated dynamic operator.

Another theme is the relationship between classes of models. A mapping between two different classes of models can have different properties when the classes are regarded as categories. For some choices of languages (read: bisimulations) such mapping will not be functorial, while for other languages the mapping will turn out to give a categorical equivalence.

Finally, the combination of these issues is composed in the problem known as tracking. We say that an update U' at the level of evidence models is 'tracked' by another update U at the level of plausibility models if updating an evidence model M with U' and then turning it into a plausibility model is the same as taking the plausibility model corresponding to M and then applying U. This matching of information dynamics highlights the possibility to reduce an update on a complex structure to an update on a simpler structure. When tracking occurs we are able to transfer results from the updates on the simpler setting to the updates on the richer setting. The main result in this chapter is a characterization of the trackable updates in the class of "simple" updates (to be defined later): for the updates that fit the description we provide a procedure to construct the corresponding update on plausibility models; for the updates that do not meet the requirements we describe how to build a counterexample to tracking.

As for the internal organization of the chapter, in Section 6.2 we describe how to arrange evidence and plausibility models into categories. Section 6.3 discusses the opportunity of regarding updates as endofunctors and considers a few examples on both plausibility and evidence models. Section 6.4 expands on the connection between the categories of plausibility models and evidence models. Finally, Section 6.5 introduces the concept of tracking, reviews some existing results and links them to the observations of the previous sections. The last result of the chapter is a characterization of the trackable updates in the class of simple updates. We conclude in Section 6.7.

6.2 Plausibility models and evidence models as categories

We begin our analysis from plausibility models, introduced in the previous chapter. For the present chapter we will only be interested in *uniform* plausibility models, namely plausibility models where the relation $\leqslant_w$ is the same for all worlds w; we will thus drop the subscript and just write $\leqslant$. Since we are interested in exploring the effects of different notions of bisimulations, we consider an array of different languages from the onset.

6.2.1 The categories of plausibility models

We consider different operators to express the features of plausibility models:

(1) belief $B\varphi$;

(2) conditional belief $B^{\psi}\varphi$;

(3) plausibility modality $[\leqslant]\varphi$;

(4) global modality $A\varphi$.

For $\emptyset \neq K \subseteq \{B, B^{\psi}, [\leqslant], A\}$, call $\mathcal{L}_K$ the propositional language enriched with the operators contained in K.

The aforementioned operators adhere to the following satisfaction clauses:

(1) $\mathcal{M}, w \vDash B\varphi$ iff for all $v \in Min(W)$ we have $\mathcal{M}, v \vDash \varphi$.

(2) $\mathcal{M}, w \vDash B^{\psi}\varphi$ iff for all $v \in Min(\llbracket\psi\rrbracket)$ we have $\mathcal{M}, v \vDash \varphi$.

(3) $\mathcal{M}, w \vDash [\leqslant]\varphi$ iff for all v if $v \leqslant w$ then $\mathcal{M}, v \vDash \varphi$.

(4) $\mathcal{M}, w \vDash A\varphi$ iff for all $v \in W$ $\mathcal{M}, v \vDash \varphi$.

The operator $B\varphi$ is a special case of $B^{\psi}\varphi$ when $\psi = \top$. To these different languages correspond different notions of bisimulation for plausibility models. We match the condition with the same number of the operator associated to it.

6.2.1. Definition. Given two plausibility models $\mathcal{M}_1$ and $\mathcal{M}_2$, a *plausibility bisimulation*^{K}, with $\emptyset \neq K \subseteq \{B, B^{\psi}, [\leqslant], A\}$, is a non-empty relation $Z \subseteq W_1 \times W_2$ such that if $(w, w') \in Z$ then $p \in V(w)$ iff $p \in V(w')$, and in addition satisfying one or more of the following properties, depending on the operators in K. If $(w, w') \in Z$:

(1) For every $x \in MinW_1$ there is $y \in MinW_2$ such that $(x, y) \in Z$, and vice versa.

(2) If two sets $X \subseteq W_1$ and $X' \subseteq W_2$ satisfy the following properties

- $Z[X] \subseteq X'$,
- $Z^{-1}[X'] \subseteq X$

then for every $x \in MinX$ there exists a $y \in MinX'$ such that $(x, y) \in Z$, and vice versa.

(3) If $(w, w') \in Z$ then for every $v \leqslant w$ there is $v' \in W_2$ such that $v' \leqslant w'$ and $(v, v') \in Z$, and vice versa.

(4) The relation Z must be total: for every $x \in W_1$ there is $y \in W_2$ such that $(x, y) \in Z$, and vice versa.

Thus, for example, if we are considering a bisimulation satisfying the first and third conditions we write plausibility bisimulation$^{B,[\leqslant]}$. We use the same superscripts to label the notion of plausibility bisimilarity.

Note that condition 2 entails condition 1 and 4; the latter is a consequence of Lemmas 5.3.15 and 5.4.8 from the previous chapter.

6.2.2. Theorem (Bisimilarity matches modal equivalence). *Given two plausibility models $\mathcal{M}_1$ and $\mathcal{M}_2$ and $\emptyset \neq K \subseteq \{B, B^{\psi}, [\leqslant], A\}$, if $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is a plausibility bisimulation^K, then w and w' are modally equivalent with respect to the language $\mathcal{L}_K$.*

On finite models, if w and w' are modally equivalent with respect to the language $\mathcal{L}_K$ then $(w, w') \in Z \subseteq W_1 \times W_2$ for some relation Z which is a plausibility bisimulation^K.

Proof:
The first part of the claim is proved as usual with an induction on the complexity of the formula, where each of the conditions from 1 to 5 takes care of the operator with the corresponding number.

The second part is proved with the standard technique: assuming two states are modally equivalent with respect to the operator $i \in \{B, B^{\psi}, [\leqslant], A\}$, we show that the relation of modal equivalence is a plausibility bisimulation satisfying condition corresponding to i.

Be begin from standard belief. For a contradiction, assume that the relation of modal equivalence violates condition 1 of Definition 6.2.1: suppose there is a minimal world x in the first model that is not modally equivalent to any $y \in MinW_2$. Since there are finitely many worlds we can enumerate such y's with a finite set I, thus we know there must be finitely many formulas ψ_i that are false at x and are true at some y_i. The formula $B \bigvee_{i \in I} \psi_i$ is then true at w' and false at w, delivering a contradiction. The claim for conditional belief was proven in the previous chapter in Theorem 6.2.2. The argument for the plausibility modality is just the usual Hennessy-Milner argument (finite models are image-finite). For the universal modality we have a similar argument: if condition 4 fails then there is a world $x \in W_1$ that is not related to any world $y \in W_2$. Since the y's are finitely many, we can again construct $\bigvee_i \psi_i$ such that $w \vDash \neg A \bigvee_i \psi_i$ and $w' \vDash A \bigvee_i \psi_i$, obtaining a contradiction. □

We now arrange the class of plausibility models into a category, using bisimulations as arrows. By choosing a subset of conditions $\emptyset \neq K \subseteq \{B, B^{\psi}, [\leqslant], A\}$, we have at our disposal $2^4 - 1 = 15$ different kinds of bisimulations and thus virtually 15 different categories.

First we check that these relations fulfill the basic requirements for a category. The notion of composition we use is relational composition. It is straightforward to check that the identity relation fulfills conditions 1 to 4 and can therefore act as the identity arrow in each of such categories.

6.2.3. Proposition. *For any $\emptyset \neq K \subseteq \{B, B^{\psi}, [\leqslant], A\}$, the evidence bisimulation^K is closed under relational composition.*

Proof:
Rather straightforward check for all conditions beside 2, for which the claim was proven in the previous chapter in Proposition 5.4.9. □

6.2.4. Definition. Call $\mathbf{PM}_K$ the category having as objects plausibility models and as arrows plausibility bisimulations^K, where $\emptyset \neq K \subseteq \{B, B^{\psi}, [\leqslant], A\}$.

We think of the category $\mathbf{PM}_K$ as the collection of evidence models looked through the modal language enriched with the operators in K.

6.2.2 The category of evidence models

We now turn our attention to evidence models, also introduced in the previous chapter. Again we only consider *uniform* evidence models, in which the collection of available evidence is the same for all worlds w; we will thus drop the notation $E(w)$ and write $\mathcal{E}$ to denote the collection of evidence sets in a model. In order to express different features of evidence models we consider the following operators:

(1) belief $B\varphi$;

(2) conditional belief $B^{\psi}\varphi$;

(3) evidence modality $\Box\varphi$;

(4) global modality $A\varphi$;

(5) instantial neighborhood modality $\Box(\Psi, \varphi)$.

where Ψ in the last operator is a finite set of formulas $\{\psi_1, \ldots, \psi_n\}$. For $\emptyset \neq N \subseteq \{B, B^{\psi}, \Box, A, \Box(\Psi)\}$, call $\mathcal{L}_N$ the propositional language enriched with the operators in N.

The satisfaction relation for these operators is defined as follows. We refer to Definition 5.5.2 in the previous chapter for the details on scenarios:

(1) $\mathcal{M}, w \vDash B\varphi$ iff for every w-scenario $\mathcal{X}$ we have $\mathcal{M}, v \vDash \varphi$ for all $v \in \bigcap \mathcal{X}$.

(2) $\mathcal{M}, w \vDash B^{\psi}\varphi$ iff every w-$[\![\psi]\!]$-scenario $\mathcal{X}$ we have $\mathcal{M}, v \vDash \varphi$ for all $v \in \bigcap \mathcal{X}^{[\![\psi]\!]}$.

(3) $\mathcal{M}, w \vDash \Box\varphi$ iff there exists $X \in \mathcal{E}$ and for all $v \in X$ we have $\mathcal{M}, v \vDash \varphi$.

(4) $\mathcal{M}, w \vDash A\varphi$ iff for all $v \in W$ $\mathcal{M}, v \vDash \varphi$.

(5) $\mathcal{M}, w \vDash \Box(\Psi, \varphi)$ iff there exists $X \in \mathcal{E}$ such that: for all $\psi_i \in \Psi$ there exist $v_i \in X$ satisfying $\mathcal{M}, v_i \vDash \psi_i$; for all $v \in X$ $\mathcal{M}, v \vDash \varphi$.

To these different languages correspond different notions of bisimulation for evidence models. Again we match the condition with the same number of the operator associated to it.

6.2.5. Definition. [From [29] and [27]] Given two evidence models $\mathcal{M}_1$ and $\mathcal{M}_2$, an *evidence bisimulation*^{N} is a non-empty relation $Z \subseteq W_1 \times W_2$ such that if $(w, w') \in Z$ then $p \in V(w)$ iff $p \in V(w')$, and in addition satisfying one or more of the following properties, depending on the operators in N. Suppose $(w, w') \in Z$:

(1) For every w-scenario $\mathcal{X}$ and $x \in \bigcap \mathcal{X}$ there is a w'-scenario $\mathcal{Y}$ and $y \in \bigcap \mathcal{Y}$ such that $(x, y) \in Z$, and vice versa.

(2) If two sets $X \subseteq W_1$ and $X' \subseteq W_2$ satisfy the following properties

- $Z[X] \subseteq X'$,
- $Z^{-1}[X'] \subseteq X$

then for every w-X-scenario $\mathcal{X}$ and $x \in \bigcap \mathcal{X}^X$ there is a w'-X'-scenario $\mathcal{Y}$ and $y \in \bigcap \mathcal{Y}^{X'}$ such that $(x, y) \in Z$, and vice versa.

(3) For every $X \in \mathcal{E}$ there is $Y \in \mathcal{E}'$ such that for all $y \in Y$ there is $x \in X$ such that $(x, y) \in Z$, and vice versa.

(4) The relation Z must be total.

(5) For every $X \in \mathcal{E}$ there is $Y \in \mathcal{E}'$ such that

- for all $y \in Y$ there is $x \in X$ such that $(x, y) \in Z$
- for all $x \in X$ there is $y \in Y$ such that $(x, y) \in Z$

and vice versa switching w and w' in the precondition.

we indicate the properties satisfied by a bisimulation with a superscript, e.g. writing evidence bisimulation$^{B,\Box,A}$. We use the same superscripts to label the different notions of evidence bisimilarity.

Note that:

- condition 5 entails condition 4, since $W_1 \in \mathcal{E}$ and $W_2 \in \mathcal{E}'$;
- condition 5 entails condition 3, since the consequent of the former entails the consequent of the latter;
- condition 2 entails condition 1, because the pair W_1 and W_2 fulfills the precondition of 2;
- condition 2 entails condition 4, because of Lemmas 5.3.15 and 5.5.11 in the previous chapter.

6.2.6. Theorem (Bisimilarity matches modal equivalence). *Given two evidence models $\mathcal{M}_1$ and $\mathcal{M}_2$ and $\emptyset \neq N \subseteq \{B, B^\psi, \Box, A, \Box(\Psi)\}$, if $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is an evidence bisimulation^N, then w and w' are modally equivalent with respect to the language $\mathcal{L}_N$.*

On finite evidence models, if w and w' are modally equivalent with respect to the language $\mathcal{L}_N$ then $(w, w') \in Z \subseteq W_1 \times W_2$, where Z is an evidence bisimulation^N.

Proof:
The first part of the theorem is proved modularly via an induction on the structure of the formula, where each of the conditions from 1 to 5 ensures the satisfaction of the corresponding operator.

For the second claim, assuming two states are modally equivalent with respect to the operator $i \in \{B, B^{\psi}, \Box, A, \Box(\Psi)\}$, we show that the relation of modal equivalence is a bisimulation satisfying the condition for operator i.

For standard belief the argument runs as usual. For a contradiction, suppose there is a w-scenario $\mathcal{X}$ and $x \in \bigcap \mathcal{X}$ such that, for any w'-scenario $\mathcal{Y}$, x is not modally equivalent to any $y \in \mathcal{Y}$. Since there are finitely many worlds we can enumerate such y's with a finite set I, thus we know there must be finitely many formulas ψ_i that are false at x and are true at some y_i. The formula $B \bigvee_{i \in I} \psi_i$ is then true at w' and false at w, delivering the contradiction we were after. The claim for conditional belief was proven in the previous chapter.

Suppose now that modal equivalence does not satisfy condition 3 (corresponding to $\Box$): then there is $X \in \mathcal{E}$ such that for every $Y \in \mathcal{E}'$ there exists $y \in Y$ such that for all $x \in X$ we have x and y not modally equivalent. Since the carriers are finite, there can be only finitely many Y's and only finitely many such y's and x's in X. Enumerate the elements of X by $x_1, \ldots, x_n$ and enumerate all the y's with the aforementioned property with $y_1, \ldots, y_m$. For each index $(i, j) \in N \times M$ (where $N = \{1, \ldots, n\}$ and $M = \{1, \ldots, m\}$), since $(x_i, y_j) \notin Z$, we get a formula $\psi_{i,j}$ such that $x_i \vDash \psi_{i,j}$ and $y_j \nvDash \psi_{i,j}$. We can now check that $w \vDash \Box \bigvee_i \bigwedge_j \psi_{i,j}$: there is an evidence set X such that for every element $x \in X$ there is an i such that $x \vDash \psi_{i,j}$ for all j's. But $w' \nvDash \Box \bigvee_i \bigwedge_j \psi_{i,j}$ because for every Y there is a $y_{j'}$ such that for every disjunct $\bigwedge_j \psi_{i,j}$ the formula $\psi_{i,j'}$ is false at $y_{j'}$: contradiction.

For condition 4 (operator A) we have an argument similar to that in the proof of Theorem 6.2.2. For the case of $\Box(\Psi, \varphi)$ see [27]. □

We want to arrange the class of evidence models into a category, using bisimulations as arrows.The preliminary step in this direction is to check that these relations fulfill the basic requirements for a category. The notion of composition is again relational composition. It is straightforward to ascertain that the identity relation fulfills conditions 1 to 5 and can therefore act as the identity arrow in each of such categories.

6.2.7. Proposition. *For any $\emptyset \neq N \subseteq \{B, B^{\psi}, \Box, A, \Box(\Psi)\}$, the evidence bisimulation^N is closed under relational composition.*

Proof:
We prove the statement separately for each condition. For the rest of the proof suppose $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ are three evidence models such that the first two are linked by a relation Z_1 and the latter two are connected by a relation Z_2. Assume moreover that $(w, w'') \in Z_1; Z_2$ and hence there must be w' such that $(w, w') \in Z_1$ and $(w', w'') \in Z_2$.

Suppose Z_1 and Z_2 satisfy condition 1. Given a w-scenario $\mathcal{X}$ and $x \in \bigcap \mathcal{X}$ in $\mathcal{M}_1$ we receive, by the property of Z_1, a w'-scenario $\mathcal{X}'$ and $y \in \bigcap \mathcal{X}'$ such that $(x, y) \in Z_1$. From $\mathcal{X}'$ and y, using the property of Z_2, we receive a w''-scenario $\mathcal{X}''$ and $z \in \bigcap \mathcal{X}''$ such that $(y, z) \in Z_2$. Thus $(x, z) \in Z_1; Z_2$, as desired.

The closure of property 2 was proved in the previous chapter in Proposition 5.5.12. As for property 3 (operator $\Box$), suppose $X \in \mathcal{E}$. By Z_1 satisfying the same property, we get $X' \in \mathcal{E}'$ such that for all $y \in X'$ there is $x \in X$ such that $(x, y) \in Z_1$. Using the property for Z_2 we obtain $X'' \in \mathcal{E}''$ such that for all $z \in X''$ there is $y \in X'$ such that $(y, z) \in Z_2$. Thus for all $z \in X''$ there is $x \in X$ such that $(x, z) \in Z_1; Z_2$.

The closure of property 4 (operator A) is straightforward. The claim for property 5 is proven along the same lines as for property 3. $\Box$

6.2.8. Definition. Call $\mathbf{EM}_N$ the category having as objects evidence models and as arrows evidence bisimulations^N, where $\emptyset \neq N \subseteq \{B, B^{\psi}, \Box, A, \Box(\Psi)\}$.

As for plausibility models, we think of the category $\mathbf{EM}_N$ as the collection of evidence models looked through the lenses of the modal language enriched with the operators in N.

6.3 Dynamic updates

Given a class of models $\mathbf{M}$ that represent some epistemic or doxastic features of an agent - such as the class of plausibility models or other kinds of models presented in the previous chapter - one may want to investigate what happens to such models when the agent is presented with new information. The transformation of the models is formalized via what is known as an 'update', namely a procedure to construct a new model in light of the new piece of information, typically encoded in a formula of a language $\mathcal{L}$. An *update* is a therefore a function $U : \mathcal{L} \times \mathbf{M} \to \mathbf{M}$, combining a model with new data and constructing a new model. We assume updates do not increase the size of the models.

One can encode the effect of an update U into an operator, usually called 'dynamic' to highlight its association with a model-changing operation (see e.g. [19]). For every two formulas $\varphi, \psi \in \mathcal{L}$ the satisfaction relation of the dynamic operator $[U(\varphi)]$ is schematically defined as

$$\mathcal{M}, w \vDash [U(\varphi)]\psi \quad \text{iff} \quad \textit{condition}(\mathcal{M}, w, \varphi) \quad \text{entails} \quad U(\varphi)(\mathcal{M}), w \vDash \psi$$

where $U(\varphi)(\mathcal{M})$ is the updated model and *condition* is a precondition involving $\mathcal{M}, w$ and φ.

We have seen that, insofar as plausibility and evidence models are concerned, such classes of models can be arranged into categories. With this observation in place then it is natural to wonder whether the functions $U(\varphi) : \mathbf{M} \to \mathbf{M}$, with $\varphi \in \mathcal{L}$, also qualify as (endo)functors. In other words, a categorical perspective raises the question: how do updates behave with respect to bisimulations?

As far as multi-agent Kripke models are concerned, an answer to this question can be traced back at least to [19] and [15]. These authors were interested in updates that preserve bisimulations, where bisimulation-preservation was defined in terms of so-called ‘transition relations’ connecting a model to its updated version. Bisimulation-preservation, which was taken to be the defining feature of an update by these authors, entails functoriality (see [15]). While the notion of bisimulation in these papers is fixed - it is the standard Aczel-Mendler notion, here presented as condition 3 on plausibility bisimulations - we are interested in studying how different kinds of bisimulations interact with different updates.

Interestingly, for several known updates we witness a *failure* of functoriality; this in turn has repercussions on the definability of the associated dynamic operators. Some updates may fail to be functorial because of how one defines their action on arrows. However, the effect of some updates can be much more radical.

6.3.1. DEFINITION. An update U *breaks bisimilarity* if there are models M_1 and M_2 linked by a bisimulation Z and there are two worlds w and w' such that

- $(w, w') \in Z$,
- $(w, w') \in U(W_1) \times U(W_2)$,
- *condition*(M_1, w, φ) and *condition*(M_2, w', φ) both hold

but the respective updated models cannot have *any* bisimulation between them.

In other words, U breaks bisimilarity when there are two bisimilar points that survive the updates but cannot be bisimilar in the updated models.[1] When this happens to an update, regardless of how we define its action on arrows, the update can never be functorial on the given category of models. We can exploit this observation to show that a certain dynamic operator cannot be reduced to the static language.

6.3.2. PROPOSITION. *Consider a category of models* $\mathbf{M}$ *where the arrows are the bisimulations for a language* $\mathcal{L}$*, in the sense that bisimilarity corresponds to modal equivalence on finite models. Suppose an update* $U(\varphi) : \mathbf{M} \to \mathbf{M}$ *breaks the bisimilarity between two finite models* M_1 *and* M_2 *connected by a bisimulation* Z*. Then there can be no reduction laws to express the associated dynamic modality in terms of the static language* $\mathcal{L}$.

Proof:
If $U(\varphi)$ breaks bisimilarity on finite models then there is a bisimulation Z between two finite models M_1 and M_2 such that $(w, w') \in Z$ and moreover there can be no bisimulation between the updated models $U(\varphi)(M_1)$ and $U(\varphi)(M_2)$. Since

[1] This definition can be understood as the opposite of bisimulation-preservation in [19].

$U(\varphi)(M_1)$ and $U(\varphi)(M_2)$ remain finite, by the correspondence between bisimulation and modal equivalence if no pair of worlds from $U(\varphi)(M_1)$ and $U(\varphi)(M_2)$ are bisimilar then we can construct a formula $\psi \in \mathcal{L}$ that is true at every world in one model and false at every world in the second model.

Now suppose there were reduction laws. Then we would be able to express the formula $[U(\varphi)]\psi$ as a formula $\alpha \in \mathcal{L}$. Because of the bisimulation Z, α will have the same truth value in worlds that are related by Z. Now observe that $M_1, w \vDash \alpha$, since by assumption $condition(M_1, w, \varphi)$ holds and $U(\varphi)(M_1), w \vDash \psi$. On the other hand $M_2, w' \nvDash \alpha$, because $condition(M_2, w', \varphi)$ holds but $U(\varphi)(M_2), w' \nvDash \psi$, contradicting the fact that bisimilarity entails modal equivalence. □

Following the blueprint of this observation we are able to exclude the existence of reduction laws for a number of updates. In the next subsections we also provide some functoriality results. To define the action of an update U on bisimulations, *qua* arrows of the category, we take a uniform approach and define $U(\varphi)(Z) = Z \cap U(W_1) \times U(W_2)$. We will say that an update is functorial if all its components $U(\varphi)$ are.

6.3.1 Updates on plausibility bisimulations

We turn our attention to the updates on plausibility models, checking whether they can be made into functors on the category $\mathbf{PM}_K$. The answer will vary depending on the subset $\emptyset \neq K \subseteq \{B, B^\psi, [\leqslant], A\}$, which indicates what kind of bisimulation we are considering and thus which static language is in the background.

Public announcement

When an agent receives information from an extremely reliable source, e.g. from direct observation, she may want to discard all the possible worlds that do not conform with the new information. This kind of update is generally known as public announcement.

6.3.3. Definition. [Public announcement, [106]] Given a formula $\varphi \in \mathcal{L}_K$, a *public announcement of* φ is a construction of type $\varphi! : \mathbf{PM}_K \to \mathbf{PM}_K$. For a plausibility model $M = \langle W, \leqslant, V\rangle$ the action on objects $\varphi!(M) = M^{\varphi!}$ is defined as:

- $W^{\varphi!} = W \cap [\![\varphi]\!]$
- $V^{\varphi!}(p) = V(p) \cap [\![\varphi]\!]$
- $\leqslant^{\varphi!}(w) = \leqslant \cap [\![\varphi]\!] \times [\![\varphi]\!]$

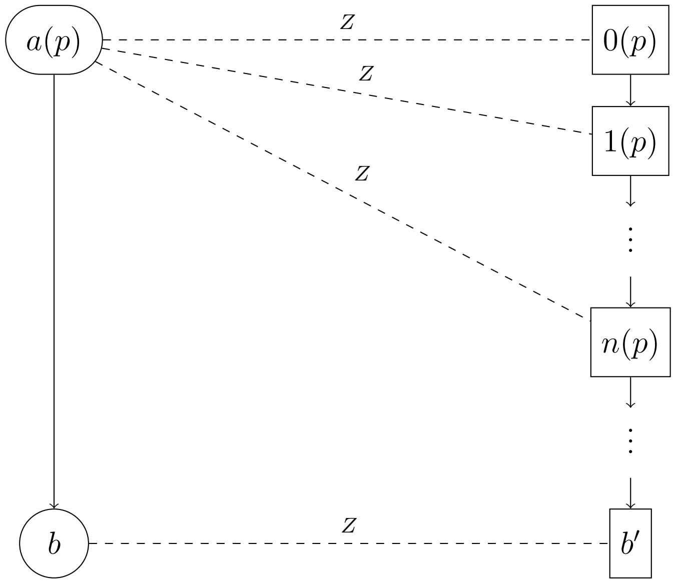
Given a plausibility bisimulation $Z \subseteq M_1 \times M_2$, define $\varphi!(Z) = Z \cap [\![\varphi]\!]_{M_1} \times [\![\varphi]\!]_{M_2}$.

The dynamic operator associated to the public announcement of φ is commonly denoted with $[\varphi!]$ and its semantics is

$$M, w \vDash [\varphi!]\psi \quad \text{iff} \quad M, w \vDash \varphi \quad \text{entails} \quad M^{\varphi!}, w \vDash \psi$$

6.3.4. Lemma. *Public announcement breaks plausibility bisimilarity*^{K} *for* $K = \{B\}, \{B, A\}$, $\{B, [\leqslant]\}$ *and* $\{B, [\leqslant], A\}$.

Proof:
Consider the two models depicted on the left- and right-hand side of the image below and connected by a relation Z, where arrows stand for $\leqslant$ and we omit reflexive and transitive arrows.



On the left we have a model with two points, where $b \leqslant a$ and only a satisfies p. On the right we have a model with a copy of the natural numbers plus an additional element b' such that $\forall n\, n+1 \leqslant n$, $\forall n\, b' \leqslant n$ and b' is the only point not satisfying p. For every n we have $(a, n) \in Z$, while b' is connected to b.

It is immediate to check that Z is a bisimulation$^{B,[\leqslant],A}$: it is total, it satisfies the 'back and forth' of condition 3 and it links together the minimal worlds. In particular this means that Z is also a bisimulation^K for for $K = \{B\}, \{B, A\}$, $\{B, [\leqslant]\}$. However, after announcing $p!$ the minimal worlds b and b' are removed. Thus on the left we have a single point a and on the right we have an infinite descending chain of natural numbers.[2] These two models cannot be linked by a bisimulation satisfying condition 1, because there is a minimal element on the left

[2] Clearly this counterexample does not work if we define plausibility models to be well-founded. In that setting, condition 3 and 4 entail condition 1 on bisimulation and thus bisimilarity$^{B,[\leqslant],A}$ is preserved; nevertheless, one can still find counterexamples to show that bisimilarity is broken in the cases $\{B\}, \{B, A\}$ and $\{B, [\leqslant]\}$.

and no minimal elements on the right. Indeed the formula $B\perp$ is now true in the right-hand model and false in the left-hand one. □

6.3.5. Proposition. *On plausibility models, the operator of public announcement cannot be reduced to any static language consisting of the propositional language enriched with one of the following sets of operators*

- *plain belief,*
- *plain belief and universal modality,*
- *plain belief and plausibility modality,*
- *plain belief, universal modality and plausibility modality.*

Proof:
Lemma 6.3.4 shows that public announcement breaks bisimilarity in all the aforementioned cases. Even though the countermodels are not finite there is a formula, namely $B\perp$, which is true everywhere in one updated model and false everywhere in the other updated model, therefore the proof of Proposition 6.3.2 still applies. □

6.3.6. Proposition. *The construction* $\varphi! : \mathbf{PM}_K \to \mathbf{PM}_K$ *is functorial for any* $\emptyset \neq K \subseteq \{B, B^{\psi}, [\leqslant], A\}$ *different from* $\{B\}, \{B, [\leqslant]\}, \{B, A\}$ *and* $\{B, [\leqslant], A\}$.

Proof:
Clearly $\varphi!$ sends identity relations to identity relations. Thanks to Theorem 6.2.2 it preserves relational composition: if three worlds are linked by two bisimulations then they are pairwise bisimilar, thus either they all satisfy φ or they do not, meaning that the $\varphi!$ construction either removes them all or leaves them in the models. This ensures that no pairs are removed from the composite relation; the fact that no pairs are added is given by the definition in terms of intersection.

We furthermore need to check that the construction does not break bisimulations, that is, if $Z \subseteq M_1 \times M_2$ is a bisimulation^K then $\varphi!(Z)$ is still a bisimulation^K between the models $\varphi!(M_1)$ and $\varphi!(M_2)$. For each condition we only describe the proof of one direction, the converse being proved symmetrically.

Suppose Z is a bisimulation$^{B^{\psi}}$ and suppose $(w, w') \in \varphi!(Z)$. Given two subsets $X \subseteq W_1^{\varphi!}$ and $X' \subseteq W_2^{\varphi!}$ satisfying the right precondition, consider $x \in MinX$ in $\varphi!(M_1)$. Because bisimilarity$^{B^{\psi}}$ entails modal equivalence as per Theorem 6.2.2, the subsets X and X' still satisfy the precondition also in the original models M_1 and M_2 (otherwise we would have a φ-world bisimilar to a world not satisfying φ). Furthermore, it is easy to see that every minimal element in X is also minimal in the original model M_1 (no worlds in X can be erased by the announcement because X is selected after the update has taken place). Thus by condition 2 we receive a $y \in MinX'$ such that $(x, y) \in Z$. Since x has survived the update we know that $x \vDash \varphi$, so by Theorem 6.2.2 $y \vDash \varphi$ too and therefore $(x, y) \in \varphi!(Z)$.

The preservation of condition 3 (corresponding to $[\leqslant]$) is folklore, while the preservation of the totality of the relation is immediate. The other claims follow from what we just proved and the inter-dependencies between conditions. □

Radical upgrade

When the source of the information φ is not infallible but still highly reliable, an agent may rearrange the possible worlds in order to have the φ ones ranked as more plausible than the $\neg\varphi$ ones.

6.3.7. DEFINITION. [Radical upgrade, [25]] Given a formula $\varphi \in \mathcal{L}_K$, a *radical upgrade of* φ is a construction of type $\Uparrow \varphi : \mathbf{PM}_K \to \mathbf{PM}_K$.[3] For a plausibility model $M = \langle W, \leqslant, V \rangle$ the action on objects $\Uparrow \varphi(M) = M^{\Uparrow\varphi}$ is defined as:

- $W^{\Uparrow\varphi} = W$
- $V^{\Uparrow\varphi}(p) = V(p)$
- $\leqslant^{\Uparrow\varphi} (w) = \{(v, v') | (v, v') \in \leqslant, v, v' \vDash \varphi\} \cup \{(v, v') | (v, v') \in \leqslant, v, v' \vDash \neg\varphi\} \cup \{(v, v') | v \vDash \varphi, v' \vDash \neg\varphi\}$

Given a plausibility bisimulation $Z \subseteq M_1 \times M_2$, define $\Uparrow \varphi(Z) = Z$.

The dynamic operator associated to radical upgrade of φ is commonly denoted with $[\Uparrow \varphi]$ and its semantics is

$$M, w \vDash [\Uparrow \varphi]\psi \quad \text{iff} \quad \Uparrow \varphi(M), w \vDash \psi$$

Note that *condition* here is equivalent to a tautology.

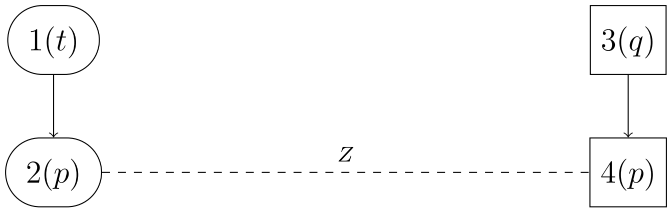
6.3.8. LEMMA. *Radical upgrade breaks bisimilarity*^{K} *for* $K = \{B\}$, $\{B, A\}$, $\{B, [\leqslant]\}$ *and* $\{B, [\leqslant], A\}$.

Proof:
Consider the first pair of model depicted in Lemma 6.3.4. The bisimulation there satisfies condition 1, 3 and 4, but after upgrading with $\Uparrow \neg p$ we have that $B\perp$ is true in one model and false in the other. □

6.3.9. LEMMA. *Radical upgrade breaks bisimulation*$^{[\leqslant]}$.

[3] This update is sometimes called 'lexicographic upgrade'.

Proof:
Consider the two models depicted in the image below and connected by a relation Z, where we omit reflexive arrows.



The bisimulation Z satisfies condition 3, corresponding to $[\leqslant]$. Updating with a radical upgrade $\Uparrow \neg p$, however, turns 1 and 3 into minimal worlds. The worlds 2 and 4 were bisimilar before the update, but after the update there can be no bisimulation$^{[\leqslant]}$ between them, since one satisfies $\langle\leqslant\rangle t$ and the other does not. □

6.3.10. PROPOSITION. *The operator of radical upgrade cannot be reduced to any static language consisting of the propositional language enriched with one of the following set of operators*

- *plain belief,*
- *plain belief and universal modality,*
- *plausibility modality,*
- *plain belief and plausibility modality,*
- *plain belief with both universal and plausibility modality.*

Proof:
The previous two Lemmas trigger Proposition 6.3.2 for radical upgrade. □

6.3.11. PROPOSITION. *The construction* $\Uparrow \varphi : \mathbf{PM}_K \to \mathbf{PM}_K$ *is functorial for any* $\emptyset \neq K \subseteq \{B, B^\psi, [\leqslant], A\}$ *different from* $\{B\}, \{[\leqslant]\}, \{B, [\leqslant]\}, \{B, A\}$ *and* $\{B, [\leqslant], A\}$.

Proof:
It takes a simple check to verify that the construction $\Uparrow \varphi$ sends identity relations to identity relations and preserves relational composition. We now ascertain that the construction preserves bisimulations.

Suppose Z is a bisimulation$^{B^\psi}$ and suppose $(w, w') \in \Uparrow \varphi(Z)$. Given two subsets $X \subseteq W_1^{\Uparrow\varphi}$ and $X' \subseteq W_2^{\Uparrow\varphi}$ satisfying the right precondition, consider a $x \in MinX$ in $\Uparrow \varphi(M_1)$. As nothing is changed in the bisimulation, the subsets

X and X' still satisfy the precondition also in the original models M_1 and M_2. The internal structure of the plausibility order within X and X', however, might be different, since we have moved down all the φ-worlds. We can make a case distinction: either $x \vDash \varphi$ or not. In the first case, consider the sets $X \cap [\![\varphi]\!]_{M_1}$ and $X' \cap [\![\varphi]\!]_{M_2}$. It is easy to see that they satisfy the precondition of condition 2, thanks to Theorem 6.2.2. Moreover, x is minimal in the former set, because the relative ranking of φ-worlds remains unchanged in the update, so by condition 2 we know there is $y \in MinX' \cap [\![\varphi]\!]_{M_2}$ such that $(x, y) \in Z$. Since the φ-worlds are pushed down in the plausibility ordering, if y was minimal in $X' \cap [\![\varphi]\!]_{M_2}$ in the original model then it must be minimal in X' in the updated model, hence $y \in MinX'$ and we obtain what required.

In the other case, if $x \vDash \neg\varphi$ then we can be sure that X contains no φ-worlds, or otherwise these would be below x after the update. Since the relation is total, bisimilarity entails modal equivalence and the sets X and X' are sent into each other by the bisimulation we can conclude that also X' contains no φ-worlds. Therefore we infer that the relative plausibility ordering within X and X' remained unaltered after the update, hence in particular x was also minimal within X in the original model. By condition 2 we receive $y \in MinX'$ such that $(x, y) \in Z$. Since X' remains untouched by the update, the condition is fulfilled.

The preservation of the totality of the relation is immediate. The only combinations left to prove is $\{[\leqslant], A\}$; we know condition 4 is always preserved so we only need to check condition 3. Suppose Z is a bisimulation$^{[\leqslant],A}$ and suppose $(w, w') \in \Uparrow \varphi(Z)$. Assume moreover that there is $v \leqslant w$ in the updated model. Here we also need to make a case distinction. If both w and v satisfy φ (or $\neg\varphi$) then we can directly apply condition 3 to obtain what we need. The only case left is when $v \vDash \varphi$ and $w \vDash \neg\varphi$ (the fourth case is ruled out by the update). By the totality of the relation we know that there must be v' in the other model such that $(v, v') \in Z$, and by Theorem 6.2.2 we get that $v' \vDash \varphi$. But then we can be sure that, by the action of the update, in the updated model we have $v' \leqslant w'$ (because clearly $w' \vDash \neg\varphi$). The other claims follow from what we just proved and the inter-dependencies between conditions. $\Box$

Suggestion

The third an last update we consider on evidence models captures a situation in which the information received by the agent is reliable enough for her to remove the plausibility link from $\neg\varphi$ to φ-worlds.

6.3.12. Definition. [Suggestion, [26] p.4] Given a formula $\varphi \in \mathcal{L}_K$, a *suggestion of* φ is a construction of type $\#\varphi : \mathbf{PM}_K \to \mathbf{PM}_K$. For a plausibility model $M = \langle W, \leqslant, V\rangle$ the action on objects $\#\varphi(M) = M^{\#\varphi}$ is defined as:

- $W^{\#\varphi} = W$

- $V^{\#\varphi}(p) = V(p)$
- $\leqslant^{\#\varphi}(w) = \leqslant \setminus \{(v, v') \in \leqslant | v \vDash \neg\varphi, v' \vDash \varphi\}$

Given a plausibility bisimulation $Z \subseteq M_1 \times M_2$, define $\#\varphi(Z) = Z$.

The dynamic operator associated to the suggestion of φ is commonly denoted with $[\#\varphi]$ and its semantics is

$$M, w \vDash [\#\varphi]\psi \quad \text{iff} \quad M^{\#\varphi}, w \vDash \psi$$

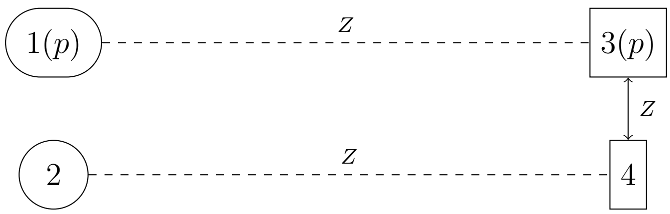
Note that *condition* here is equivalent to a tautology.

6.3.13. LEMMA. *Suggestion breaks bisimilarity^K for $K = \{B, [\leqslant]\}$.*

Proof:
The counterexample is the same as in Lemma 6.3.9. The bisimulation there satisfies condition 1 and 3, but after upgrading with $\#\neg p$ we cut the link between the two world on top and the lower world, in both models. This creates two new minimal worlds that cannot be bisimilar: again $B\neg q$ is true in one model and false in the other. □

6.3.14. LEMMA. *Suggestion breaks bisimilarity^K for all $K \subseteq \{B, B^\psi, A\}$ except for K empty or $K = \{A\}$.*

Proof:
Consider the two models depicted in the image below and connected by a relation Z.



It is straightforward to check that Z is a bisimulation$^{B,B^\psi,A}$ between the models. If both models are updated with $\#p$ then the first model remains unaltered while in the second model we remove the edge $4 \leqslant 3$.

This change cracks the symmetry between the two models: now 3 is the only minimal world in the second model. But this means that the world 2 in the first model is not connected by Z to any minimal world, violating condition 1. The impossibility of a bisimulation between the two updated models is witnessed by the fact that Bp is now true in the second model and false in the first. □

6.3.15. PROPOSITION. *The operator of suggestion cannot be reduced to any static language consisting of the propositional language enriched with one of the following sets of operators*

- *plain belief,*
- *plain belief and plausibility modality,*
- *conditional belief,*
- *plain belief and universal modality.*

Proof:
The previous Lemma provides the counter-models that trigger Proposition 6.3.2 for this update. □

6.3.16. PROPOSITION. *The construction* $\#\varphi : \mathbf{PM}_K \to \mathbf{PM}_K$ *is functorial for* $\{B^\psi, [\leqslant]\} \subseteq K \subseteq \{B, B^\psi, [\leqslant], A\}$, $K = \{[\leqslant]\}$ *and* $\{A\}$.

Proof:
The construction $\#\varphi$ sends identity relations to identity relations and preserves relational composition. We check that the construction preserves bisimulations.

The preservation of the totality of the relation is immediate. For condition 3 (operator $[\leqslant]$), suppose $(w, w') \in \#\varphi(Z)$ and $v \leqslant w$ in $\#\varphi(M_1)$. Then this was also the case in the original model, as we have not added any link, thus there is $v' \leqslant w'$ such that $(v, v') \in Z$. The latter link is not removed during the update, since w' and v' satisfy the same formulas as w and v and $(w, w') \in \#\varphi(Z)$.

Suppose Z is a bisimulation$^{B^\psi,[\leqslant]}$ and suppose $(w, w') \in \#\varphi(Z)$. Given two subsets $X \subseteq W_1^{\Uparrow\varphi}$ and $X' \subseteq W_2^{\Uparrow\varphi}$ satisfying the right precondition, consider $x \in MinX$ in $\#\varphi(M_1)$. As nothing is changed in the bisimulation, the subsets X and X' still satisfy the precondition also in the original models M_1 and M_2.

We can distinguish two cases: either $x \in MinX$ also in the original model or not. In the first case, by the property of the bisimulation, we can conclude that there is $y \in MinX'$ in M_2 such that $(x, y) \in Z$. If y is not minimal in X' in the updated model this means that we have removed a link $y \leqslant y'$ for some y' (breaking a tie), thus there is $y' < y$ such that $y' \in X'$, $y' \vDash \varphi$ and $y \vDash \neg\varphi$. But then by condition 3 (which is preserved as we have seen) there must be x' in the first model such that $x' \leqslant x$ and $(x', y') \in Z$; by the precondition on X, X' we must have $x' \in X$. Since bisimulation preserves modal equivalence we must have $x' \vDash \varphi$ and $x \vDash \neg\varphi$. But then the link $x \leqslant x'$, if it existed, has been removed so now $x' < x$ and thus x is not minimal in X, contradiction. Hence y is minimal in X' in the updated model.

For the second case, suppose x is not minimal in X in the original model. Then this means that x became minimal after the update: with the update we have removed some links $x' \leqslant x$ for x''s below x. This in turn means that $x \vDash \varphi$ and $x' \vDash \neg\varphi$ for all such x''s. Consider $X \cap [\![\varphi]\!]$ and $X' \cap [\![\varphi]\!]$, it is easy to see that this pair also fulfills the prerequisites of condition 2 on bisimulations. The element x is minimal in $X \cap [\![\varphi]\!]$, so by condition 2 there must be y that is minimal in $X' \cap [\![\varphi]\!]$ and such that $(x, y) \in Z$. If there is an element z below y in X' then z must satisfy $\neg\varphi$ and thus after the update the link $z \leqslant y$ is removed. So after the update y is minimal in X'. The other claims follow from what we just proved and the inter-dependencies between conditions. □

6.3.2 Updates on evidence bisimulations

In this section we analyze some dynamic updates on evidence models and consider to what extent they can be regarded as endofunctors on $\mathbf{EM}_N$. We will only focus on the interesting cases, rather than giving a complete account.

Public announcement

The update known as public announcement can also be defined on evidence models.

6.3.17. Definition. [Public announcement, [29]] Given a formula $\varphi \in \mathcal{L}_N$, a *public announcement of* φ is a construction of type $\varphi! : \mathbf{EM}_N \to \mathbf{EM}_N$. For an evidence model $M = \langle W, E, V \rangle$ the action on objects $\varphi!(M) = M^{\varphi!}$ is defined as:

- $W^{\varphi!} = W \cap [\![\varphi]\!]$
- $V^{\varphi!}(p) = V(p) \cap [\![\varphi]\!]$
- $E^{\varphi!}(w) = \{Y \cap [\![\varphi]\!] | Y \cap [\![\varphi]\!] \neq \emptyset, Y \in E(w)\}$

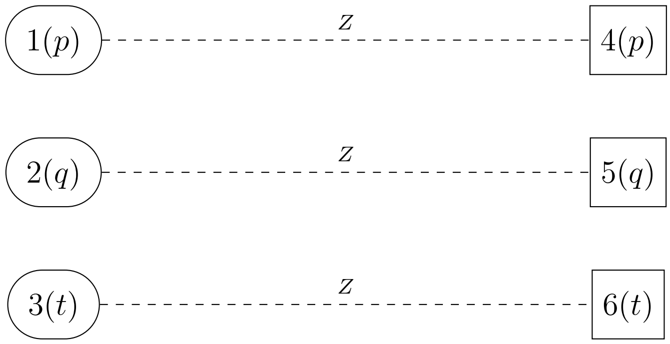
Given an evidence bisimulation $Z \subseteq M_1 \times M_2$, define $\varphi!(Z) = Z \cap [\![\varphi]\!]_{M_1} \times [\![\varphi]\!]_{M_2}$.

The dynamic operator associated to public announcement of φ has the same semantics as for plausibility models:

$$M, w \vDash [\varphi!]\psi \quad \text{iff} \quad M, w \vDash \varphi \quad \text{entails} \quad M^{\varphi!}, w \vDash \psi$$

6.3.18. Lemma. *Public announcement breaks bisimilarity^N for any $N \subseteq \{B, \Box, A\}$ beside the empty set and the singleton $\{A\}$.*

Proof:
Consider the two models depicted in the image below and connected by the bisimulation$^{B,\Box,A}$ Z.



The only evidence available in the model on the left, at all worlds, is $\{\{1\},\{1,2\},W_1\}$; the only evidence available in the second model, at all worlds, is $\{\{4\},\{4,6\},W_2\}$. It is easy to check that Z is a bisimulation$^{B,\Box,A}$ between the models, thus in particular it is also a bisimulation^{B,A}, a bisimulation^{B}, a bisimulation$^{\Box}$, a bisimulation$^{B,\Box}$ and a bisimulation$^{\Box,A}$. If both models undergo a public announcement of $\neg p!$ then the evidence sets will become just $\{\{2\},W_1\}$ and $\{\{6\},W_2\}$. This means that the resulting bisimulation does not satisfy condition 1 and 3. Indeed, the formula Bq is now true in the first model and false in the second; likewise after the update we can see that $\Box q$ is true in the first model but false in the second. □

6.3.19. Proposition. *On evidence models, the operator of public announcement cannot be reduced to any static language consisting of the propositional language enriched with one of the following set of operators*

- *plain belief,*
- *evidence modality,*
- *plain belief and universal modality,*
- *plain belief and evidence modality,*
- *plain belief together with both evidence modality and universal modality.*

Proof:
The last Lemma provides the counter-models that triggers Proposition 6.3.2 for public announcement. □

6.3.20. Proposition. *The construction* $\varphi! : \mathbf{EM}_N \to \mathbf{EM}_N$ *is functorial for* $\{B^{\psi},\Box(\Psi)\} \subseteq N \subseteq \{B,B^{\psi},\Box,A,\Box(\Psi)\}$ *or* $\{B^{\psi}\} \subseteq N \subseteq \{B,B^{\psi},A\}$ *or* $\{\Box(\Psi)\} \subseteq N \subseteq \{\Box,A,\Box(\Psi)\}$.

Proof:
Clearly $\varphi! : \mathbf{EM}_N \to \mathbf{EM}_N$ sends identity relations to identity relations. Thanks to Theorem 6.2.6 it preserves relational composition: if three worlds are linked by two bisimulations then they are pairwise bisimilar, thus either they all satisfy φ or they do not, meaning that the $\varphi!$ construction either removes them all or leaves them in the models. This ensures that no pairs are removed from the composite relation; the fact that no pairs are added is given by the definition in terms of intersection.

We furthermore need to check that the construction preserves bisimulations, that is, if $Z \subseteq M_1 \times M_2$ is a bisimulation^N then $\varphi!(Z)$ is still a bisimulation^N between the models $\varphi!(M_1)$ and $\varphi!(M_2)$. For each condition we only describe the proof of one direction, the converse being proved symmetrically.

Suppose Z is a bisimulation$^{B^\psi}$ and suppose $(w, w') \in \varphi!(Z)$. Given two subsets $X \subseteq W_1^{\varphi!}$ and $X' \subseteq W_2^{\varphi!}$ satisfying the right precondition, consider a w-X-scenario $\mathcal{X}$ in $\varphi!(M_1)$ and $x \in \bigcap \mathcal{X}^X$. Because bisimilarity$^{B^\psi}$ entails modal equivalence as per Theorem 6.2.6, the subsets X and X' still satisfy the precondition also in the extended models M_1 and M_2 (otherwise we would have a φ-world in the bisimulation with a world not satisfying φ). Furthermore, it is easy to see that every w-X-scenario $\mathcal{X}$ in $\varphi!(M_1)$ is the restriction of some w-X-scenario $\mathcal{X}'$ in M_1. Clearly x still belongs to $x \in \bigcap \mathcal{X}'^X$, hence by the property of the bisimulation we obtain a w'-X'-scenario $\mathcal{Y}'$ in M_2 and $y \in \mathcal{Y}'$ such that $(x, y) \in Z$. Again by Theorem 6.2.6 we know that if x satisfies φ then the same must hold for y, so we can conclude $(x, y) \in \varphi!(Z)$. Intersecting the evidence sets in $\mathcal{Y}'$ with the extension of φ we obtain a w'-X'-scenario $\mathcal{Y}$ in $\varphi!(M_2)$ such that $y \in \bigcap \mathcal{Y}$.

The fourth condition is rather immediate: removing all the worlds not satisfying φ we still retain a total relation. As for the fifth condition, suppose Z is a bisimulation$^{\Box(\Psi)}$. Consider $X \in E^{\varphi!}(w)$: by construction we know that $X = X' \cap [\![\varphi]\!]_{M_1}$ for some $X' \in E(w)$. The assumption $(w, w') \in \varphi!(Z)$ entails $(w, w') \in Z$, which in turn by the condition on bisimulation$^{\Box(\Psi)}$ entails that there must be some $Y' \in E'(w')$. By the property of Z can conclude that for every $y \in Y'$ there is $x \in X'$ such that $(x, y) \in (Z)$ and vice versa. Now take $Y = Y' \cap [\![\varphi]\!]_{M_2}$ and $y \in Y$. By what we just stated there will be $x \in X'$ such that $(x, y) \in (Z)$. By Theorem 6.2.6 we know that also $x \vDash \varphi$, so $x \in X$. The converse switching X and Y is proved analogously; the other direction switching w and w' likewise. The other claims follow from what we just proved and the inter-dependencies between the conditions. □

Evidence weakening

The next update encodes the idea that the agent receives an information undermining the credibility of φ, as a result she adds the worlds satisfying $\neg\varphi$ to all her evidence sets.

6.3.21. Definition. Given a formula $\varphi \in \mathcal{L}_N$, *evidence weakening of* φ is a construction of type $\cup\varphi : \mathbf{EM}_N \to \mathbf{EM}_N$. For an evidence model $M = \langle W, E, V\rangle$ the action on objects $\cup\varphi(M)$ is defined as:

- $W^{\cup\varphi} = W$
- $V^{\cup\varphi}(p) = V(p)$
- $E^{\cup\varphi}(w) = \{X \cup [\![\neg\varphi]\!] | X \in E(w)\}$

Given an evidence bisimulation $Z \subseteq M_1 \times M_2$, define $\cup\varphi(Z) = Z$.

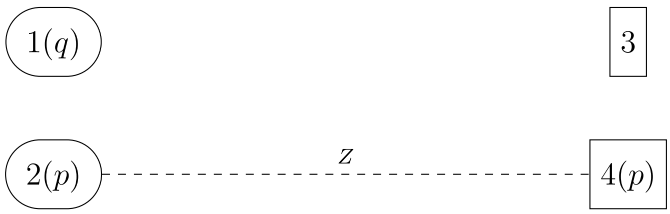
The dynamic operator associated to the evidence weakening of φ is denoted with $[\cup\varphi]$ and its semantics is

$$M, w \models [\cup\varphi]\psi \quad \text{iff} \quad \cup\,\varphi(M), w \models \psi$$

Note that *condition* here is equivalent to a tautology.

6.3.22. Lemma. *Evidence weakening breaks bisimilarity^B, bisimulation$^{B,\Box}$ and bisimulation$^\Box$.*

Proof:
Consider the two models depicted in the image below and connected by a the bisimulation Z.



The evidence available in the model on the left, at all worlds, is $\{\{2\}, W_1\}$; the evidence available in the second model, at all worlds, is $\{\{4\}, W_2\}$. It is easy to check that Z is a bisimulation^B between the models: essentially the only possible scenarios are the singletons of 2 and 4. If both models undergo an evidence weakening operation $\cup p$ then the evidence sets will become just $\{W_1\}$ and $\{W_2\}$. As a result there cannot be a bisimulation^B between the two models: since in $\cup p(M_1)$ the formula $B(p \vee q)$ is true at all worlds and in $\cup p(M_2)$ the same formula is false at all worlds, by Theorem 6.2.6 we know that there cannot be any bisimulation between them.

For bisimulation$^\Box$, notice that Z is also a bisimulation$^\Box$, but after the update we have that in the second model $\Box\neg q$ is true at all worlds but the same formula is everywhere false in the first model. Notice that $\Box\neg q$ was true in both models before the update: removing the evidence for p also influenced the evidence concerning other propositional variables. $\Box$

6.3.23. PROPOSITION. *On evidence models, the operator of evidence weakening cannot be reduced to any static language consisting of the propositional language enriched with one of the following set of operators*

- *plain belief,*
- *evidence modality,*
- *plain belief and evidence modality.*

Proof:
Directly from the last Lemma and Proposition 6.3.2. □

However, it is enough to add one of the other operators, either the global modality or conditional belief or the instantial neighborhood modality, to make evidence weakening functorial.

6.3.24. PROPOSITION. *For* $\{A\} \subseteq N \subseteq \{B, \Box, A\}$ *the construction* $\cup\varphi : \mathbf{EM}_N \to \mathbf{EM}_N$ *is functorial.*

Proof:
If a bisimulation between two evidence models is total clearly in the updated model the relation will remain total. Now suppose Z is a bisimulation^{B,A} and consider a w-scenario $\mathcal{X}$ in $\cup\varphi(M_1)$. The latter must be obtained from a w-scenario $\mathcal{X}'$ in M_1 by adding $[\![\neg\varphi]\!]$ to all the sets in $\mathcal{X}'$. Thus if $x \in \bigcap \mathcal{X}$ it is either $x \in \bigcap \mathcal{X}'$ or $x \in [\![\neg\varphi]\!]$. In the first case we know by condition 1 that there is a w'-scenario $\mathcal{Y}'$ in M_2 and $y \in \bigcap \mathcal{Y}'$ such that $(x, y) \in Z$. Weakening all the evidences in $\mathcal{Y}'$ with $[\![\neg\varphi]\!]$ we obtain a w'-scenario $\mathcal{Y}'$ in $\cup\varphi(M_2)$ that does the job. If $x \in [\![\neg\varphi]\!]$ then by totality and Theorem 6.2.6 there is $y \in [\![\neg\varphi]\!]$ such that $(x, y) \in Z$. Such y belongs to the intersection of every scenario in $\cup\varphi(M_2)$, by construction.

Assume now Z is a bisimulation$^{\Box,A}$ and consider $X \in E^{\cup\varphi}(w)$. By construction we know $X = X' \cup [\![\neg\varphi]\!]$ for some $X' \in E(w)$. By condition 3 there exists $Y' \in E'(w')$ such that for all $y \in Y'$ there is $x \in X'$ such that $(x, y) \in Z$. By totality and Theorem 6.2.6 for every $y \in [\![\neg\varphi]\!]$ there is $x \in [\![\neg\varphi]\!]$ such that $(x, y) \in Z$, thus taking $Y = Y' \cup [\![\neg\varphi]\!]$ we fulfill condition 3. The claim for bisimulation$^{B,\Box,A}$ follows from what we already proved. □

6.3.25. PROPOSITION. *The construction* $\cup\varphi : \mathbf{EM}_N \to \mathbf{EM}_N$ *is functorial for any* N *such that* $\{\Box(\Psi)\} \subseteq N \subseteq \{B, \Box, A, \Box(\Psi)\}$ *or* $\{B^\psi\} \subseteq N \subseteq \{B, B^\psi, \Box, A\}$ *or* $N = \{B, B^\psi, \Box, A, \Box(\Psi)\}$.

Proof:
Suppose Z is bisimulation$^{\Box(\Psi)}$. Take $X \in E^{\cup\varphi}(w)$, it must come from $X = X' \cup [\![\neg\varphi]\!]$. By property of Z we know there is a $Y' \in E'(w)$ that matches with X'.

Since the relation is total, we also know that for every $v \in [\![\neg\varphi]\!]_{M_1}$ there will be a v' in M_2 such that $(v, v') \in Z$, and by Theorem 6.2.6 we will also have $v' \models \neg\varphi$. So every $x \in X = X' \cup [\![\neg\varphi]\!]$ has a corresponding $y \in Y = Y' \cup [\![\neg\varphi]\!]$ such that $(x, y) \in Z$ and every $y \in Y = Y' \cup [\![\neg\varphi]\!]$ has a corresponding $x \in X = X' \cup [\![\neg\varphi]\!]$ such that $(x, y) \in Z$. The converse direction works symmetrically.

Suppose Z is bisimulation$^{B^\psi}$. Let a pair of sets X and X' fulfill the relevant preconditions in the updated models. Since the carriers and the relations are the same as the original models, we can conclude that the same sets fulfill the preconditions in the original models. It is easy to see that a w-X-scenario $\mathcal{X}$ in $\cup\varphi(M_1)$ must come from a w-X-scenario $\mathcal{X}'$ in M_1 by adding $[\![\neg\varphi]\!]$ to all the evidence sets in $\mathcal{X}'$. Thus if $x \in \bigcap \mathcal{X}^X$ it is either $x \in \bigcap \mathcal{X}'^X$ or $x \in [\![\neg\varphi]\!]$. In the first case we obtain by the property of Z a matching w'-X'-scenario $\mathcal{Y}'$ in M_2 that does the job. In the second case we know that, since the relation is total, we must have y in M_2 such that $(x, y) \in Z$, and by Theorem 6.2.6 we will also have $y \models \neg\varphi$. Note that in the updated model $\cup\varphi(M_2)$ every evidence sets contains $[\![\neg\varphi]\!]$, thus y will be in the intersection of any scenario in the updated model (and there is always at least one). The converse is proved analogously.

We have showed that bisimulation$^{B^\psi}$ and bisimulation$^{\Box(\Psi)}$ are preserved by evidence weakening; the rest of the claim follows from Proposition 6.3.24 and the inter-dependencies between conditions. □

Evidence addition

Finally, we analyze the case in which the agent receives a piece of evidence that is as reliable as the ones she already possesses.

6.3.26. Definition. [Evidence addition, [29] p.70] Given a formula $\varphi \in \mathcal{L}_N$, the *evidence addition of* φ is a construction of type $\varphi! : \mathbf{EM}_N \to \mathbf{EM}_N$. For an evidence model $M = \langle W, E, V\rangle$ the action on objects $+\varphi(M)$ is defined as:

- $W^{+\varphi} = W$
- $V^{+\varphi}(p) = V(p)$
- $E^{+\varphi}(w) = E(w) \cup \{[\![\varphi]\!]\}$ if $[\![\varphi]\!] \neq \emptyset$, $E^{+\varphi}(w) = E(w)$ otherwise

Given an evidence bisimulation $Z \subseteq M_1 \times M_2$, define $+\varphi(Z) = Z$.

The case distinction in the definition prevents us from be adding $\emptyset$ to $E(w)$, which would be against the definition of evidence model. We assume that the agent does not accept contradictions as reliable evidence and thus the update works as the identity on the models where φ is everywhere false.

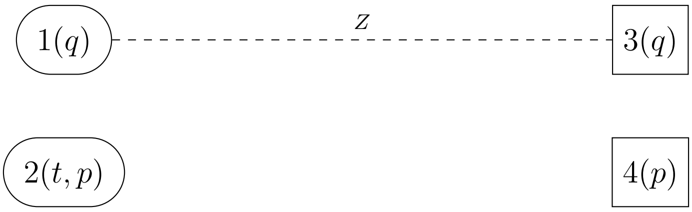
The dynamic operator associated to the evidence addition of φ is denoted with $[+\varphi]$ and its semantics is

$$M, w \vDash [+\varphi]\psi \quad \text{iff} \quad M, w \vDash E\varphi \quad \text{entails} \quad +\varphi(M), w \vDash \psi$$

where E is the dual of the universal modality A; the precondition here encodes the existence of a world where φ holds.

6.3.27. Lemma. *Evidence addition breaks bisimilarity^N for $\emptyset \neq N \subseteq \{B, \Box\}$.*

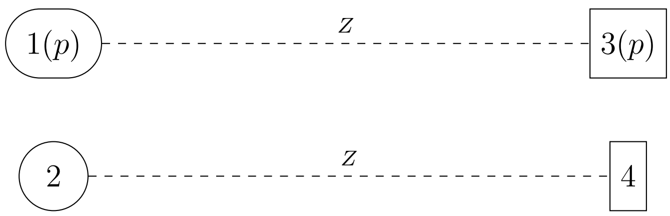
Proof:
Consider the two models depicted in the image below and connected by a bisimulation Z.



The evidence available in the model on the left, at all worlds, is $\{\{1\}, W_1\}$; the only evidence available in the second model, at all worlds, is $\{\{3\}, W_2\}$. It is straightforward to check that Z is a bisimulation$^{B,\Box}$ between the models. If both models are updated with $+p$ then the evidence sets will be enriched with $\{2\}$ and $\{4\}$ in the first and second model respectively. To have a bisimulation$^{\Box}$ between the two models there should be some world in $+p(M_2)$ in bisimulation with 2, which is now the only world in an evidence set. But there is no world in $+p(M_2)$ satisfying the same propositional letters, so there can be no such bisimulation. This is also witnessed by the fact that now $\Box t$ is false in $+p(M_2)$ but true in $+p(M_1)$. An analogous argument works for bisimulation^B: Z is a bisimulation^B but after the update $B\neg t$ is false in the first model but true in the second. $\Box$

6.3.28. Lemma. *Evidence addition breaks bisimilarity^N for all $N \subseteq \{B, B^{\psi}, A\}$ except for N empty or $N = \{A\}$.*

Proof:
Consider the two models depicted in the image below and connected by a bisimulation Z.



The evidence available in the model on the left, at all worlds, is $\{\{1\},\{2\},W_1\}$; the only evidence available in the second model, at all worlds, is $\{W_2\}$. It is straightforward to check that Z is a bisimulation$^{B,B^{\psi},A}$ between the models. If both models are updated with $+\neg p$ then the collection of evidence sets in the second model will be enriched with $\{4\}$, while the first model remains unaltered.

This apparently little change in the second model cracks the symmetry between scenarios in the two models: now there is a scenario in the first model, namely $\{\{1\},W_1\}$, and a world in its intersection, world 1, for which we have to find a scenario in the second model and a world in its intersection such that it is in the bisimulation with 1. But there is no such scenario, because now the only scenario available in the second model is $\{\{4\},W_2\}$. The impossibility of a bisimulation between the two updated models is witnessed by the fact that $B\neg p$ is now true in the second model and false in the first. □

6.3.29. Proposition. *On evidence models, the operator of evidence addition cannot be reduced to any static language consisting of the propositional language enriched with one of the following set of operators*

- *plain belief,*
- *evidence modality,*
- *plain belief and evidence modality,*
- *conditional belief,*
- *plain belief and universal modality.*

Proof:
Directly from the last two Lemmas and Proposition 6.3.2. □

6.3.30. Proposition. *The construction* $+\varphi : \mathbf{EM}_N \to \mathbf{EM}_N$ *is functorial for* $\emptyset \neq N \subseteq \{A, \Box(\Psi)\}$.

Proof:

The update $+\varphi$ sends identity relations to identity relations and preserves relational compositions, thus we turn our attention to the preservation of bisimulations. If Z is total then clearly it remains total after the update, thus condition 4 is preserved.

Suppose Z is a bisimulation$^{\Box(\Psi)}$ and suppose $(w,w') \in +\varphi(Z)$. Let $X \in E^{+\varphi}(w)$: if $X \neq [\![\varphi]\!]$ then by the property of Z we have a corresponding $Y \in E^{+\varphi}(w')$ in the other models fulfilling the required conditions. If $X = [\![\varphi]\!]_{M_1}$

then, since the bisimulation is total, we know that any world $x \in [\![\varphi]\!]_{M_1}$ will have a counterpart y in the second model; moreover, due to Theorem 6.2.6, $y \in [\![\varphi]\!]_{M_2}$ as well. The same holds in the other direction, thus $[\![\varphi]\!]_{M_1}$ and $[\![\varphi]\!]_{M_2}$ are sent into each other by the bisimulation Z. An analogous reasoning takes care of the other direction. □

6.4 Connecting evidence to plausibility models

We now address the possibility to connect the category of evidence model with the category of plausibility models via functors. Evidently evidence models are richer and it can be easily shown that they 'collapse' into plausibility models. However, the picture gets more complicated if we take into account which languages we use to describe such models, or in other words, if we consider also the bisimulations between them. We stress two points:

(1) If we choose a strong language to describe plausibility models and a weak language to characterize evidence models it might be the case that, during the 'collapse', we break the indistinguishability relation between models. This constitutes a breach of functoriality, as in the previous sections on update endofunctors.

(2) If we choose equally expressible languages then we might be able to describe only features that are preserved during the collapse: in such cases the two categories may turn out to have a strong connection, as an adjunction or an equivalence.

These two points substantiate the idea that the differences between the two level emerge only when we adopt a stronger language to describe evidence models.

6.4.1. Definition. [[29]] Given an evidence model $\mathcal{M} = \langle W, \mathcal{E}, V \rangle$ construct the plausibility model $ORD(M) = \langle W, \leqslant_{\mathcal{E}}, V \rangle$ where $\leqslant_{\mathcal{E}}$ is defined as[4]

$$w \leqslant_{\mathcal{E}} v \text{ iff } \forall X \in \mathcal{E}, v \in X \text{ implies } w \in X$$

Define ORD on bisimulations as $ORD(Z) = Z$.

A reader with some knowledge in Topology or Duality Theory will recognize this construction as (the converse of) the *specialization preorder* obtainable from a neighborhood structure. The following proposition showcases our first point.

[4]This definition only applies to a subcategory of evidence models, the uniform ones. The 'natural' counterpart of regular evidence models would be plausibility models where the relation $\leqslant$ is relativized to possible worlds. Then we would define

$$w \leqslant^k_E v \text{ iff } \forall X \in E(k), v \in X \text{ implies } w \in X$$

6.4.2. Proposition. *No construction ORD : $\mathbf{EM}_{\{B,\Box\}} \to \mathbf{PM}_{\{B^{\psi}\}}$ that sends an evidence model M to $ORD(M)$ can be functorial.*

Proof:
Consider the following pair of models:

$$M_1 = \langle \{w_1, w_2\}, \mathcal{E}_1 = \{\{w_1, w_2\}, \{w_1\}\}, V(w_1) = \{p\}, V(w_2) = \{t\}\rangle$$

and

$$M_2 = \langle \{v\}, \mathcal{E}_2 = \{\{v\}\}, V(v) = \{p\}\rangle$$

There is an evidence bisimulation$^{\{B,\Box\}}$ between these two models: $Z = \{(w_1, v)\}$. After performing the construction ORD on both models we have $w_1 \leqslant_1^{\mathcal{E}_1} w_2$, plus the obvious reflexive edges on both models. Looking at the resulting plausibility models it is immediate to see that there can not be a plausibility bisimulation$^{\{B^{\psi}\}}$ between them: such relation would have to be total and there is no counterpart satisfying the same propositional letters as w_2. Since we cannot construct a plausibility bisimulation between $ORD(M_1)$ and $ORD(M_2)$, we cannot properly define an image for the evidence bisimulation Z. Therefore no functor can be defined on these premises. $\Box$

Note that this counterexample works regardless of how we define the action on arrows for the construction ORD. Clearly the same counterexample still works if we consider only evidence bisimulation for operator $\{B\}$ or operator $\{\Box\}$.

6.4.3. Definition. [[29]] Call a set $X \subseteq W$ *downward closed* with respect to $\leqslant$ if $w \in X$ and $v \leqslant w$ entails $v \in X$. A set with such property is called *down-set.*

Given a plausibility model $\mathcal{M} = \langle W, \leqslant, V\rangle$ construct the evidence model $EV(M) = \langle W, \mathcal{E}_{\leqslant}, V\rangle$ where $\mathcal{E}_{\leqslant}$ is the set of non-empty downward closed subsets of W. Define EV on arrows as $EV(Z) = Z$.

We now address our second point.

6.4.4. Theorem. *The constructions ORD : $\mathbf{EM}_{\{B^{\psi}\}} \to \mathbf{PM}_{\{B^{\psi}\}}$ and EV : $\mathbf{PM}_{\{B^{\psi}\}} \to \mathbf{EM}_{\{B^{\psi}\}}$ are both functorial and constitute a categorical equivalence.*

Proof:
We begin checking functoriality, starting from ORD. Clearly identities and composition are preserved. Suppose M_1 and M_2 are two uniform evidence models and say there is an evidence bisimulation$^{\{B^{\psi}\}}$ Z between them. Let $(w, w') \in ORD(Z) = Z$ and take X and X' in $ORD(M_1)$ and $ORD(M_2)$ satisfying the prerequisites for $ORD(Z) = Z$. Since the domains of the structures remained the same, the same sets must also satisfy the precondition in the models M_1 and M_2. Thus for every w-X-scenario $\mathcal{X}$ and $x \in \bigcap \mathcal{X}^X$ there is a w'-X'-scenario $\mathcal{Y}$ and $y \in \bigcap \mathcal{Y}^{X'}$ such that $(x, y) \in Z$, and vice versa.

Now suppose $x \in MinX$ in $ORD(M_1)$: being minimal with respect to $\leqslant_{\mathcal{E}_1}$ means that there is no $v \in X$ such that $v <_{\mathcal{E}_1} x$, thus there is no $Y \in \mathcal{E}_1$ such that $v \in Y$ and $x \notin Y$. We claim this means that there is a w-X-scenario $\mathcal{X}$ and $x \in \bigcap \mathcal{X}^X$. Clearly $x \in X$ is given by $x \in MinX$. Now consider the w-X-scenario $\mathcal{X}$ consisting of all evidence sets in $E(w)$ that contain x. This is clearly consistent with X, thanks to x itself. Now suppose $\mathcal{X}$ does not have the f.i.p. with respect to X: then there is a set Y that has non-empty intersection with the sets in $\mathcal{X}$. This would entail that there is a $v \in X$ such that $v \in Y$ and $w \notin Y$ (otherwise Y would be already in $\mathcal{X}$). But this contradicts $x \in MinX$ in $ORD(M_1)$. Thus we apply the property of Z and receive a w'-X'-scenario $\mathcal{Y}$ and $y \in \bigcap \mathcal{Y}^{X'}$ such that $(x, y) \in Z$. We claim this means that $y \in MinX'$. By contradiction suppose there is $v' <_{\mathcal{E}_1} y$ with $v' \in X'$, then there must be $K' \in E(w')$ such that $v' \in K'$ and $y \notin K'$ and moreover all the evidence sets containing y also contain v'. But then we could add consistently K' to $\mathcal{Y}$, which contradicts the fact that it is maximal with the f.i.p. with respect to X'. We thus conclude $y \in MinX'$. The converse is proved analogously.

We now prove the functoriality of EV. Let M_1 and M_2 be two plausibility models and say there is a plausibility bisimulation$^{\{B^\psi\}}$ Z between them. Let $(w, w') \in EV(Z) = Z$ and take X and X' in $EV(M_1)$ and $EV(M_2)$ satisfying the prerequisites. Consider a w-X-scenario $\mathcal{X}$ and $x \in \bigcap \mathcal{X}^X$, we first show that $x \in MinX$ in M_1, in order to apply the property of Z. By contradiction suppose there is $v < x$ with $v \in X$. This would mean that the down-set generated by v is a set with non empty intersection with $\mathcal{X}$ that excludes x, thus either $\mathcal{X}$ does not have the f.i.p. with respect to X or $x \notin \bigcap \mathcal{X}^X$, contradiction. Hence $x \in MinX$ and thus there is $y \in MinX'$ such that $(x, y) \in Z$.

We now claim that there is a w'-X'-scenario $\mathcal{Y}$ such that $y \in \bigcap \mathcal{Y}^{X'}$. Define $\mathcal{Y}$ as the collection of all down-sets containing y. This is consistent with X' because of y. To see it is maximal with the f.i.p., suppose there is a set $K' \notin \mathcal{Y}$ such that it can be added to $\mathcal{Y}$ preserving consistency with respect to X'. This in particular would mean that K' that does not contain y (otherwise would be already in $\mathcal{Y}$). Notice however that, since y is minimal in X' the down-set generated by y, call it Y, contains only one element in X', y itself, and clearly $Y \in \mathcal{Y}$. Thus K' and Y would have empty intersection within X', contradicting the fact that we can add K' to $\mathcal{Y}$ consistently. This proves the claim; the converse direction is symmetric.

It remains to prove the equivalence between the two categories. It is sufficient to show that there are two natural isomorphisms $\eta : ORD \circ EV \to Id_{\mathbf{PM}_{\{B^\psi\}}}$ and $\epsilon : EV \circ ORD \to Id_{\mathbf{EM}_{\{B^\psi\}}}$.

Given a plausibility model $\mathcal{M} = \langle W, \leqslant, V\rangle$, the operation $ORD \circ EV$ turns it into $ORD \circ EV(\mathcal{M}) = \langle W, \leqslant_{\mathcal{E}_{\leqslant}}, V\rangle$, where $\leqslant_{\mathcal{E}_{\leqslant}}$ is the relation defined as

$$\begin{aligned} w \leqslant_{\mathcal{E}_{\leqslant}} v \quad & \text{iff} \quad \forall X \in \mathcal{E}_{\leqslant}, v \in X \quad \text{implies} \quad w \in X \\ & \text{iff} \quad \forall X \quad \text{downward closed}, v \in X \quad \text{implies} \quad w \in X \\ & \text{iff} \quad w \leqslant v \end{aligned}$$

This shows that $\mathcal{M}$ and $ORD \circ EV(\mathcal{M})$ are the same structure: we can then just define $\eta_{\mathcal{M}}$ to be the identity of the object $\mathcal{M}$. Since both ORD and EV send a bisimulation to itself, the commutation of the naturality diagram is given by the identity laws.

We now turn our attention to the other construction, $\epsilon : EV \circ ORD \to Id_{\mathbf{EM}^u_{\{B^\psi\}}}$. Given a uniform evidence model $\mathcal{M} = \langle W, \mathcal{E}, V\rangle$, the operation $EV \circ ORD$ turns it into $EV \circ ORD(\mathcal{M}) = \langle W, \mathcal{E}_{\leqslant_{\mathcal{E}}}, V\rangle$, where $\mathcal{E}_{\leqslant_{\mathcal{E}}}$ is the closure under intersection of $\mathcal{E}$.

Define $\epsilon_{\mathcal{M}}$ to be the identity of $\mathcal{M}$. This definition makes sense because, even though the evidence in the two models might be different, the two models have the same carrier W. We need to show that such identity is a bisimulation$^{\{2\}}$, if this is the case then, being a bijection, the relation will automatically be an isomorphism in the category.

Any pair in the relation is of the form (w, w), and two pairs X and X' in the two models will satisfy the precondition of condition 2 with respect to the identity iff $X = X'$, thus what we have to verify is that for any w-X-scenario $\mathcal{X}$ in M and $x \in \bigcap \mathcal{X}^X$ there is a w-X-scenario $\mathcal{Y}$ in $EV \circ ORD(\mathcal{M})$ and $y \in \bigcap \mathcal{Y}^X$ such that $x = y$, and vice versa.

Let $\mathcal{X}$ be a w-X-scenario in M and $x \in \bigcap \mathcal{X}^X$. Since $\mathcal{E}_{\leqslant_{\mathcal{E}}}$ is the closure under intersection of $\mathcal{E}$, the same scenario $\mathcal{X}$, enriched with all the intersections, will still be a w-X-scenario in $EV \circ ORD(\mathcal{M})$ (otherwise, if there were some set K that could be added to it, it would already fail to be a w-X-scenario in M). Thus x still lies at the intersection and we have what we need.

For the converse, suppose there is a w-X-scenario $\mathcal{X}$ in $EV \circ ORD(\mathcal{M})$ and $x \in \bigcap \mathcal{X}^X$. We claim that $\mathcal{X}$ is the closure under intersection of a w-X-scenario $\mathcal{Y}$ already existing in M. Define $\mathcal{Y} = \mathcal{X} \cap \mathcal{E}$. Clearly it has non-empty intersection with X. We want to show it is maximal with the f.i.p. with respect to X in M. Suppose we could add a set K to $\mathcal{Y}$ preserving the consistency with X: then K would also be in $\mathcal{E}_{\leqslant_{\mathcal{E}}}$ and thus in $\mathcal{X}$ by the fact that the latter is maximal. But then by construction it is already in $\mathcal{Y}$. It is also easy to see that by taking the sub-collection $\mathcal{Y}$ we still retain $x \in \bigcap \mathcal{Y}^X$, hence we have the desired conclusion.

This proves that $\epsilon_{\mathcal{M}}$ is a bisimulation$^{\{2\}}$ and therefore an isomorphism in the category. Since both ORD and EV send a bisimulation to itself, the commutation of the naturality diagram is immediate by the identity laws. □

6.5 Tracking information dynamics

The notion of tracking was introduced in [26] to describe the matching of information dynamics on different structures. The author considers the case study of evidence and plausibility models, investigating when an update on evidence models is mirrored by another update at the level of plausibility models. In said paper an update is regarded as a purely semantical operation, meaning that a set is used as the parameter for the update instead of the extension of a formula.

6.5.1. Definition. [Tracking, [26]] We indicate with $U(X)$ the update that uses as a parameter the set X. We say that a function $U(X) : \mathbf{PM}_K \to \mathbf{PM}_K$ *tracks* a function $U'(X) : \mathbf{EM}_N \to \mathbf{EM}_N$ if $U(X)(ORD(M)) = ORD(U'(X)(M)$ for all X, or equivalently if the following square commutes on objects for any X:

$$
\begin{array}{ccc}
\mathbf{PM}_K & \xrightarrow{U(X)} & \mathbf{PM}_K \\
{\scriptstyle ORD}\uparrow & & \uparrow{\scriptstyle ORD} \\
\mathbf{EM}_N & \xrightarrow[U'(X)]{} & \mathbf{EM}_N
\end{array}
$$

We sometimes omit the parameter X when it is clear from the context.

Notice that the previous definition does not require the functoriality of any of the constructions involved, as we are only interested in the commutation on objects.

The definition of tracking highlights the fact that we are interested in tracking updates on the richer structures (evidence models) with updates on the poorer structures (plausibility models). The other direction, from poor to rich structures, is less interesting since every update on plausibility models has a canonical counterpart on evidence models, as the next proposition shows.

6.5.2. Proposition (See [26]). *For every update $U(X) : \mathbf{PM}_K \to \mathbf{PM}_K$ there is an update $U'(X) : \mathbf{EM}_N \to \mathbf{EM}_N$ that is tracked by $U(X)$.*

Proof:
Define $U'(X) := EV \circ U(X) \circ ORD$. We have seen in the proof of Theorem 6.4.4 that on objects $ORD \circ EV = Id_{\mathbf{PM}_K}$. Thus

$$
\begin{aligned}
ORD \circ U'(X) =& ORD \circ EV \circ U(X) \circ ORD \\
=& Id_{\mathbf{PM}_K} \circ U(X) \circ ORD \\
=& U(X) \circ ORD
\end{aligned}
$$

□

The next proposition provides an equivalent condition for tracking.

6.5.3. Proposition. *The existence of an update U tracking an update U' is equivalent to the following: for every pair of evidence models M_1, M_2, if $ORD(M_1) = ORD(M_2)$ then $ORD(U'(M_1)) = ORD(U'(M_2))$.*

Proof:
From left to right, suppose U tracks U' and $ORD(M_1) = ORD(M_2)$. Then $ORD(U'(M_1)) = U(ORD(M_1)) = U(ORD(M_2)) = ORD(U'(M_2))$.

For the other direction suppose that for every pair of evidence models M_1, M_2, if $ORD(M_1) = ORD(M_2)$ then $ORD(U'(M_1)) = ORD(U'(M_2))$. Notice that, for any evidence model M, the models $EV(ORD(M)$ and M do fulfill the antecedent: since $ORD \circ EV = Id$ we have $ORD(EV(ORD(M)) = ORD(M)$. Thus by assumption we can conclude $ORD(U'(EV(ORD(M))) = ORD(U'(M))$. We can now define $U := ORD \circ U' \circ EV$ and the proof is easy to complete: $U(ORD(M)) = ORD(U'(EV(ORD(M)))) = ORD(U'(M)$. □

The left-to-right direction of this proposition suggests how to prove that an update on evidence models *cannot* be tracked: it is sufficient to find two models for which the condition of Proposition 6.5.3 fails. This strategy is adopted in [26] to prove that some updates cannot be tracked.

The other direction of the proposition may at first glance seem to trivialize the problem of tracking: given an update U' on evidence models, we can just verify that the condition of Proposition 6.5.3 is fulfilled and then we immediately have an update $U := ORD \circ U' \circ EV$ that tracks U'. Such definition, however, is only partially satisfactory: even though it fits the bill from a semantic perspective, the interest of tracking lies in the possibility to rewrite an update on a complex structure in the language of a poorer structure. We expand on this matter in Subsection 6.5.3. The definition $U := ORD \circ U' \circ EV$ circumvents this problem altogether and is therefore not very informative. For this reason it is still noteworthy to obtain positive tracking results.

6.5.4. Proposition (See [26]). *The following statements hold:*

- *Public announcement at the plausibility level tracks public announcement on evidence models.*
- *Suggestion tracks evidence addition.*
- *Radical upgrade tracks the upgrade called "up".*[5]

6.5.1 A new tracking result

We propose an update on plausibility models that tracks evidence weakening. This tracking result is actually a direct consequence of the more general Theorem

[5]We did not cover this update here, see [26] p. 12 for details.

6.5.15, proved later; notwithstanding, we include an explicit proof for the readers that may want to see a concrete instance of a tracking proof.

6.5.5. Definition. Given a formula $\varphi \in \mathcal{L}_K$, *collapse of* φ is a construction of type $coll(\varphi) : \mathbf{PM}_K \to \mathbf{PM}_K$. For a plausibility model $M = \langle W, \leqslant, V \rangle$ the action on objects $coll(\varphi)(M)$ is defined as:

- $coll(W) = W$
- $coll(V)(p) = V(p)$
- $coll(\leqslant)$ is defined via a case distinction:
  (1) If $w, v \in [\![\varphi]\!]$ then $(w, v) \in coll(\leqslant)$;
  all φ-worlds are equi-plausible.
  (2) If $w, v \in [\![\neg\varphi]\!]$ then $(w, v) \in coll(\leqslant)$ iff $w \leqslant v$;
  the relation is unaltered on $\neg\varphi$-worlds.
  (3) If $w \in [\![\neg\varphi]\!]$ and $v \in [\![\varphi]\!]$ then $(w, v) \in coll(\leqslant)$ iff $\forall k \in W\ w \leqslant k$;
  a $\neg\varphi$-world is at least as plausible as a φ-world iff the former was the bottom element of $\leqslant$.
  (4) if $v \in [\![\neg\varphi]\!]$ and $w \in [\![\varphi]\!]$ then $(w, v) \in coll(\leqslant)$;
  all φ-worlds are at least as plausible as $\neg\varphi$-worlds.

Given a plausibility bisimulation $Z \subseteq M_1 \times M_2$ we put $coll(\varphi)(Z) = Z$.

6.5.6. Theorem (Tracking of evidence weakening). *The evidence weakening update on evidence models is tracked by the collapse update, making the following diagram commute on objects:*

$$
\begin{array}{ccc}
\mathbf{PM}_K & \xrightarrow{coll(X)} & \mathbf{PM}_K \\
\uparrow{\scriptstyle ORD} & & \uparrow{\scriptstyle ORD} \\
\mathbf{EM}_N & \xrightarrow[\cup X]{} & \mathbf{EM}_N
\end{array}
$$

Proof:
Consider an evidence model $\mathcal{M} = \langle W, \mathcal{E}, V \rangle$. The functor ORD and the two updates leave the set of worlds W and the valuation V unaltered, thus there is nothing to check there. Applying first ORD and then $coll(X)$ we obtain the relation $coll(\leqslant_{\mathcal{E}})$, while applying the update $\cup X$ and then ORD we get the relation $\leqslant_{\cup X(\mathcal{E})}$: we need to show that the two coincide, that is

$$(w, v) \in coll(\leqslant_{\mathcal{E}}) \quad \text{iff} \quad w \leqslant_{\cup X(\mathcal{E})} v$$

We do so by a case distinction:

- Suppose $w, v \in X$. Then by definition $(w, v) \in coll(\leqslant_{\mathcal{E}})$ is always the case. But also $w \leqslant_{\cup X(\mathcal{E})} v$ must always be the case: since $w, v \in X$ the condition $\forall Y \in \mathcal{E}$ if $v \in Y \cup X$ then $w \in Y \cup X$ is always fulfilled.

- Assume $w, v \in \overline{X}$. Then $(w, v) \in coll(\leqslant_{\mathcal{E}})$ iff, by definition, $w \leqslant_{\mathcal{E}} v$, which means that for all $Y \in \mathcal{E}$ if $v \in Y$ then $w \in Y$. Since we assumed $w, v \in \overline{X}$, the last condition is equivalent to the following: for all $Y \in \mathcal{E}$ if $v \in Y \cup X$ then $w \in Y \cup X$. But this is just the definition of $w \leqslant_{\cup X(\mathcal{E})} v$.

- Suppose now that $w \in \overline{X}$ and $v \in X$. By the definition of collapse, $(w, v) \in coll(\leqslant_{\mathcal{E}})$ is the case iff w is below every element in W with respect to relation $\leqslant_{\mathcal{E}}$. This latter condition is the case iff w is contained in all the evidence sets in $\mathcal{E}$: if it does then clearly it is below every other element by the definition of $\leqslant_{\mathcal{E}}$; for the other direction consider that every evidence set Y is not empty (by definition of evidence model $\emptyset \notin \mathcal{E}$) so there is $k \in Y$ but because $w \leqslant_{\mathcal{E}} k$ we get $w \in Y$.

  If for all $Y \in \mathcal{E}$ we have $w \in Y$ then for all $Y \in \mathcal{E}$ we have that if $v \in Y \cup \overline{X}$ then $w \in Y \cup \overline{X}$, because the consequent always holds. Hence $w \leqslant_{\cup X(\mathcal{E})} v$. Conversely, under the assumption $v \in X$ and $w \in \overline{X}$, the condition $\forall Y \in \mathcal{E}$ if $v \in Y \cup X$ then $w \in Y \cup X$ entails that $w \in Y$ for all $Y \in \mathcal{E}$.

- For the last case assume that $v \in \overline{X}$ and $w \in X$. Then $(w, v) \in coll(\leqslant_{\mathcal{E}})$ is always the case by definition. Note that the same holds for $w \leqslant_{\cup X(\mathcal{E})} v$: since $w \in X$, we have $w \in Y \cup X$ for every $Y \in \mathcal{E}$. □

6.5.2 Tracking and reduction laws

The definition of tracking given at the beginning of this section did not require the commutation on arrows, i.e. we did not take into account the effect of updates on bisimulations. However, when two updates are in a tracking relation then their effects on bisimulations are entangled.

6.5.7. Proposition. *Suppose a function $U : \mathbf{PM}_K \to \mathbf{PM}_K$ tracks a function $U'(X) : \mathbf{EM}_N \to \mathbf{EM}_N$, ORD and EV are functorial and U breaks bisimilarity on plausibility models. Then U' breaks bisimilarity on evidence models.*

Proof:
Suppose U breaks bisimilarity between two models M_1, M_2, call Z the plausibility bisimulation^K between them. If U' does not break bisimilarity then we could apply $ORD \circ U' \circ EV$ to $Z \subseteq M_1 \times M_2$ and obtain a bisimulation between $ORD \circ U' \circ EV(M_1)$ and $ORD \circ U' \circ EV(M_2)$. By tracking we have $ORD \circ U' \circ EV(M_1) = U \circ ORD \circ EV(M_1) = U(M_1)$ and likewise $ORD \circ U' \circ EV(M_2) = U(M_2)$, thus there is a bisimulation between $U(M_1)$ and

$U(M_2)$. But this contradicts the assumption of U breaking bisimilarity, so U' must break bisimilarity. □

This proposition can be handy to prove the non-existence of reduction laws for complicated updates on rich structures: if we have a matching update on a poorer structures we can show that the latter breaks bisimilarity, which is typically easier on simpler models. For example, we could have derived the non-existence of reduction laws for updates on evidence models exploiting the tracking results of Proposition 6.5.4 and the results of Section 6.3.1.

What about the other way around, can we transfer results from updates on evidence models to updates on plausibility models? This is possible if there is an equivalence between categories.[6]

6.5.8. PROPOSITION. *Suppose a function* $U : \mathbf{PM}_K \to \mathbf{PM}_K$ *tracks a function* $U'(X) : \mathbf{EM}_N \to \mathbf{EM}_N$*, ORD and EV constitute an equivalence and* U' *breaks bisimilarity on evidence models. Then U breaks bisimilarity on plausibility models.*

Proof:
Suppose U' breaks the bisimilarity between two evidence models M_1, M_2, call Z the evidence bisimulation^N between them. If U does not break bisimulation then we can apply $EV \circ U \circ ORD$ to $Z \subseteq M_1 \times M_2$ and obtain a bisimulation $EV \circ U \circ ORD(Z)$ between $EV \circ U \circ ORD(M_1)$ and $EV \circ U \circ ORD(M_2)$. By tracking we have $EV \circ U \circ ORD(M_1) = EV \circ ORD \circ U'(M_1)$ and likewise $EV \circ U \circ ORD(M_2) = EV \circ ORD \circ U'(M_2)$. Since ORD and EV form an equivalence, there are natural isomorphism $EV \circ ORD \circ U'(M_1) \simeq U'(M_1)$ and $EV \circ ORD \circ U'(M_2) \simeq U'(M_2)$. But then we can compose the bisimulation $EV \circ U \circ ORD(Z)$ with these isomorphisms to obtain a bisimulation between $U'(M_1)$ and $U'(M_2)$, contradicting that U' breaks bisimilarity on these models. □

We provided an example of equivalence in Theorem 6.4.4: this result could be used with the last proposition to derive, e.g., the fact that suggestion breaks plausibility bisimilarity$^{B^\psi}$ from the tracking result of Proposition 6.5.4 and the fact that evidence addition breaks evidence bisimilarity$^{B^\psi}$.

6.5.3 Tracking as a definability problem

As we mentioned previously, the interesting part of tracking is the reduction of updates cast in a complex language to updates cast in a poor language, typically a fragment. In other words, tracking is ultimately a *definability* issue.

[6]In fact the following proof goes through with a weaker assumption, namely the existence of a natural transformation $\eta : Id_{\mathbf{EM}_N} \to EV \circ ORD$; since we gave no examples of such transformations we state a slightly stronger proposition to highlight the connection with Theorem 6.4.4.

We begin by making explicit what we mean by saying that an update is defined in a certain language. For this section we focus exclusively on updates that preserve the carrier of the structure.

6.5.9. Definition. [Definability] An n-ary relation R in a model M is definable in a language $\mathcal{L}$ iff there is a formula $\varphi(\overline{x}) \in \mathcal{L}$ with n open variables such that:

$$R = \{(a_1, \dots, a_n) | M \vDash \varphi(\overline{x})[(a_1, \dots, a_n)]\}$$

The signature of plausibility models is FOL with a binary relational symbol $\leqslant$ which is meant to be interpreted on the plausibility relation. For evidence models we need a stronger language in order to quantify over evidence sets.

6.5.10. Definition. [Evidence language] Consider the grammar

$$\varphi ::= \quad \mathcal{E}(n) \,|\, x \in n \,|\, \neg\varphi \,|\, \varphi \wedge \varphi \,|\, \forall x\, \varphi \,|\, \forall n\, \varphi$$

where variables the $n, n', \dots$ are for subsets and variables $x, y, \dots$ are for elements. To the signature we add a unary predicate $\mathcal{E}$ on subsets, denoting whether a subset is a piece of evidence, and a binary relation $\in$ denoting elementhood. We adopt the standard conventions for free and bound variables, as well as the classical abbreviations for defined propositional connectives.

We will use 'plausibility language' or 'evidence language' to refer to such languages. The semantics of these languages are just the standard first and second-order semantics; the former language is meant to be interpreted over the class of plausibility models, while the second over the class of evidence models.

To define an update we define its output via a formula containing the suitable parameters. On plausibility models, given an update U, a plausibility model M and a set P, we will be interested in defining the plausibility relation on the updated model $U(P)(M)$, thus we will need a formula $\beta(x, y, P, \leqslant)$ such that:

- it depends on P, a unary predicate interpreted on the set P;[7]
- it depends on $\leqslant$, a binary relational symbol interpreted on the relation $\leqslant$;
- it has two open variables in order to define a binary relation;
- it is in the signature of plausibility models.

In the case of evidence models we define an update $U'(P)$ with a formula $\alpha(n, P, \mathcal{E})$, where n is an open variable of sort 'subset', P is again a unary predicate for worlds and $\mathcal{E}$ is the aforementioned unary predicate for subsets; a formula $\alpha(n, P, \mathcal{E})$ will denote the evidence sets of the updated model. We can now state precisely what the problem of tracking amounts to in the case of evidence and plausibility models.

[7]We ambiguously use the same symbol for the corresponding semantic and the syntactic objects; the context will disambiguate.

6.5.11. QUESTION. [Tracking] Given an update U' on evidence models defined by a formula $\alpha(n, P, \mathcal{E})$ in the evidence language, can we find an update U on plausibility models that tracks U' and is defined by a formula $\beta(x, y, P, \leqslant)$ in the plausibility language?

The optimal answer to this problem would be a characterization result giving:

- the syntactic shape of all and only the trackable updates;
- for the updates of that shape, an algorithm that manipulates syntactically the corresponding formulas α and produces the definitions β of their tracking companion;
- for the updates that are not of that shape, a procedure to construct a counter-example to tracking.

Notice that the evidence language is strong enough to express the action of the function ORD: given an evidence collection $\mathcal{E}$ the relation $x \leqslant_{\mathcal{E}} y$ is defined by the formula with two open variables

$$x \leqslant_{\mathcal{E}} y \quad := \quad \forall n(\mathcal{E}(n) \to (y \in n \to x \in n))$$

Call $FOL(P, \leqslant_{\mathcal{E}})$ the language of FOL enriched with two additional symbols for P and $\leqslant_{\mathcal{E}}$. Note that this is a fragment of the evidence language (enriched with the unary predicate P), namely a fragment where the quantification over subsets occurs only within $\leqslant_{\mathcal{E}}$. The next proposition points to the fact that any update in the language of evidence models that is definable with a formula of $FOL(P, \leqslant_{\mathcal{E}})$ is in fact trackable.

6.5.12. PROPOSITION (DEFINABILITY ENTAILS TRACKING). *Given an update U' on evidence models and a set P, assume $U'(P)$ preserves the domain of the models unchanged. Suppose that for any model M the relation $ORD(U'(P)(\mathcal{E}))$, namely the plausibility relation in the model $ORD(U'(P)(M))$, is definable by a formula $\beta(x, y, P, \leqslant_{\mathcal{E}}) \in FOL(P, \leqslant_{\mathcal{E}})$. Then $U'(P)$ is tracked by an update $U(P)$ defined by $\beta(x, y, P, \leqslant)$.*

Proof:
Given a plausibility model $M = \langle W, \leqslant, V \rangle$, define U to be the identity on the domain and put $U(P)(\leqslant) = \{(a, b) | M \vDash \beta(x, y, P, \leqslant)[(a, b)]\}$. Notice that, having replaced $\leqslant_{\mathcal{E}}$ with $\leqslant$, now β is in the signature of plausibility models. Since all functions involved preserve the carrier, we only need to check that the plausibility relations $ORD(U'(P)(\mathcal{E}))$ and $U(P)(ORD(\mathcal{E}))$ are the same. By assumption, for every evidence model M we have

$$ORD(U'(P)(\mathcal{E})) = \{(a, b) | M \vDash \beta(x, y, P, \leqslant_{\mathcal{E}})[a, b]\}$$

and by construction the latter is equal to $U(P)(\leqslant_{\mathcal{E}})$. By definition of ORD, $U(P)(\leqslant_{\mathcal{E}})$ is the same as $U(P)(ORD(\mathcal{E}))$, so the models $ORD(U'(P)(M))$ and $U(P)(ORD(M))$ coincide. □

If an update $U'(P)$ on evidence models is defined by $\alpha(n, P, \mathcal{E})$ in the evidence language (that is, such formula denotes the subsets that are pieces of evidence after the update) then the relation $ORD(U'(P)(\mathcal{E}))$ is also defined as

$$\forall n(\alpha(n, P, \mathcal{E}) \rightarrow (y \in n \rightarrow x \in n)) \tag{6.1}$$

Therefore the last proposition is guaranteeing that if we can reduce 6.1 to the fragment $FOL(P, \leqslant_{\mathcal{E}})$ then we know that U' is trackable.

6.5.4 Characterizing trackable updates

In the final part of this section we isolate a class of updates for which we can prove a characterization result. We begin with a preliminary definition and some notation.

6.5.13. Definition. [Simple formulas] Given a predicate P on elements, a formula $\psi(n, x, P)$ in the evidence language is *simple* if it is built from the atomic formulas $x \in n$ and Px using only negations, conjunctions and disjunctions.

Simple formulas are essentially just boolean combinations of the two atomic formulas $x \in n$ and Px.

Notation. We use the following abbreviations in the evidence language:

- $n \subseteq n' := \forall y\,[y \in n \rightarrow y \in n']$
- $n = n' := n \subseteq n' \wedge n' \subseteq n$
- $n \subset n' := n \subseteq n' \wedge \neg(n' \subseteq n)$
- $n \subseteq P := \forall y\,[y \in n \rightarrow Py]$
- $n = P := n \subseteq P \wedge P \subseteq n$
- $n \subset P := n \subseteq P \wedge \neg(P \subseteq n)$
- $n \subseteq \varphi(n', P) := \forall y\,[y \in n \rightarrow \varphi(n', y, P)]$
- $n = \varphi(n', P) := (n \subseteq \varphi(n', P)) \wedge (\varphi(n', P) \subseteq n)$
- $n \subset \varphi(n', P) := (n \subseteq \varphi(n', P)) \wedge \neg(\varphi(n', P) \subseteq n)$

Note how we remove the variable y from $\varphi(n', P)$ to stress that this variable has been quantified over. We use the same notation with other formulas such as $\theta(n', x, P)$ in the same fashion.

A first observation, proved in Lemma 6.6.1, is that all updates that are defined with a formula of the following shape

$$\alpha(n, P) := \exists n'(\mathcal{E}(n') \wedge n = \varphi(n', P))$$

turn out to be trackable. In this cases all evidence sets are modified uniformly by φ. An example of such updates is evidence weakening, in which case $\varphi(n', y, P) := x \in n' \vee Px$. We thus seek to enlarge this class of updates to a more diverse one, including some non-trackable updates.

As witnessed by some examples treated in [26], counterexamples to tracking seem to occur when we break this uniformity, that is, we modify some evidence sets while we leave some other unchanged. This suggests the introduction of a 'precondition' θ, which may be triggered or not by an evidence set; to keep things under control we require θ to also be a simple formulas. This leads us to the definition of simple updates.

6.5.14. Definition. [Simple updates] An update on evidence models is *simple* if it is definable with a formula of the following shape:

$$\begin{aligned}\alpha(n, P) :=& \exists n'(\mathcal{E}(n') \wedge \exists x\, \theta(n', x, P) \wedge n = \varphi(n', P)) \vee \\ & (\mathcal{E}(n') \wedge \neg\exists x\, \theta(n', x, P) \wedge n = n')\end{aligned}$$

where both $\theta(n, x, P)$ and $\varphi(n', y, P)$ are simple formulas.

Simple updates can be intuitively understood as follows: the new collection of evidences keeps all the old evidence sets n' for which the condition θ 'fails', namely when $\neg\exists x\, \theta(n', x, P)$ is the case, while it replaces with $\varphi(n', P)$ all the old evidence sets n' for which the precondition $\exists x\, \theta(n', x, P)$ holds. If $\theta(n, x, P)$ is equivalent to $\top$ then a simple update performs a uniform operation on all evidence sets, thus we recover all updates of the form $\alpha(n, P) := \exists n'(\mathcal{E}(n') \wedge n = \varphi(n', P))$. If θ has more structure then it can be exploited to select the evidence sets that we intend to manipulate.

The class of simple updates contains both examples and counterexamples to tracking, therefore it is amenable for a characterization result as described in the previous section. Despite being defined in terms of simple formulas, simple updates already exhibit a certain level of complexity due to the interaction between the 'precondition' θ and the 'effect' of the update φ.

Now for some further terminology. The elements that belong to a subset n but do not belong to its updated version, namely $\varphi(n, x, P)$, are called *separated*. We call *adopted* elements those elements that do not belong to a subset n but belong to $\varphi(n, x, P)$. Finally, we call *witnesses* for n the elements that belong to $\theta(n, x, P)$, since they will be witnesses for the truth of the existential $\exists x\, \theta(n, x, P)$. We encode these notions in the following formulas:

- $Sep(n, x, P) := x \in n \wedge \neg\varphi(n, x, P)$
- $Adop(n, x, P) := x \notin n \wedge \varphi(n, x, P)$

With this terminology in place we can state our characterization result for tracking of simple updates.

6.5.15. THEOREM. *A simple update U' is trackable if and only if one of the following conditions hold.*

(1) All separated points and all adopted points are witnesses: $\forall n\, Sep(n, P) \subseteq \theta(n, P) \wedge Ado(n, P) \subseteq \theta(n, P)$ is a tautology.

(2) The formula $\forall n\, \mathcal{E}(n) \to \forall x(\gamma(n, x, P) \to \theta(n, x, P))$ is a tautology on evidence models, where $\gamma(n, x, P)$ is one of the following formulas:

- $x \in n$
- $x \notin n$
- Px
- $\neg Px$

(3) $\exists x\, \theta(n, x, P)$ is equivalent to $\perp$.

If one of the aforementioned conditions hold then we have a procedure to construct the tracking companion of U'; if they all fail we have a procedure to construct a counterexample to tracking.

The proof of this result is given in details in the next section.

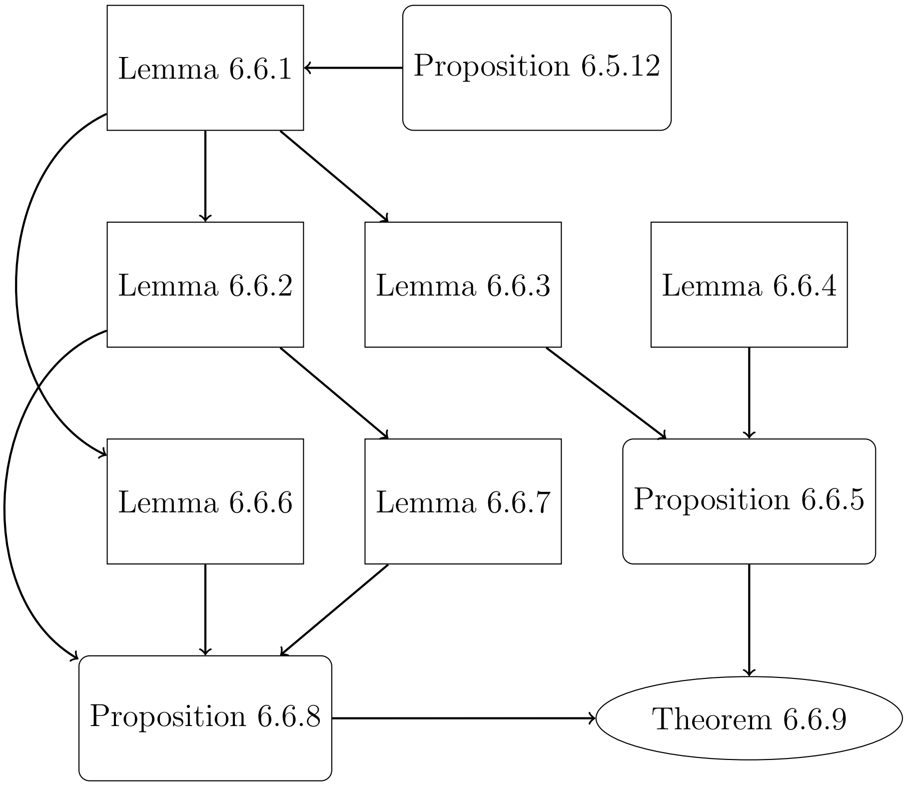
6.6 Proof of Theorem 6.5.15

We split the proof by addressing one direction at the time. We first show that each of the premises of Theorem 6.5.15 is sufficient for tracking, while in the following subsection we demonstrate that they are actually necessary, meaning that if they all fail then we can concoct a counterexample to tracking.

6.6.1 Sufficient conditions for tracking of simple updates

The crux of this half of the proof is a procedure to re-arrange and eliminate second-order quantifiers, enabling us to reduce the definition of a simple update to a formula in $FOL(P, \leqslant_{\mathcal{E}})$. After this is achieved, the application of Proposition 6.5.12 ensures that the update is trackable.

We provide a flowchart to help the reader navigate through the different parts of the proof. Recall that Proposition 6.5.12 was proved in the previous section, while Theorem 6.6.9 constitute one direction of Theorem 6.5.15.



The first crucial observation is that, for φ and θ simple formulas appearing in the definition of a simple update, if θ is trivial then any choice of φ will give a trackable update.

6.6.1. Lemma. *Given an update U' on evidence models and a set P, assume $U'(P)$ preserves the domain of the models unchanged. If U' is defined by a formula $\alpha(n, P, \mathcal{E}) := \exists n'(\mathcal{E}(n') \wedge \forall y\,[y \in n \leftrightarrow \varphi(n', y, P)])$ where $\varphi(n', y, P)$ is a simple formula then U' is trackable and we have an effective procedure to find its tracking companion.*

Proof:
We show that, for α of this particular shape, the formula

$$\forall n(\alpha(n, P, \mathcal{E}) \to (y \in n \to x \in n))$$

can be reduce to the fragment $FOL(P, \leqslant_{\mathcal{E}})$; this triggers Proposition 6.5.12 and gives the desired result. We first rearrange the formula into an equivalent one

$$\begin{aligned}
&\forall n(\alpha(n, P, \mathcal{E}) \to (y \in n \to x \in n)) = \\
&\forall n(\exists n'(\mathcal{E}(n') \wedge \forall z\,[z \in n \leftrightarrow \varphi(n', z, P)]) \to (y \in n \to x \in n)) = \\
&\forall n, \forall n'([\mathcal{E}(n') \wedge \forall z\,[z \in n \leftrightarrow \varphi(n', z, P)]] \to (y \in n \to x \in n)) = \\
&\forall n'(\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))
\end{aligned}$$

We show that the formula in the last line can be reduced to $FOL(P, \leqslant_{\mathcal{E}})$ by considering the structure of φ. We take φ to be in disjunctive normal form, that is, we take it to be a disjunction of conjunctions of literals, namely atomic formulas or their negations.

If $\varphi(n', z, P) = z \in n'$ then the formula becomes

$$\forall n'(\mathcal{E}(n') \to (y \in n' \to x \in n'))$$

which is by definition just $x \leqslant_{\mathcal{E}} y$. If $\varphi(n', z, P) = z \notin n'$ then the formula becomes

$$\begin{aligned} &\forall n'(\mathcal{E}(n') \to (y \notin n' \to x \notin n')) = \\ &\forall n'(\mathcal{E}(n') \to (x \in n' \to y \in n')) \end{aligned}$$

which is by definition just $y \leqslant_{\mathcal{E}} x$.[8]

If $\varphi(n', z, P) = Pz$ then the formula becomes

$$\begin{aligned} &\forall n'(\mathcal{E}(n') \to (Pz \to Pz)) = \\ &\exists n'\mathcal{E}(n') \to (Pz \to Pz) = \\ &Pz \to Pz \end{aligned}$$

where the last inference is due to the fact that $\exists n'\mathcal{E}(n')$ is always true on evidence models: there is always at least one evidence set, the whole carrier W. The formula $Pz \to Pz$ is clearly in $FOL(P, \leqslant_{\mathcal{E}})$. The case $\varphi(n', z, P) = \neg Pz$ is analogous. This covers the case of the literals.

Now suppose φ is a conjunction of literals. Since we have only two atomic formulas, without loss of generality we can assume $\varphi(n', z, P) = \varphi_1(n', z, P) \wedge \varphi_2(n', z, P)$ where both conjuncts are literals. If the conjunction is equivalent to $\bot$ then the formula

$$\forall n'(\mathcal{E}(n') \to (\bot \to \bot))$$

is equivalent to $\top$ and we are done. We thus assume wlog that $\varphi_1(n', z, P)$ is either Pz or $\neg Pz$ and $\varphi_2(n', z, P)$ is either $z \in n'$ or $z \notin n'$ (if they are the same then we fall back into the base case of the literals). We then have

$$\begin{aligned} &\forall n'(\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P))) = \\ &\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P) \wedge \varphi_2(n', y, P)] \to [\varphi_1(n', x, P) \wedge \varphi_2(n', x, P)])) = \\ &\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P) \wedge \varphi_2(n', y, P)] \to [\varphi_1(n', x, P)])) \wedge \\ &\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P) \wedge \varphi_2(n', y, P)] \to [\varphi_2(n', x, P)])) \end{aligned}$$

In the last step we have split the conjunction on the consequent into a conjunction of two implications and also distributed the universal quantifiers over the conjunction.

[8]This last line shows that the update "take the complement of all evidence sets" is tracked by the update on plausibility models that reverses the plausibility order.

We thus have two conjuncts that can now be reduced separately. Consider the first one

$$\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P) \wedge \varphi_2(n', y, P)] \to \varphi_1(n', x, P)))$$

Since $\varphi_1(n', z, P)$ is either Pz or $\neg Pz$, we can extract $\varphi_1(n', y, P)$ from the implication because n' is not free in it. We obtain

$$\varphi_1(n', y, P) \to \forall n'(\mathcal{E}(n') \to (\varphi_2(n', y, P) \to \varphi_1(n', x, P)))$$

Now we can flip the innermost implication

$$\varphi_1(n', y, P) \to \forall n'(\mathcal{E}(n') \to (\neg\varphi_1(n', x, P) \to \neg\varphi_2(n', y, P)))$$

and again extract $\neg\varphi_1(n', x, P)$ from the quantifier, since n' is not free in this formula (recall $\varphi_1(n', z, P)$ is either Pz or $\neg Pz$).

$$\varphi_1(n', y, P) \to \neg\varphi_1(n', x, P) \to \forall n'(\mathcal{E}(n') \to \neg\varphi_2(n', y, P)) \tag{6.2}$$

The first two antecedent are already in $FOL(P, \leqslant_{\mathcal{E}})$; as for the consequent $\forall n'(\mathcal{E}(n') \to \neg\varphi_2(n', y, P))$ we have two cases: either $\varphi_2(n', y, P)$ is $y \in n'$ or $y \notin n'$. In the first case we get $\forall n'(\mathcal{E}(n') \to \neg y \in n')$ which is equivalent to $\bot$ on evidence models, because the whole carrier is always a piece of evidence. Thus in this case the whole formula 6.2 becomes $\top$. In the second case we get $\forall n'(\mathcal{E}(n') \to y \in n')$ which on evidence models is equivalent to $\forall x\, y \leqslant_{\mathcal{E}} x$. Thus formula 6.2 becomes

$$\varphi_1(n', y, P) \to \neg\varphi_1(n', x, P) \to \forall x\, y \leqslant_{\mathcal{E}} x$$

which is in $FOL(P, \leqslant_{\mathcal{E}})$.

We now consider the other conjunct, namely

$$\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P) \wedge \varphi_2(n', y, P)] \to \varphi_2(n', x, P)))$$

Again we extract $\varphi_1(n', y, P)$ from the implication

$$\varphi_1(n', y, P) \to \forall n'(\mathcal{E}(n') \to (\varphi_2(n', y, P) \to \varphi_2(n', x, P)))$$

If $\varphi_2(n', z, P) = z \in n'$ then the consequent is just $x \leqslant_{\mathcal{E}} y$, if $\varphi_2(n', y, P) = y \notin n'$ then the consequent becomes $y \leqslant_{\mathcal{E}} x$. In both cases we have successfully reduced the formula to $FOL(P, \leqslant_{\mathcal{E}})$. This concludes the case of the conjunctions of literals.

Finally, suppose φ is a disjunction of conjunctions of literals, $\varphi(n', z, P) = \varphi_1(n', z, P) \vee \varphi_2(n', z, P)$. It is easy to show that, since there are only two atomic formulas, every such disjunction is equivalent to a disjunction with two disjuncts. We can thus assume that $\varphi_1(n', z, P)$ and $\varphi_2(n', z, P)$ are themselves conjunctions of literals.

Consider the following manipulation

$$\forall n'(\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P))) =$$
$$\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P) \vee \varphi_2(n', y, P)] \to [\varphi_1(n', x, P) \vee \varphi_2(n', x, P)])) =$$
$$\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P) \to [\varphi_1(n', x, P) \vee \varphi_2(n', x, P)]]$$
$$\wedge [[\varphi_2(n', y, P)] \to [\varphi_1(n', x, P) \vee \varphi_2(n', x, P)]]])) =$$
$$\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P)] \to [\varphi_1(n', x, P) \vee \varphi_2(n', x, P)]))\wedge$$
$$\forall n'(\mathcal{E}(n') \to ([\varphi_2(n', y, P)] \to [\varphi_1(n', x, P) \vee \varphi_2(n', x, P)]))$$

where in the second-to-last step we split the disjunction in the antecedent using the law $[(p \vee q) \to r] \leftrightarrow [(p \to r) \wedge (q \to r)]$, and in the last step we split the whole implication over this conjunction and then distribute the universal quantifier. Now consider the first conjunct (the procedure is symmetric for the other conjunct):

$$\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P)] \to [\varphi_1(n', x, P) \vee \varphi_2(n', x, P)]))$$
$$\forall n'(\mathcal{E}(n') \to ([\varphi_1(n', y, P)] \to [(\neg\varphi_1(n', x, P)) \to \varphi_2(n', x, P)]))$$

Since $\varphi_1(n', x, P)$ is a conjunction, its negation will be a disjunction; since such disjunction is the antecedent of the innermost implication we can repeat the last manipulation: split this disjunction using the law $[(p \vee q) \to r] \leftrightarrow [(p \to r) \wedge (q \to r)]$, then split the whole implication and distribute the universal quantifier as before. At the end of this process we have removed all the disjunctions and we are left with conjunctions of literals appearing in the inner-most consequent and in some antecedents. This is then handled with the procedure we described above in the conjunction case. □

6.6.2. Lemma. *Suppose U' is a simple update on evidence models defined by $\alpha(n, P)$ and assume that on evidence model $\exists x\, \theta(n, x, P)$ is equivalent to $\top$ for any n, where θ is the simple formula featuring in α. Then U' is a trackable update and we have an effective procedure to obtain its tracking companion.*

Proof:
In this circumstance, since $\neg\exists x\, \theta(n', x, P)$ is equivalent to $\bot$, the formula defining the update becomes

$$\alpha(n, P) := \exists n'(\mathcal{E}(n') \wedge \forall y\, [y \in n \leftrightarrow \varphi(n', y, P)])$$

Since $\varphi(n', y, P)$ is simple, by Lemma 6.6.1 we have an algorithm to obtain the corresponding update on plausibility models. □

6.6.3. Lemma. *If, for any model and any subset* n*,* $\neg\exists x\,\theta(n,x)$ *entails* $n = \varphi(n,P)]$ *then the update is trackable.*

Proof:
Recall that we abbreviate $\forall z\,[z \in n \leftrightarrow \varphi(n',z,P)]$ with $n = \varphi(n')$. Under the assumption of the statement the definition of the update can be transformed

$$\begin{aligned}\alpha(n,P) :=& \exists n'(\mathcal{E}(n') \wedge \exists x\,\theta(n',x,P) \wedge n = \varphi(n'))\vee \\ & (\mathcal{E}(n') \wedge \neg\exists x\,\theta(n',x,P) \wedge n = n') \\ =& \exists n'(\mathcal{E}(n') \wedge \exists x\,\theta(n',x,P) \wedge n = \varphi(n'))\vee \\ & (\mathcal{E}(n') \wedge \neg\exists x\,\theta(n',x,P) \wedge n = n' \wedge n' = \varphi(n')) \\ =& \exists n'(\mathcal{E}(n') \wedge \exists x\,\theta(n',x,P) \wedge n = \varphi(n'))\vee \\ & (\mathcal{E}(n') \wedge \neg\exists x\,\theta(n',x,P) \wedge n = \varphi(n')) \\ =& \exists n'(\mathcal{E}(n') \wedge n = \varphi(n')) \\ =& \exists n'(\mathcal{E}(n') \wedge \forall y\,[y \in n \leftrightarrow \varphi(n,y,P)])\end{aligned}$$

By Lemma 6.6.1 we know the last line defines a trackable update. □

The next Lemma states that $\varphi(n,x,P)$, when n is fixed, is the set of elements contained in n plus the adopted elements and minus the separated elements.

6.6.4. Lemma. *The following formula is a validity:*

$$\forall n\,\forall x\,\varphi(n,x,P) \leftrightarrow (x \in n \vee \mathit{Ado}(n,x,P)) \wedge \neg \mathit{Sep}(n,x,P)$$

Proof:
Directly by the definitions. □

The next proposition explains what kind of connection between θ and φ is sufficient to ensure tracking.

6.6.5. Proposition. *Consider a simple update* U' *and the* φ *and* θ *in its definition. If* $\forall n\,\forall x\,\mathit{Sep}(n,x,P) \rightarrow \theta(n,x,P)$ *and* $\forall n\,\forall x\,\mathit{Ado}(n,x,P) \rightarrow \theta(n,x,P)$ *are both valid on evidence models then the update is tracked.*

Proof:
If $\forall n\,\forall x\,\mathit{Sep}(n,x,P) \rightarrow \theta(n,x,P)$ and $\forall n\,\forall x\,\mathit{Ado}(n,x,P) \rightarrow \theta(n,x,P)$ are both valid on evidence models then $\neg\exists x\,\theta(n,x,P)$ entails both $\neg\exists x\,\mathit{Sep}(n,x,P)$ and $\neg\exists x\,\mathit{Ado}(n,x,P)$. The latter two formulas, together with Lemma 6.6.4 entail that $\forall n\,\forall x\,\varphi(n,x,P) \leftrightarrow x \in n$.

We have thus showed that $\neg\exists x\,\theta(n,x,P)$ entails $\forall n\,\varphi(n,P) = n$. This triggers the premise of Lemma 6.6.3, from which we conclude that U' is trackable. □

6.6.6. LEMMA. *Let U' be a simple update. If the $\theta(n, x, P)$ featuring in the definition of U' is equivalent to either Px or $\neg Px$ then the update is trackable.*

Proof:
Suppose $\theta(n, x, P)$ is equivalent to Px. Then the update is defined by

$$\begin{aligned}\alpha(n, P) :=& \exists n'(\mathcal{E}(n') \wedge \exists x\, Px \wedge n = \varphi(n', P)) \vee \\ & (\mathcal{E}(n') \wedge \neg\exists x\, Px \wedge n = n') = \\ & \exists n'(\mathcal{E}(n') \wedge \exists x\, Px \wedge n = \varphi(n', P)) \vee \\ & \exists n'(\mathcal{E}(n') \wedge \neg\exists x\, Px \wedge n = n') = \\ & [\exists x\, Px \wedge \exists n'(\mathcal{E}(n') \wedge n = \varphi(n', P))] \vee \\ & [\neg\exists x\, Px \wedge \exists n'(\mathcal{E}(n') \wedge n = n')]\end{aligned}$$

We can now plug this latter $\alpha(n, P)$ into ORD and perform some manipulation of first-order logic.

$$\begin{aligned}& \forall n([\exists x\, Px \wedge \exists n'(\mathcal{E}(n') \wedge n = \varphi(n', P)) \vee \\ & \neg\exists x\, Px \wedge \exists n'(\mathcal{E}(n') \wedge n = n')] \rightarrow (y \in n \rightarrow x \in n)) = \\ & \forall n([\exists x\, Px \wedge \exists n'(\mathcal{E}(n') \wedge n = \varphi(n', P))] \rightarrow (y \in n \rightarrow x \in n)) \wedge \\ & \forall n([\neg\exists x\, Px \wedge \exists n'(\mathcal{E}(n') \wedge n = n')] \rightarrow (y \in n \rightarrow x \in n)) = \\ & \exists x\, Px \rightarrow \forall n([\exists n'(\mathcal{E}(n') \wedge n = \varphi(n', P))] \rightarrow (y \in n \rightarrow x \in n)) \wedge \\ & \neg\exists x\, Px \rightarrow \forall n([\exists n'(\mathcal{E}(n') \wedge n = n')] \rightarrow (y \in n \rightarrow x \in n))\end{aligned}$$

Note that the formulas

$$\forall n([\exists n'(\mathcal{E}(n') \wedge n = \varphi(n', P))] \rightarrow (y \in n \rightarrow x \in n))$$

and

$$\forall n([\exists n'(\mathcal{E}(n') \wedge n = n')] \rightarrow (y \in n \rightarrow x \in n))$$

in the consequents are both reducible thanks to Lemma 6.6.1, thus the whole formula is reducible. A similar argument covers the case of $\neg Px$. □

6.6.7. LEMMA. *If $\forall n\, \mathcal{E}(n) \rightarrow \forall x(x \notin n \rightarrow \theta(n, x, P))$ is a tautology on evidence models, then the formula*

$$\forall n'[[(\mathcal{E}(n') \wedge \exists z \theta(n', z, P))] \rightarrow (\varphi(n', y, P) \rightarrow \varphi(n', x, P))]$$

is reducible to $FOL(P, \leqslant_{\mathcal{E}})$.

Proof:
Since $\forall n\, \mathcal{E}(n) \rightarrow \forall x(x \notin n \rightarrow \theta(n, x, P))$ is a tautology it must be that $\theta(n, x, P)$ is $x \notin n$ or (equivalent to) a disjunction of $x \notin n$ and another literal ξ. If $\xi = x \in n$

then θ is a tautology and Lemma 6.6.2 gives us the desired result. If $\xi = Px$ then we have

$$\begin{aligned}
&\forall n'[[(\mathcal{E}(n') \land \exists z[z \notin n' \lor Pz])] \to (\varphi(n', y, P) \to \varphi(n', x, P))] = \\
&\forall n'[\exists z[z \notin n' \lor Pz] \to (\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))] = \\
&\forall n', \forall z[[z \notin n' \lor Pz] \to (\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))] = \\
&\forall n', \forall z[z \notin n' \to (\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))] \land \\
&\forall n', \forall z[Pz \to (\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))] = \\
&\forall n', [\exists z\, z \notin n' \to (\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))] \land \\
&\forall n'[\exists z\, Pz \to (\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))] = \\
&\forall n', [\exists z\, z \notin n' \to (\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))] \land \\
&\exists z\, Pz \to \forall n'[(\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))]
\end{aligned}$$

Note that the second line of the last formula is reducible to $FOL(P, \leqslant_\mathcal{E})$: $\exists z\, Pz$ is already in $FOL(P, \leqslant_\mathcal{E})$ and the consequent is reducible thanks to Lemma 6.6.2. The same argument works for $\xi = \neg Px$, so we only need to argue that the first line of the last formula is reducible and we are done, since this also covers the case of $\theta(n, x, P)$ being $x \notin n$.

To reduce $\forall n', [\exists z\, z \notin n' \to (\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))]$ we first apply the same procedure described in Lemma 6.6.2 until we have pushed out from the implication all literals of shape Px' or $\neg Px'$ (where x' could be either x or y). We will then be left with a number of implications whose consequents will then have one of the following four shapes (where $\ldots$ stand for the formulas that we pushed out of the quantification over n'):

(1) $\cdots \to \forall n'\, [\exists z\, z \notin n' \land \mathcal{E}(n') \to (y \in n' \to x \in n'))]$;

(2) $\cdots \to \forall n'\, [\exists z\, z \notin n' \land \mathcal{E}(n') \to (y \notin n' \to x \notin n'))]$;

(3) $\cdots \to \forall n'\, [\exists z\, z \notin n' \land \mathcal{E}(n') \to (x' \in n')]$, (where x' could be either x or y);

(4) $\cdots \to \forall n'\, [\exists z\, z \notin n' \land \mathcal{E}(n') \to (x' \notin n')]$, (where x' could be either x or y).

Notice that the first consequent is equivalent to $\forall n', [\mathcal{E}(n') \to (y \in n' \to x \in n'))]$: the only set for which the antecedent does not hold is the whole set, and in that case the consequent $(y \in n' \to x \in n')$ holds trivially. Thus the first case reduces to $x \leqslant_\mathcal{E} y$.

The second consequent $\forall n'\, [\exists z\, z \notin n' \land \mathcal{E}(n') \to (y \notin n' \to x \notin n'))]$ is tantamount to $\forall n', [\exists z\, z \notin n' \land \mathcal{E}(n') \to (x \in n' \to y \in n'))]$; for the same argument as above this reduces to $y \leqslant_\mathcal{E} x$.

Now for the third consequent $\forall n'\, [\exists z\, z \notin n' \land \mathcal{E}(n') \to (x' \in n')]$. Since x' trivially also belongs to the whole set, this formula is equivalent to $\forall n', [\mathcal{E}(n') \to (x' \in n')]$, which is in turn equivalent to $\forall z\, x' \leqslant_\mathcal{E} z$. Hence this part also reduces to $FOL(P, \leqslant_\mathcal{E})$.

Finally, $\forall n' [\exists z\, z \notin n' \wedge \mathcal{E}(n') \to (x' \notin n')]$ is equivalent to $\forall n' [x' \in n' \to (\mathcal{E}(n') \to \forall z\, z \in n')$. A little reflection shows that on evidence models this last formula is equivalent to $\forall z\, z \leqslant_{\mathcal{E}} x'$: if the formula is the case then, for any z and any evidence X, if $x' \in X$ then X must the whole carrier so $z \in X$, so $\forall z\, z \leqslant_{\mathcal{E}} x'$. On the other hand if $x' \in X$ for some evidence X then $z \leqslant_{\mathcal{E}} x'$ entails $z \in X$.Since $\forall z\, z \leqslant_{\mathcal{E}} x'$ is in $FOL(P, \leqslant_{\mathcal{E}})$, the last implication has been reduced successfully. This conclude the explanation on how to reduce $\forall n', [\exists z\, z \notin n' \to (\mathcal{E}(n') \to (\varphi(n', y, P) \to \varphi(n', x, P)))]$ and also concludes the main case distinction. □

6.6.8. Proposition. *Let U' be a simple update and $\theta(n, x, P)$ be the simple formula featuring in its definition. If $\forall n\, \mathcal{E}(n) \to \forall x(\gamma(n, x, P) \to \theta(n, x, P))$ is a tautology on evidence models, where $\gamma(n, x, P)$ is one of the following formulas:*

- $x \in n$
- $x \notin n$
- Px
- $\neg Px$

then U' is trackable.

Proof:
For the first item, recall that the assumption on evidence models is that evidence sets are non-empty, hence $\exists x\, x \in n$ is always true. Thus $\exists x\, \theta(n, x, P)$ must also be a tautology and therefore Lemma 6.6.2 applies. For the second item, we have that $\neg\exists x\theta(n, x, P)$ entails $\neg\exists z\, z \notin n = \forall z\, z \in n$, therefore the second part of the definition of α only applies to the whole set W. Then the update definition can be rearranged as follows.

$$\begin{aligned}
\alpha(n, P) :=& \exists n'(\mathcal{E}(n') \wedge \exists x\theta(n', x, P) \wedge n = \varphi(n', P)]) \vee \\
& \exists n'(\mathcal{E}(n') \wedge \neg\exists x\theta(n', x, P) \wedge n = n') \\
=& \exists n'(\mathcal{E}(n') \wedge \exists x\theta(n', x, P) \wedge n = \varphi(n', P)]) \vee \\
& \exists n'(\mathcal{E}(n') \wedge \neg\exists x\theta(n', x, P) \wedge \forall z\,(z \in n') \wedge n = n') \\
=& \exists n'(\mathcal{E}(n') \wedge \exists x\theta(n', x, P) \wedge n = \varphi(n', P)]) \vee \\
& \exists n'(\mathcal{E}(n') \wedge \neg\exists x\theta(n', x, P) \wedge \forall z\,(z \in n))
\end{aligned}$$

The second part of this formula is essentially stating that n is the whole set. Now plug this into the definition of ORD:

$$\begin{aligned}ORD(\alpha) :=& \forall n\, \alpha(n, P) \rightarrow (y \in n \rightarrow x \in n)\\ =& \forall n[\exists n'(\mathcal{E}(n') \wedge \exists x\theta(n', x, P) \wedge n = \varphi(n', P)]) \vee \\ & \exists n'(\mathcal{E}(n') \wedge \neg\exists x\theta(n', x, P) \wedge \forall z\, (z \in n))] \rightarrow (y \in n \rightarrow x \in n)\\ =& \forall n[[\exists n'(\mathcal{E}(n') \wedge \exists x\theta(n', x, P) \wedge n = \varphi(n', P))] \rightarrow (y \in n \rightarrow x \in n)] \wedge \\ & \forall n[[\exists n'(\mathcal{E}(n') \wedge \neg\exists x\theta(n', x, P) \wedge \forall z\, (z \in n))] \rightarrow (y \in n \rightarrow x \in n)]\end{aligned}$$

Notice that the second conjunction is a tautology: if n is the whole set then $(x \in n \rightarrow y \in n)$ is always the case. Hence the last line of this derivation is equivalent to

$$\forall n[[\exists n'(\mathcal{E}(n') \wedge \exists x\theta(n', x, P) \wedge n = \varphi(n', P))] \rightarrow (y \in n \rightarrow x \in n)]$$

Which is in turn equivalent to

$$\forall n'[[(\mathcal{E}(n') \wedge \exists x\theta(n', x, P))] \rightarrow (\varphi(n', y, P) \rightarrow \varphi(n', x, P))]$$

Thanks to Lemma 6.6.7 this last formula is reducible to $FOL(P, \leqslant_{\mathcal{E}})$, thus the update is trackable.

Now for the third case. If $\exists x\theta(n, x, P)$ is equivalent to $\exists x Px$ then by Lemma 6.6.6 the update is trackable. So suppose $\exists x Px$ entails $\exists x\theta(n, x, P)$ but not viceversa. If the extension of P is strictly contained in $\theta(n, x, P)$ then, since θ is a simple formula, a simple induction in the structure of θ shows that either $\exists x\, x \in n$ entails $\exists x\theta(n, x, P)$ or $\exists x\, x \notin n$ entails $\exists x\theta(n, x, P)$. In both cases we have already proved that the update is trackable. An analogous argument covers the case of $\exists x\, \neg Px$. □

6.6.9. THEOREM. *Suppose an update U' is simple. If one of the following three options is the case then the update is tracked and we have an effective procedure to compute the corresponding update on plausibility models.*

(1) All separated points and all adopted points are witnesses: $\forall n\, Sep(n, P) \subseteq \theta(n, P) \wedge Ado(n, P) \subseteq \theta(n, P)$ is a tautology.

(2) The formula $\forall n\, \mathcal{E}(n) \rightarrow \forall x(\gamma(n, x, P) \rightarrow \theta(n, x, P))$ is a tautology on evidence models, where $\gamma(n, x, P)$ is one of the following formulas:

- $x \in n$
- $x \notin n$
- Px
- $\neg Px$

(3) $\exists x\, \theta(n, x, P)$ is equivalent to $\bot$.

Proof:
The first two items entail tracking directly by Proposition 6.6.8 and Proposition 6.6.5. In the third case the update reduces to the identity, which is tracked by the identity on plausibility models. The proofs of the propositions and related lemmas explain how to manipulate to syntactic shape of the update to obtain the update on plausibility models. □

6.6.2 Necessary conditions for tracking of simple updates

We begin by noticing that the failure of premise 2 and 3 of Theorem 6.6.9 (which are the same as those in Theorem 6.5.15) enforces a particular behaviour for the sets P and $\neg P$: either one or the other are left untouched by the update. Moreover, either the subsets or the supersets of P (respectively, $\neg P$) trigger the precondition θ.

6.6.10. LEMMA. *Let U' be a simple update. Suppose $\theta(n, x, P)$ is not entailed by any of the following formulas:*

- $x \in n$
- $x \notin n$
- Px
- $\neg Px$

and $\exists x\, \theta(n, x, P)$ is not equivalent to $\bot$. Then one of four cases must occur, for any model.

- *The update acts as the identity on P: $\forall n\, (n = P \to \neg\exists x\, \theta(n, x, P))$ is a tautology on evidence models. Moreover, either*
  - *Strict subsets of P trigger θ: $\exists n\, (n \subset P \to \exists x\, \theta(n, x, P))$ is a tautology on evidence models.*
  - *Strict supersets of P trigger θ: $\exists n\, (P \subset n \to \exists x\, \theta(n, x, P))$ is a tautology on evidence models.*
- *The update acts as the identity on $\neg P$: $\forall n\, (n = \neg P \to \neg\exists x\, \theta(n, x, P))$ is a tautology on evidence models. Moreover, either*
  - *Strict subsets of $\neg P$ trigger θ: $\exists n\, (n \subset \neg P \to \exists x\, \theta(n, x, P))$ is consistent on evidence models, or*
  - *Strict supersets of $\neg P$ trigger θ: $\exists n\, (\neg P \subset n \to \exists x\, \theta(n, x, P))$ is consistent on evidence models.*

Proof:
By the fact that θ is simple, under the restrictions above only six possibilities are left for $\theta(n, x, P)$:

(1) $x \in n \wedge Px$

(2) $x \in n \wedge \neg Px$

(3) $x \notin n \wedge Px$

(4) $x \notin n \wedge \neg Px$

(5) $(x \in n \wedge Px) \vee x \notin n \wedge \neg Px$

(6) $(x \in n \wedge \neg Px) \vee x \notin n \wedge Px$

In cases 2,3 and 6 if $\forall z\,(z \in n \leftrightarrow Pz)$ is the case then $\theta(n, x, P)$ becomes a contradiction, thus $\neg\exists\theta(n, x, P)$ will hold. In cases 3 and 6 the formula $n \subset P$ entails $\exists x\, \theta(n, x, P)$; in cases 2 and 6 the formula $P \subset n$ entails $\exists x\, \theta(n, x, P)$.

In cases 1,4 and 5 if $\forall z\,(z \in n \leftrightarrow \neg Pz)$ is the case then again $\theta(n, x, P)$ becomes a contradiction, hence $\neg\exists\theta(n, x, P)$ does hold. For cases 4 and 5 we have that $n \subset \neg P$ entails $\exists x\, \theta(n, x, P)$, while in cases 1 and 5 $\neg P \subset n$ entails $\exists x\, \theta(n, x, P)$. □

We can then use this observation together with the failure of the first premise of Theorem 6.6.9 to build a counterexample to tracking.

6.6.11. THEOREM. *Let U' be a simple update for which all the following assumptions do* not *hold:*

(1) All separated points and all adopted points are witnesses: $\forall n\, Sep(n, P) \subseteq \theta(n, P) \wedge Ado(n, P) \subseteq \theta(n, P)$ is a tautology.

(2) $\exists x\, \theta(n, x, P)$ is entailed by one of the following formulas:

- $\exists x\, x \in n$
- $\exists x\, x \notin n$
- $\exists x\, Px$
- $\exists x\, \neg Px$

(3) $\exists x\, \theta(n, x, P)$ is equivalent to $\perp$.

then the update is not trackable.

Proof:
For any update that does not satisfy all the above requirements we build a counterexample to tracking in the fashion of Proposition 6.5.3: we construct two evidence models that produce the same plausibility models and show that, after the update, the two new models generate different plausibility models.

Due to Lemma 6.6.10 we know that if assumption 2 and 3 do not hold then one of four cases must occur. Proposition 6.6.12 explains how to construct a counterexample to tracking in the first two cases. The counterexamples for the remaining cases are obtained symmetrically by switching the sets P and $\neg P$. □

The next proposition describes how to construct counterexamples to tracking. The proof is a rather long case distinction. The good news is that only five counterexamples are enough to prove the non-trackability of all the aforementioned updates - although the same counterexample might work for different reasons for different updates.

Moreover, every counterexample is based on pair of models both having four elements. The reason why four elements are sufficient is that, from a semantic point of view, simple formulas for a fixed subset n are essentially constructed by taking unions of the four areas $n \cap P$, $n \cap \neg P$, $\neg n \cap P$ and $\neg n \cap \neg P$. Thus every two points that belong to the same area behave uniformly with respect to the same n: either they will be both witnesses for $\exists x \theta(n, x, P)$ or neither of them will be, they will both be separated for n (or adopted) or neither will be, and so on. Thus having four elements, two of which are in P, gives us enough leeway to construct evidence sets that stand in all possible relationships with P while at the same time interacting with the simple formulas θ and φ.[9]

6.6.12. PROPOSITION. *Let U' be a simple update that does not fulfill item 1, 2 and 3 of Theorem 6.6.11. Suppose moreover that*

$$\forall n\, (n = P \rightarrow \neg \exists x\, \theta(n, x, P))$$

is a tautology on evidence models. Then U' is not trackable.

Proof:
Suppose $\forall n\, (n = P \rightarrow \neg \exists x\, \theta(n, x, P))$ is a tautology on evidence models. This means that the set corresponding to P is left untouched by the update: if P is a piece of evidence then $\neg \exists x\, \theta(n, x, P)$ is the case so the update acts as the identity on it. This means that, for any n, there are no witnesses for $\theta(n, x, P)$ in $\neg n \cap \neg P$ and $n \cap P$, or otherwise there would be models where P would trigger $\exists x\, \theta(n, x, P)$. Moreover we know that either of the following holds.

[9] In fact, some counterexamples also work by taking models with three elements, e.g. some instances of Counterexample 1 and 3, depending on θ and φ. We decided that a uniform presentation of the counterexamples was preferable in order to simplify the case distinction.

(1) Strict subsets of P trigger θ: $\exists n\,(n \subset P \to \exists x\,\theta(n, x, P))$ is a tautology on evidence models.

(2) Strict supersets of P trigger θ: $\exists n\,(P \subset n \to \exists x\,\theta(n, x, P))$ is a tautology on evidence models.

Case 1. Let us consider the first scenario (call this **assumption (a)**). If strict subsets of P do trigger $\exists x\,\theta(n, x, P)$ then this means that elements in $\neg n \cap P$ must be witnesses. Moreover, if item 1 fails then for some n there is either an adopted point that is not a witness or a separated point that is not a witness.

- In the first case, the adopted points that are not witnesses must be in $\neg n \cap \neg P$, since points in $\neg n \cap P$ are witnesses and adopted points must lie outside of n. Thus any strict subset n of P will trigger $\exists x\,\theta(n, x, P)$ due to the witnesses in $\neg n \cap P$ and adopt all the points in $\neg n \cap \neg P$ after the update, while P will remain stable under the update. These considerations suggest that the two evidence models depicted on the left and right side of the figure Counterexample 1 will produce a counterexample to tracking. We draw the existing evidence in both models as well as the propositional variable p used to interpret the unary predicate P.

Figure 6.1: Counterexample 1, before the update

Before the update, in Figure 6.1, the evidence models in Counterexample 1 produce the same plausibility model, as the reader can easily check. After the update, in Figure 6.2, the strict subsets of P have adopted all the $\neg p$-worlds, while P remained stable. These two models generate different plausibility relations, because on the left the p-worlds are separated from the others by an evidence set and on the right-hand side they are not.

- Consider now the other case in which there are separated points that are not witnesses (call this **assumption (b)**). Suppose there are no adopted points. If the elements in $n \cap P$ are separated and not witnesses then the strict subsets of P will disappear after the update. Thus the models of Counterexample 1 will still work, albeit for a different reason: after the

Figure 6.2: Counterexample 1, after the update

Figure 6.3: Counterexample 1, after the update (different case)

update we will have the two models of Figure 6.3, which clearly generate different plausibility relations.

If the separated points that are not witnesses are in $n \cap \neg P$ then the subsets of $\neg P$ will disappear. Consider the two models of Figure 6.4. They produce the same ordering, since the additional evidence on the left is just the union of the other evidence sets.

Figure 6.4: Counterexample 2, before the update

After the update we obtain the models depicted in Figure 6.5. Due to the update the singleton evidence disappears, since points in $\neg n \cap P$ trigger θ and points in $n \cap \neg P$ are separated (and we assumed there are no adopted points). The other two sets will survive unchanged, since they contain all the P-points they cannot trigger θ (since we are under the assumption that elements in $n \cap \neg P$ are not witnesses). These two models clearly generate a different ordering between the two right-most points.

Figure 6.5: Counterexample 2, after the update

Thus we are left with the case where there are separated points that are not witnesses, there are adopted points and all adopted points are witnesses (call this **assumption (c)**). Together with assumption (a), namely that strict subsets of P trigger θ, this means that the points in $\neg n \cap P$ are adopted and witnesses. If the elements that are separated and not witnesses are in $n \cap \neg P$ then Counterexample 2 still works: the singleton will be sent to P while the other two sets will remain unchanged.

So finally suppose that the elements that are separated and not witnesses are in $n \cap P$ (**assumption (d)**). Again we have two cases. If elements in $n \cap \neg P$ are also separated then they must be witnesses or otherwise the previous case applies. But this means that supersets of P have all their elements separated, i.e. they are mapped to the empty set by the update. Hence the two models of Figure 6.6 will give a counterexample, since both supersets of P get erased by the update while P is left untouched.

Figure 6.6: Counterexample 3, before the update

These two models produce the same ordering, since the additional evidence set on the left is just the intersection of the other two, but after the update they result in the two models of Figure 6.7, which clearly correspond to different plausibility models.

Figure 6.7: Counterexample 3, after the update

Ruling out the previous cases we have the following group of assumptions to work with:

- Elements in $n \cap \neg P$ are not separated.
- Strict subsets of P trigger θ: **assumption (a)**.
- There are separated points that are not witnesses: **assumption (b)**.
- There are adopted points and all adopted points are witnesses **assumption (c)**, i.e., the points in $\neg n \cap P$ are adopted and witnesses.
- The elements that are separated and not witnesses are in $n \cap P$: **assumption (d)**.

From all this we can infer that the strict subsets of $\neg P$ adopt all the P-points and the strict subsets of P are sent to their complement within P. Consider the models in Figure 6.8. A little reflection shows that they produce the same ordering.

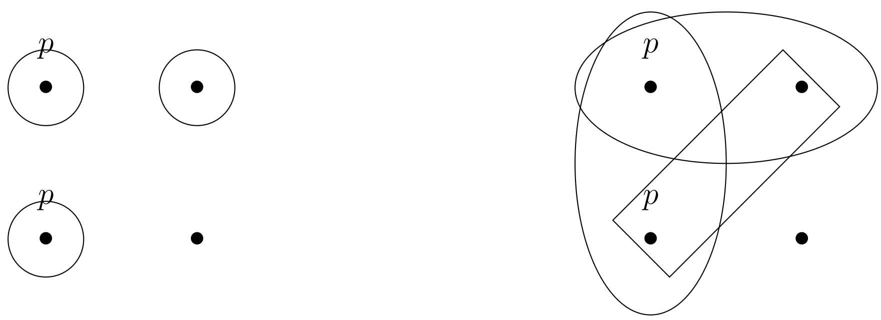


Figure 6.8: Counterexample 4

After the update, on the left of Figure 6.9, the singleton outside of P is enriched with the P-points while the other two singletons are mapped to each other; on the right, the set P is stable while the other two sets are mapped to each other. Hence the evidence separating the top-right element from the P-elements is lost in the model on the left and kept in the model on the right. This concludes the case distinction under **assumption (a)**.

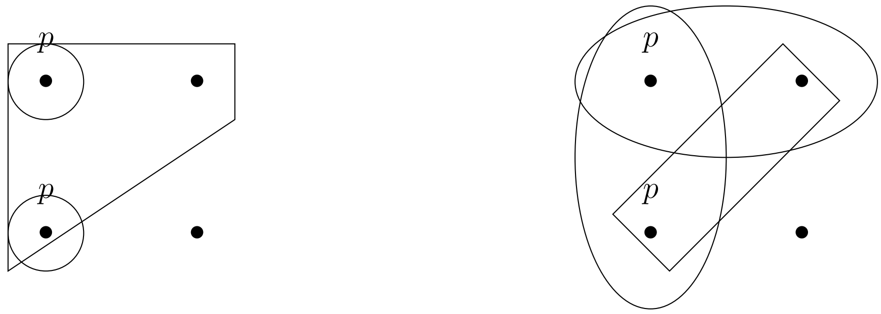


Figure 6.9: Counterexample 4, after the update

Case 2. Suppose now that P does not fulfill θ and the strict supersets of P trigger θ (**assumption (a')**). As before, the former assumption entails that, there are no witnesses in $\neg n \cap \neg P$ and $n \cap P$, or otherwise there would be models where P would trigger θ. By the latter assumption we can conclude that for any n the elements in $n \cap \neg P$ must be witnesses, or otherwise a strict superset of P could not trigger θ. By the failure of item 1 we know that there is either an adopted point that is not a witness or a separated point that is not a witness.

- Suppose there are adopted points that are not witnesses (**assumption (b')**). Assume that there are no separated points (**assumption (c')**). Then one of two cases occurs.

  If the elements in $\neg n \cap \neg P$ are adopted then every superset of P will be mapped to the whole set. Thus the third counterexample will work: the supersets of P will be mapped to the whole set while P will remain unchanged, acting as a division between two points that are not divided in the other updated model.

  If the elements in $\neg n \cap \neg P$ are not adopted then the ones in $\neg n \cap P$ must be adopted and not be witnesses, by **assumption (b')**. Then every set that intersects $\neg P$ adopts all the P elements. Thus both supersets and subsets of P are fixpoints of this update; the former because they adopt no new elements (and by **assumption (c')** there are no separated elements), the latter because they do not trigger θ. This suggests that the fourth counterexample does the job: in the model on the right every piece of evidence is sent to a superset of P, while in the model on the left the subsets of P are left untouched; therefore there are divisions within P in one updated model but not in the other.

Suppose now that there are separated points (we thus drop **assumption (c')**) and that all separated points are witnesses (**assumption (d')**). Together with **assumption (a')** this entails that points in $n \cap \neg P$ are separated and witnesses (and are the only separated points). If the elements in $\neg n \cap P$ are also witnesses then the subsets of P will adopt all the $\neg P$- elements. Thus the first counterexample works: on the left side, after the update the set P will divide P-elements from $\neg P$ elements, while on the other side this division is lost because both singletons adopt all the $\neg P$- elements.

If the elements in $\neg n \cap P$ are not witnesses but are adopted then every set that intersects $\neg P$ adopts all the P elements and loses the $\neg P$ ones. Subsets of P, however, are fixpoints. If we take the fourth counterexample we will reach the desired result: the singletons subsets of P that feature in the model on the left are left undisturbed by the update, while all the evidence sets that intersect $\neg P$ is sent to P. Thus again we have divisions within P in the updated model on the left but not in the updated model on the right.

Finally, if the elements in $\neg n \cap P$ are not witnesses and not adopted, by **assumption (b')** it must be that the points in $\neg n \cap \neg P$ are adopted and not witnesses. Here subsets of P are fixpoints and supersets of P just trade places, in the sense that they lose their $\neg P$ points and adopt the $\neg P$ point that they do not contain. To build a counterexample we have to build a new pair of models. In this circumstance supersets of $\neg P$ are mapped to subsets of P, because they lose all their $\neg P$ points, while subsets of $\neg P$ swap places, which is the leading idea for Counterexample 5, depicted in Figure 6.10.

Figure 6.10: Counterexample 5, before the update

After the update, in the model on the left of Figure 6.11, the singleton switches to the other $\neg P$ element and the two-elements set replaces its $\neg P$ element with the $\neg P$ element outside of it. Thus there is no division between the top-left and bottom-right elements after the update. In the model on the right, after the update the trapezoid loses all the $\neg P$ elements, shrinking to the top-left singleton. But now in this updated model there is a division between the top-left and bottom-right elements, hence the two updated models will produce a different ordering. This concludes the cases under **assumption (b')**.

Figure 6.11: Counterexample 5, after the update

- In the other main case, the one given by the failure of item 1, we know that there are separated elements that are not witnesses. Since separated points belong to n and we have seen that under **assumption (a')** the points in $n \cap \neg P$ are witnesses, it must be that the points in $n \cap P$ are separated and not witnesses. Thus any superset of P will lose all the P-points after the update, while P itself will remain stable. Hence the third counterexample will prevent tracking: in the first updated model there is a set separating P-worlds from the others, while in the second updated both evidence sets lose all the P-worlds. □

This concludes our demonstration of how every counterexample can be constructed from five simple patterns. The last proposition also constitutes the last missing piece for the proof of Theorem 6.5.15.

6.7 Conclusions

In this chapter we described how evidence and plausibility models can be arranged into categories by means of different notions of bisimulations. We then surveyed several updates on these structures and investigated the possibility of conceiving them as endofunctors. This highlighted the connection between the impossibility of making these construction functorial and the non-existence of reduction laws for the dynamic modalities associated to the updates.

We successively expanded on the connection between the categories of plausibility models and the categories of evidence models, showcasing different results that emphasize the effect of the choice of bisimulations in the two categories. Lastly, in Section 6.5 we dived into the topic of tracking. We described a new tracking result and discussed the significance of tracking in light of our previous study on the functoriality of updates. We concluded with an if and only if characterization of the trackable updates in the class of simple updates.

The groundwork put forward in this chapter provides a basis to further study the issues connected to dynamic updates and tracking, as well as a methodology that can be adapted to other classes of models different from evidence or plausibility models. A prominent example would be probability spaces and plausibility models.

The interface between these two structures, the first tied to Bayesian epistemology and the second to epistemic and doxastic logics, has received a lot of attention in recent years; we believe our framework could be successfully applied in this context too.

As for open problems, a natural continuation along our path would be to extend the characterization of Theorem 6.5.15 to a wider class of updates. We have also seen the sense in which tracking, at least in the case study we analyzed, connects to the reduction of second-order formulas to first-order ones. Updates whose definition can be reduced in such a way are, loosely speaking, treating a second order structure as if it were first-order. This perspective becomes particularly interesting if we consider that evidence models are examples of neighborhood models. Even though our results are tailored to work on the class of evidence models specifically, the techniques employed in this chapter could be tested in the general case, namely the tracking of operations on neighborhood models by operations on preorders.

We conclude this chapter observing how the problem of tracking can itself be categorically motivated. Categories of models such as preorders and neighborhood models have been extensively studied in the context of Duality Theory, showing different connections between spaces and relational structures. These kind of results typically impose conditions on the objects of the categories or on the morphisms to obtain such a duality. One way to see tracking is as a search for the right conditions to establish a correspondence between endofunctors on two separate categories, i.e. moving the first steps towards a duality between endofunctor categories.

Chapter 7

Conclusions

The time has come to summarize what we see as the theoretical underpinning of this thesis, and what our separate results have taught us about it. For a start, there are clear running themes in all that we have done. Perhaps the most obvious one is the key role of bisimulation.

The reader may have noticed that many categories we dealt with had some version of bisimulation as morphisms. In the case of the category of typed transition systems in Chapter 2, we adopted a notion of bisimulation that was the natural counterpart of the standard bisimulation. There is ample supply of examples in the literature where the latter concept (or its functional version) is used to provide arrows for a category of relational structures. Chapter 6 however goes along a less well-trodden path, describing categories in which arrows are given by different notions of bisimulations.

This suggestion is not accidental; in fact, it connects to a conceptual issue at the heart of the topic of this thesis, namely the parallelism between languages and morphisms. Suppose given a class of models $\mathcal{C}$ and a powerful expressive language L' interpreted on such models.[1] One of the major insights provided by the field of Modal Logic is the idea of characterizing a language L interpreted on $\mathcal{C}$ via a tuple $\langle \mathcal{C}, L', R \rangle$, where R is some notion of relationship between models. The idea is that L is characterized as the fragment of L' that is invariant with respect to R, over the class $\mathcal{C}$. The prototypical example is van Benthem's theorem, stating that basic modal logic is the fragment of FOL that is invariant under bisimulations.

One may wonder what is the adequate notion R for a given language L or, conversely, search for the fragment L corresponding to a given R. Modal Logic is often concerned with the first question (see the overture of Chapter 5), while a category-theoretic point of view turns the attention to the choice of transformations or morphisms, leading to the second question. Every pick of transformation indirectly defines a class of transformation-invariant properties,

[1] Typically such language would be the one used to describe the models themselves, thus a first or higher-order language.

namely those that are preserved by such transformations. This in turn prompts the question: what is the fragment of L' in which we can express all and only the invariant properties?

Thus, for a fixed $\mathcal{C}$ and L', fragments of L' and notions of morphisms seem to go hand in hand, linked by invariance results. This phenomenon is especially evident for the categories of models and bisimulations, where the latter are coupled with modal languages via Hennessy-Milner results, and its significance is amplified by the plurality of languages proposed in the Modal Logic literature. We believe this theoretical knot is an entanglement of many threads traversing old and new issues, some of which we have addressed in this thesis.

With these considerations in mind, we retrace our steps and take stock of where we are with respect to our initial goal. We begin from the title. Beside being a word play on the title of the influential book by MacLane [84], our headline is meant to hint at the overarching theme of this dissertation, namely the placement of heuristics and techniques from Category Theory in the toolkit of modal logicians.

Beside providing general considerations such as the ones we just offered, one would like to show that certain formal tools can be useful in solving the problems at hand, or can tilt such problems in an interesting and original direction. The extreme level of generality of Category Theory can sometimes constitute a fog screen, concealing the import of the theory to specific applications.

We thus attempted to build a ladder from the heavens of general abstract nonsense[2] to the interests of modal logicians, in particular those concerned with modelling tasks and therefore involved with particular models and languages. Rather than providing a general argument or theory, we set out to collect enough evidence for this fruitful interaction. Our Chapter 4, for example, linked the general work on presheaves models to a modal logic for social choice functions, unveiling the importance of a logic for varying coalitions. Chapter 5 did not establish a formal connection, but displayed how the issues of choosing the right notion of morphism and ensuring its closure under composition can become very relevant for models of modal languages. This theme was further developed in Chapter 6, where categories of models were employed in the analysis of current problems in Modal Logic.

As we briefly mentioned in the Introduction, the examples we analyzed witness two possible modes of interaction. The first is the study of hybrid models, namely structures that are on one hand significant from a category-theoretic perspective and on the other hand lend themselves to a treatment with modal languages. We showed how presheaf models can be seen as particular relational structures and developed a hierarchy of modal languages to express their properties. In Chapter 2 and 3 we highlighted how in this setting some of the traditional issues of Modal Logic, e.g. completeness, expressivity and decidability, receive an original twist

[2]This periphrasis is often used to jokingly refer to Category Theory.

and can be resolved with alternative solutions.

A second mode, more heuristic in nature, consists of regarding a given class of models as a category. The benefit of this stance is the cluster of questions that come with it. What is the most sensible notion of morphism for these models? Is it closed under composition? Given a uniform construction on such models, is it functorial? These are some of the basic issues that get raised in this context; in Chapter 5 and 6 we saw how they can shape our research and how they are intertwined with problems such as definability, existence of reduction laws and preservation of bisimulations.

We believe that the examples we treated and the techniques we introduced are not isolated success stories, but rather an indication that the interaction between Category Theory and Modal Logic can be further developed and give rise to a broad scale of further applications.

Bibliography

[1] Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. On the logic of preference and judgment aggregation. *Journal of Autonomous Agents and Multiagent Systems*, 22(1):4–30, 2011.

[2] Samson Abramsky. Arrow's theorem by arrow theory. In Andrés Villaveces, Roman Kossak, Juha Kontinen, and Åsa Hirvonen, editors, *Logic Without Borders: Essays on Set Theory, Model Theory, Philosophical Logic and Philosophy of Mathematics*, pages 15–30. De Gruyter, 2015.

[3] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield. Contextuality, cohomology and paradox. pages 211–228.

[4] Samson Abramsky and Adam Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13(11):113036, 2011.

[5] Jirí Adámek. Algebra ∩ coalgebra= presheaves. In *International Conference on Algebra and Coalgebra in Computer Science*, pages 67–82. Springer, 2005.

[6] Natasha Alechina, Michael Mendler, Valeria De Paiva, and Eike Ritter. Categorical and Kripke semantics for constructive S4 modal logic. In *International Workshop on Computer Science Logic*, pages 292–307. Springer, 2001.

[7] Alon Altman and Moshe Tennenholtz. Axiomatic foundations for ranking systems. *Journal of Artificial Intelligence Research*, 31:473–495, 2008.

[8] Mikkel Birkegaard Andersen, Thomas Bolander, and Martin Holm Jensen. Don't plan for the unexpected: Planning based on plausibility models. *Logique et Analyse*, 1(1), 2014.

[9] Mikkel Birkegaard Andersen, Thomas Bolander, Hans van Ditmarsch, and Martin Holm Jensen. Bisimulation for single-agent plausibility models. In Stephen Cranefield and Abhaya Nayak, editors, *AI 2013: Advances in Artificial Intelligence*, volume 8272 of *Lecture Notes in Computer Science*, pages 277–288. Springer International Publishing, 2013.

[10] Mikkel Birkegaard Andersen, Thomas Bolander, Hans van Ditmarsch, and Martin Holm Jensen. Bisimulation and expressivity for conditional belief, degrees of belief, and safe belief. *Synthese*, pages 1–41, 2015.

[11] Kenneth J Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 2nd edition, 1963. First edition published in 1951.

[12] Steve Awodey. *Category theory*, volume 49. Oxford University Press, 2006.

[13] Steve Awodey and Kohei Kishida. Topology and modality: The topological interpretation of first-order modal logic. *The Review of Symbolic Logic*, 1(02):146–166, 2008.

[14] Philippe Balbiani, Andreas Herzig, and Nicolas Troquard. Dynamic logic of propositional assignments: a well-behaved variant of PDL. In *Proceedings of the 2013 28th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 143–152. IEEE Computer Society, 2013.

[15] Alexandru Baltag. A coalgebraic semantics for epistemic programs. *Electronic Notes in Theoretical Computer Science*, 82(1):17–38, 2003.

[16] Alexandru Baltag and Giovanni Cinà. Modal logics for presheaves. Manuscript, 2014.

[17] Alexandru Baltag and Giovanni Cinà. Bisimulation for conditional modalities. Technical report, ILLC Technical Notes X-2016-01, 2016.

[18] Alexandru Baltag and Giovanni Cinà. Bisimulation for conditional modalities. *Studia Logica*, 2017. Forthcoming.

[19] Alexandru Baltag and Lawrence S Moss. Logics for epistemic programs. *Synthese*, 139:165–224, 2004.

[20] Alexandru Baltag and Sonja Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science*, 165:5–21, 2006.

[21] Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. In M. Woolridge G. Bonanno, W. van der Hoek, editor, *Texts in Logic and Games*, volume 3, pages 9–58. Amsterdam University Press, 2008.

[22] Bernhard Beckert, Rajeev Goré, Carsten Schürmann, Thorsten Bormer, and Jian Wang. Verifying voting schemes. *Journal of Information Security and Applications*, 19(2):115–129, 2014.

[23] John L Bell. Types, Sets, and Categories. In *Handbook of the History of Logic. Volume 6. Sets and Extensions in the Twentieth Century*, pages 633–687. Elsevier, 2012.

[24] Johan van Benthem. *Modal correspondence theory*. PhD thesis, Mathematical Institute, University of Amsterdam, 1977.

[25] Johan van Benthem. Dynamic logic for belief revision. *Journal of applied non-classical logics*, 17(2):129–155, 2007.

[26] Johan van Benthem. Tracking information. In K. Bimbo, editor, *J. Michael Dunn on Information Based Logics*, pages 363–389. Springer, 2016.

[27] Johan van Benthem, Nick Bezhanishvili, Sebastian Enqvist, and Junhua Yu. Instantial neighborhood logic. *The Review of Symbolic Logic*, 10(1):116–144, 2017.

[28] Johan van Benthem, Jan van Eijck, and Barteld Kooi. Logics of communication and change. *Information and computation*, 204(11):1620–1662, 2006.

[29] Johan van Benthem and Eric Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99(1-3):61–92, 2011.

[30] Gavin M Bierman and Valeria CV De Paiva. Intuitionistic necessity revisited. *School of Computer Science research reports - University of Birmingham CSR*, 1996.

[31] Gavin M Bierman and Valeria CV de Paiva. On an intuitionistic modal logic. *Studia Logica*, 65(3):383–416, 2000.

[32] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2002.

[33] Patrick Blackburn, Johan van Benthem, and Frank Wolter. *Handbook of modal logic*, volume 3. Elsevier, 2006.

[34] Oliver Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49(1):49–80, 2004.

[35] Elise Bonzon, Marie-Christine Lagasquie-Schiex, Jérôme Lang, and Bruno Zanuttini. Boolean games revisited. In *Proc. 17th European Conference on Artificial Intelligence (ECAI-2006)*, 2006.

[36] Felix Brandt, Vincent Conitzer, and Ulle Endriss. Computational social choice. In G. Weiss, editor, *Multiagent Systems*, pages 213–283. MIT Press, 2013.

[37] Felix Brandt and Christian Geist. Finding strategyproof social choice functions via SAT solving. In *Proc. 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2014)*, 2014.

[38] Glen E Bredon. *Sheaf theory*, volume 170. Springer Science & Business Media, 1997.

[39] Roberto Bruni, Hernán Melgratti, and Ugo Montanari. A connector algebra for p/t nets interactions. In *CONCUR 2011–Concurrency Theory*, pages 312–326. Springer, 2011.

[40] Gian Luca Cattani and Glynn Winskel. Profunctors, open maps and bisimulation. *Mathematical Structures in Computer Science*, 15(03):553–614, 2005.

[41] Siddharth Chatterjee and Arunava Sen. Automated reasoning in social choice theory: Some remarks. *Mathematics in Computer Science*, 8(1):5–10, 2014.

[42] Brian F Chellas. Basic conditional logic. *Journal of philosophical logic*, 4(2):133–153, 1975.

[43] Giovanni Cinà. Tracking ‘tracking’ with categories. Manuscript, 2016.

[44] Giovanni Cinà and Ulle Endriss. A syntactic proof of Arrow’s theorem in a modal logic of social choice functions. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1009–1017, 2015.

[45] Giovanni Cinà and Ulle Endriss. Proving classical theorems of social choice theory in modal logic. *Autonomous Agents and Multi-Agent Systems*, 30(5):963–989, 2016.

[46] Giovanni Cinà and Sebastian Enqvist. Bisimulation and path logic for sheaves: contextuality and beyond. Technical report, ILLC Technical Notes X-2015-01, 2015.

[47] Valeria de Paiva and Eike Ritter. Basic constructive modality. *Logic without frontiers - Festschrift for Walter Alexandre Carnielli on the occasion of his 60th birthday. College Publications*, pages 411–428, 2011.

[48] Lorenz Demey. Some remarks on the model theory of epistemic plausibility models. *Journal of Applied Non-Classical Logics*, 21(3-4):375–395, 2011.

[49] Lorenz Demey. Looking for the right notion of epistemic plausibility model. *Logic, Philosophy and History of Science in Belgium II. Proceedings of the Young Researchers Days 2010*, pages 73–78, 2012.

[50] Jan van Eijck. Update, probability, knowledge and belief. Draft available from homepages.cwi.nl/ jve/papers/15/pdfs/upkb.pdf, 2015.

[51] Jan van Eijck and François Schwarzentruber. Epistemic probability logic simplified. *Advances in Modal Logic*, 10:158–177, 2014.

[52] Samuel Eilenberg and Saunders MacLane. General theory of natural equivalences. *Transactions of the American Mathematical Society*, 58(2):231–294, 1945.

[53] Herbert B. Enderton. *A mathematical introduction to logic. 2nd ed.* San Diego, CA: Harcourt/Academic Press, 2nd ed. edition, 2001.

[54] Ulle Endriss. Logic and social choice theory. In A. Gupta and J. van Benthem, editors, *Logic and Philosophy Today*, volume 2, pages 333–377. College Publications, 2011.

[55] Robert Furber, Mathys Rennela, and Sam Staton. Infinite-dimensionality in quantum foundations: W^*-algebras as presheaves over matrix algebras. *Electronic Proceedings in Theoretical Computer Science, QPL*, 2016.

[56] Dov M Gabbay. *Theoretical foundations for non-monotonic reasoning in expert systems*. Springer, 1985.

[57] Christian Geist and Ulle Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011.

[58] Jelle Gerbrandy. Logics of propositional control. In *Proc. 5th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2006)*, 2006.

[59] Silvio Ghilardi. Incompleteness results in Kripke semantics. *Journal of Symbolic Logic*, pages 517–538, 1991.

[60] Silvio Ghilardi and Giancarlo Meloni. Modal and tense predicate logic: Models in presheaves and categorical conceptualization. In *Categorical algebra and its applications*, pages 130–142. Springer, 1988.

[61] Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.

[62] Robert I Goldblatt and Steve K Thomason. Axiomatic classes in propositional modal logic. In *Algebra and logic*, pages 163–173. Springer, 1975.

[63] Daniel Gorín and Lutz Schröder. Simulations and bisimulations for coalgebraic modal logics. In *Algebra and Coalgebra in Computer Science*, pages 253–266. Springer, 2013.

[64] Umberto Grandi and Ulle Endriss. First-order logic formalisation of impossibility theorems in preference aggregation. *Journal of Philosophical Logic*, 42(4):595–618, 2013.

[65] Davide Grossi, Emiliano Lorini, and François Schwarzentruber. The ceteris paribus structure of logics of game forms. *Journal of Artificial Intelligence Research*, pages 91–126, 2015.

[66] Adam Grove. Two modellings for theory change. *Journal of philosophical logic*, 17(2):157–170, 1988.

[67] Joseph Y Halpern. *Reasoning about uncertainty*, volume 21. MIT Press Cambridge, 2003.

[68] Paul Harrenstein, Wiebe van der Hoek, John-Jules Meyer, and Cees Witteveen. Boolean games. In *Proc. 8th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2001)*, 2001.

[69] Wilfrid Hodges. *A shorter model theory.* Cambridge university press, 1997.

[70] Bart Jacobs. *Categorical logic and type theory*, volume 141. Elsevier, 1999.

[71] Bart Jacobs. The temporal logic of coalgebras via galois algebras. *Mathematical Structures in Computer Science*, 12(06):875–903, 2002.

[72] Bart Jacobs. *Introduction to Coalgebra: Towards Mathematics of States and Observation*, volume 59. Cambridge University Press, 2016.

[73] André Joyal and Ieke Moerdijk. A completeness theorem for open maps. *Annals of Pure and Applied Logic*, 70(1):51–86, 1994.

[74] André Joyal, Mogens Nielsen, and Glynn Winskel. Bisimulation from open maps. *Information and Computation*, 127(2):164–185, 1996.

[75] Kohei Kishida. Stochastic relational presheaves and dynamic logic for contextuality. In B. Coecke et al., editor, *Proceedings of QPL 2014*, pages 115–132.

[76] Barteld P Kooi. *Knowledge, chance, and change.* PhD thesis, Groningen University, 2003.

[77] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial intelligence*, 44(1):167–207, 1990.

[78] Alexander Kurz. Coalgebras and modal logic. *Course Notes for ESSLLI*, 2001.

[79] Christoph Lange, Colin Rowat, and Manfred Kerber. The ForMaRE Project: Formal mathematical reasoning in economics. In *Intelligent Computer Mathematics*, pages 330–334. Springer-Verlag, 2013.

[80] Sławomir Lasota. Coalgebra morphisms subsume open maps. *Theoretical Computer Science*, 280(1):123–135, 2002.

[81] James J Leifer and Robin Milner. Deriving bisimulation congruences for reactive systems. In *International Conference on Concurrency Theory*, pages 243–258. Springer, 2000.

[82] Clarence Irving Lewis. *A survey of symbolic logic*. University of California Press, 1918.

[83] David Lewis. Counterfactuals. *Harvard University Press*, 1973.

[84] Saunders MacLane. *Categories for the Working Mathematician. 2nd edition.* Springer, 1998.

[85] Saunders MacLane and Ieke Moerdijk. *Sheaves in geometry and logic: A first introduction to topos theory.* Springer, 1992.

[86] Michael Makkai and Gonzalo E. Reyes. Completeness results for intuitionistic and modal logic in a categorical setting. *Annals of Pure and Applied Logic*, 72(1):25–101, 1995.

[87] Michael Makkai and Gonzalo E Reyes. *First order categorical logic: Model-theoretical methods in the theory of topoi and related categories*, volume 611. Springer, 2006.

[88] Octavio Malherbe, Philip Scott, and Peter Selinger. Presheaf models of quantum computation: an outline. In *Computation, Logic, Games, and Quantum Foundations. The Many Facets of Samson Abramsky*, pages 178–194. Springer, 2013.

[89] Andrew Mao, Ariel D. Procaccia, and Yiling Chen. Better human computation through principled voting. In *Proc. 27th AAAI Conference on Artificial Intelligence (AAAI-2013)*, 2013.

[90] Jean-Pierre Marquis. *From a geometrical point of view: a study of the history and philosophy of category theory*, volume 14. Springer Science & Business Media, 2008.

[91] Jean-Pierre Marquis. Category Theory. In Edward N Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2015 edition, 2015.

[92] Kenneth O May. A set of independent necessary and sufficient conditions for simple majority decisions. *Econometrica*, 20(4):680–684, 1952.

[93] John CC McKinsey and Alfred Tarski. The algebra of topology. *Annals of mathematics*, pages 141–191, 1944.

[94] John CC McKinsey and Alfred Tarski. Some theorems about the sentential calculi of Lewis and Heyting. *The Journal of Symbolic Logic*, 13(01):1–15, 1948.

[95] Peter D Mosses. Modular structural operational semantics. *The Journal of Logic and Algebraic Programming*, 60:195–228, 2004.

[96] Eitan Muller and Mark A Satterthwaite. The equivalence of strong positive association and strategy-proofness. *Journal of Economic Theory*, 14(2):412–418, 1977.

[97] Tobias Nipkow. Social choice theory in HOL: Arrow and Gibbard-Satterthwaite. *Journal of Automated Reasoning*, 43(3):289–304, 2009.

[98] Rohit Parikh. The logic of games and its applications. In *Topics in the Theory of Computation*, volume 24 of *Annals of Discrete Mathematics*. North-Holland, 1985.

[99] Rohit Parikh. Social software. *Synthese*, 132(3):187–211, 2002.

[100] Dirk Pattinson. Semantical principles in the modal logic of coalgebras. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 514–526. Springer, 2001.

[101] Marc Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.

[102] Marc Pauly. On the role of language in social choice theory. *Synthese*, 163(2):227–243, 2008.

[103] David M Pennock, Eric Horvitz, and C Lee Giles. Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *Proc. 17th National Conference on Artificial Intelligence (AAAI-2000)*, 2000.

[104] Tin Perkov. Natural deduction for a fragment of modal logic of social choice. Presented at ESSLLI-2014 Workshop on Information Dynamics in Artificial Societies, 2014.

[105] Tin Perkov. Natural deduction for modal logic of judgment aggregation. *Journal of Logic, Language and Information*, 25(3):335–354, 2016.

[106] Jan Plaza. Logics of public communications. *Synthese*, 158(2):165–179, 2007.

[107] Slavian Radev. Infinitary propositional normal modal logic. *Studia Logica*, 46(4):291–309, 1987.

[108] Jan JMM Rutten. Universal coalgebra: a theory of systems. *Theoretical Computer Science*, 249(1):3–80, 2000.

[109] Mark A Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

[110] Amartya K Sen. The impossibility of a Paretian liberal. *The Journal of Political Economics*, 78(1):152–157, 1970.

[111] Amartya K Sen. Social Choice Theory. In K. J. Arrow and M. D. Intriligator, editors, *Handbook of Mathematical Economics*, volume 3. North-Holland, 1986.

[112] Pawel Sobociński. Relational presheaves as labelled transition systems. In *Coalgebraic Methods in Computer Science (CMCS '12)*, volume 7399 of *LNCS*, pages 40–50. Springer, 2012.

[113] Robert C Stalnaker. A theory of conditionals. In N. Rescher, editor, *Studies in logical theory*. Blackwell, Oxford, 1968.

[114] Holger Sturm. A complete axiomatization for $\mathcal{ML}_{\omega_1}$. Available online at http://www.cis.lmu.de/download/cis-berichte/98-113.pdf.

[115] Pingzhong Tang and Fangzhen Lin. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

[116] Alan D Taylor. *Social Choice and the Mathematics of Manipulation*. Cambridge University Press, 2005.

[117] Nicolas Troquard. Logics of social choice and perspectives on their software implementation. Presented at Dagstuhl Seminar 11101 on Reasoning about Interaction: From Game Theory to Logic and Back, 2011.

[118] Nicolas Troquard, Wiebe van der Hoek, and Michael Wooldridge. Reasoning about social choice functions. *Journal of Philosophical Logic*, 40(4):473–498, 2011.

[119] Jouko Väänänen. Many-sorted logic, July 2014. Course notes.

[120] Wiebe van der Hoek and Michael Wooldridge. On the logic of cooperation and propositional control. *Artificial Intelligence*, 164(1):81–119, 2005.

[121] Frank Veltman. Logic for conditionals. *PhD thesis, University of Amsterdam*, 1985.

[122] Freek Wiedijk. Arrow's Impossibility Theorem. *Formalized Mathematics*, 15(4):171–174, 2007.

[123] Glynn Winskel and Mogens Nielsen. Models for concurrency. *DAIMI Report Series*, 22(463), 1993.

[124] Glynn Winskel and Mogens Nielsen. Presheaves as transition systems. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 29:129–140, 1997.

[125] James Worrell. A note on coalgebras and presheaves. *Electronic Notes in Theoretical Computer Science*, 65(1):358–364, 2002.

Samenvatting

In dit proefschrift trachten wij een ladder te construeren die de hogere sferen van de Categorietheorie verbindt met de interessen van modale logici, in het bijzonder diegenen die zich bezighouden met modelleringstaken en daarmee het ontwerp van specifieke modellen en talen. In plaats van een algemene a priori argumentatie of theorie inzake dit verband streven wij naar het verzamelen van voldoende concrete evidentie die het nut van deze interactie aantoont.

Het werk dat wordt gepresenteerd in dit proefschrift geeft twee vormen van interactie te zien. De eerste vorm is de studie van hybride modellen, dat wil zeggen, structuren die enerzijds significant zijn vanuit een categorietheoretisch gezichtspunt, maar die zich anderzijds ook lenen voor behandeling met modale talen. Zo laten we met name zien hoe pre-schoof modellen beschouwd kunnen worden als een speciaal soort relationele modellen, en we ontwikkelen een hiërarchie van modale talen die hun eigenschappen kunnen weergeven. Om de flexibiliteit van dit raamwerk aan te tonen geven we verschillende toepassingen, waaronder in het bijzonder een modale logica voor sociale keuzefuncties. Ook laten we zien hoe in onze benadering traditionele vragen in de Modale Logica, zoals volledigheid, uitdrukkingskracht, en beslisbaarheid, een nieuwe draai krijgen en op nieuwe manieren kunnen worden aangetoond.

Een tweede vorm van interactie, meer heuristisch van aard, beschouwt gegeven klassen van modellen zelf als categorieën. Het voordeel van deze zienswijze is het pakket van vragen dat hiermee gepaard gaat. Enkele fundamentele vragen die op deze manier rijzen zijn de juiste keuze van morfismen voor de gegeven modellen, afsluiting onder compositie van deze morfismen, en de functorialiteit van uniforme constructies op de modellen. In het tweede deel van het proefschrift leggen we uit hoe zulke vragen richting kunnen geven aan modaal-logisch onderzoek en hoe ze een natuurlijk mengsel vormen met reeds bestaande vragen op modaal gebied.

Wij menen dat de voorbeelden behandeld in dit proefschrift, en de nieuwe technieken die zijn geïntroduceerd, geen geïsoleerde succesjes zijn, maar een indicatie dat de interactie van Categorietheorie en Modale Logica verder kan worden uitgebouwd, en aanleiding kan geven tot een breed scala aan verdere toepassingen.

Abstract

We attempt to build a ladder connecting the heavens of Category Theory to the interests of modal logicians, in particular those concerned with modelling tasks and therefore involved with specific models and languages. Rather than providing a general argument or theory, we set out to collect enough evidence for this fruitful interplay.

The body of work presented in this theses witnesses two possible modes of interaction. The first is the study of hybrid models, namely structures that are on one hand significant from a category-theoretic perspective and on the other hand lend themselves to a treatment with modal languages. We show how presheaf models can be seen as particular relational structures and develop a hierarchy of modal languages to express their properties. To argue in favor of the flexibility of this framework we review several applications; we especially dive into the details of a modal logic for social choice functions. Furthermore, we highlight how in this setting some of the traditional issues of Modal Logic, e.g. completeness, expressivity and decidability, receive an original twist and can be resolved with alternative solutions.

A second mode, more heuristic in nature, consists of regarding a given class of models as a category. The benefit of this stance is the baggage of questions that come with it. The right notion of morphism for these models, its closure under composition, the functoriality of some uniform constructions, these are some of the basic issues that get raised in this context. In the second half of the thesis we explicate how they can shape research in Modal Logic and how they are intertwined with existing problems.

We believe that the examples we treated and the techniques we introduced are not isolated success stories, but rather an indication that the interaction between Category Theory and Modal Logic can be further developed and give rise to a broad scale of further applications.

Titles in the ILLC Dissertation Series:

ILLC DS-2009-01: **Jakub Szymanik**
Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language

ILLC DS-2009-02: **Hartmut Fitz**
Neural Syntax

ILLC DS-2009-03: **Brian Thomas Semmes**
A Game for the Borel Functions

ILLC DS-2009-04: **Sara L. Uckelman**
Modalities in Medieval Logic

ILLC DS-2009-05: **Andreas Witzel**
Knowledge and Games: Theory and Implementation

ILLC DS-2009-06: **Chantal Bax**
Subjectivity after Wittgenstein. Wittgenstein's embodied and embedded subject and the debate about the death of man.

ILLC DS-2009-07: **Kata Balogh**
Theme with Variations. A Context-based Analysis of Focus

ILLC DS-2009-08: **Tomohiro Hoshi**
Epistemic Dynamics and Protocol Information

ILLC DS-2009-09: **Olivia Ladinig**
Temporal expectations and their violations

ILLC DS-2009-10: **Tikitu de Jager**
"Now that you mention it, I wonder...": Awareness, Attention, Assumption

ILLC DS-2009-11: **Michael Franke**
Signal to Act: Game Theory in Pragmatics

ILLC DS-2009-12: **Joel Uckelman**
More Than the Sum of Its Parts: Compact Preference Representation Over Combinatorial Domains

ILLC DS-2009-13: **Stefan Bold**
Cardinals as Ultrapowers. A Canonical Measure Analysis under the Axiom of Determinacy.

ILLC DS-2010-01: **Reut Tsarfaty**
Relational-Realizational Parsing

ILLC DS-2010-02: **Jonathan Zvesper**
Playing with Information

ILLC DS-2010-03: **Cédric Dégremont**
The Temporal Mind. Observations on the logic of belief change in interactive systems

ILLC DS-2010-04: **Daisuke Ikegami**
Games in Set Theory and Logic

ILLC DS-2010-05: **Jarmo Kontinen**
Coherence and Complexity in Fragments of Dependence Logic

ILLC DS-2010-06: **Yanjing Wang**
Epistemic Modelling and Protocol Dynamics

ILLC DS-2010-07: **Marc Staudacher**
Use theories of meaning between conventions and social norms

ILLC DS-2010-08: **Amélie Gheerbrant**
Fixed-Point Logics on Trees

ILLC DS-2010-09: **Gaëlle Fontaine**
Modal Fixpoint Logic: Some Model Theoretic Questions

ILLC DS-2010-10: **Jacob Vosmaer**
Logic, Algebra and Topology. Investigations into canonical extensions, duality theory and point-free topology.

ILLC DS-2010-11: **Nina Gierasimczuk**
Knowing One's Limits. Logical Analysis of Inductive Inference

ILLC DS-2010-12: **Martin Mose Bentzen**
Stit, Iit, and Deontic Logic for Action Types

ILLC DS-2011-01: **Wouter M. Koolen**
Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice

ILLC DS-2011-02: **Fernando Raymundo Velazquez-Quesada**
Small steps in dynamics of information

ILLC DS-2011-03: **Marijn Koolen**
The Meaning of Structure: the Value of Link Evidence for Information Retrieval

ILLC DS-2011-04: **Junte Zhang**
System Evaluation of Archival Description and Access

ILLC DS-2011-05: **Lauri Keskinen**
Characterizing All Models in Infinite Cardinalities

ILLC DS-2011-06: **Rianne Kaptein**
Effective Focused Retrieval by Exploiting Query Context and Document Structure

ILLC DS-2011-07: **Jop Briët**
Grothendieck Inequalities, Nonlocal Games and Optimization

ILLC DS-2011-08: **Stefan Minica**
Dynamic Logic of Questions

ILLC DS-2011-09: **Raul Andres Leal**
Modalities Through the Looking Glass: A study on coalgebraic modal logic and their applications

ILLC DS-2011-10: **Lena Kurzen**
Complexity in Interaction

ILLC DS-2011-11: **Gideon Borensztajn**
The neural basis of structure in language

ILLC DS-2012-01: **Federico Sangati**
Decomposing and Regenerating Syntactic Trees

ILLC DS-2012-02: **Markos Mylonakis**
Learning the Latent Structure of Translation

ILLC DS-2012-03: **Edgar José Andrade Lotero**
Models of Language: Towards a practice-based account of information in natural language

ILLC DS-2012-04: **Yurii Khomskii**
Regularity Properties and Definability in the Real Number Continuum: idealized forcing, polarized partitions, Hausdorff gaps and mad families in the projective hierarchy.

ILLC DS-2012-05: **David García Soriano**
Query-Efficient Computation in Property Testing and Learning Theory

ILLC DS-2012-06: **Dimitris Gakis**
Contextual Metaphilosophy - The Case of Wittgenstein

ILLC DS-2012-07: **Pietro Galliani**
The Dynamics of Imperfect Information

ILLC DS-2012-08: **Umberto Grandi**
Binary Aggregation with Integrity Constraints

ILLC DS-2012-09: **Wesley Halcrow Holliday**
Knowing What Follows: Epistemic Closure and Epistemic Logic

ILLC DS-2012-10: **Jeremy Meyers**
Locations, Bodies, and Sets: A model theoretic investigation into nominalistic mereologies

ILLC DS-2012-11: **Floor Sietsma**
Logics of Communication and Knowledge

ILLC DS-2012-12: **Joris Dormans**
Engineering emergence: applied theory for game design

ILLC DS-2013-01: **Simon Pauw**
Size Matters: Grounding Quantifiers in Spatial Perception

ILLC DS-2013-02: **Virginie Fiutek**
Playing with Knowledge and Belief

ILLC DS-2013-03: **Giannicola Scarpa**
Quantum entanglement in non-local games, graph parameters and zero-error information theory

ILLC DS-2014-01: **Machiel Keestra**
Sculpting the Space of Actions. Explaining Human Action by Integrating Intentions and Mechanisms

ILLC DS-2014-02: **Thomas Icard**
The Algorithmic Mind: A Study of Inference in Action

ILLC DS-2014-03: **Harald A. Bastiaanse**
Very, Many, Small, Penguins

ILLC DS-2014-04: **Ben Rodenhäuser**
A Matter of Trust: Dynamic Attitudes in Epistemic Logic

ILLC DS-2015-01: **María Inés Crespo**
Affecting Meaning. Subjectivity and evaluativity in gradable adjectives.

ILLC DS-2015-02: **Mathias Winther Madsen**
The Kid, the Clerk, and the Gambler - Critical Studies in Statistics and Cognitive Science

ILLC DS-2015-03: **Shengyang Zhong**
Orthogonality and Quantum Geometry: Towards a Relational Reconstruction of Quantum Theory

ILLC DS-2015-04: **Sumit Sourabh**
Correspondence and Canonicity in Non-Classical Logic

ILLC DS-2015-05: **Facundo Carreiro**
Fragments of Fixpoint Logics: Automata and Expressiveness

ILLC DS-2016-01: **Ivano A. Ciardelli**
Questions in Logic

ILLC DS-2016-02: **Zoé Christoff**
Dynamic Logics of Networks: Information Flow and the Spread of Opinion

ILLC DS-2016-03: **Fleur Leonie Bouwer**
What do we need to hear a beat? The influence of attention, musical abilities, and accents on the perception of metrical rhythm

ILLC DS-2016-04: **Johannes Marti**
Interpreting Linguistic Behavior with Possible World Models

ILLC DS-2016-05: **Phong Lê**
Learning Vector Representations for Sentences - The Recursive Deep Learning Approach

ILLC DS-2016-06: **Gideon Maillette de Buy Wenniger**
Aligning the Foundations of Hierarchical Statistical Machine Translation

ILLC DS-2016-07: **Andreas van Cranenburgh**
Rich Statistical Parsing and Literary Language

ILLC DS-2016-08: **Florian Speelman**
Position-based Quantum Cryptography and Catalytic Computation

ILLC DS-2016-09: **Teresa Piovesan**
Quantum entanglement: insights via graph parameters and conic optimization

ILLC DS-2016-10: **Paula Henk**
Nonstandard Provability for Peano Arithmetic. A Modal Perspective

ILLC DS-2017-01: **Paolo Galeazzi**
Play Without Regret

ILLC DS-2017-02: **Riccardo Pinosio**
The Logic of Kant's Temporal Continuum

ILLC DS-2017-03: **Matthijs Westera**
Exhaustivity and intonation: a unified theory