Title
What Can You Say? Measuring the Expressive Power of Languages

Author
Kocurek, Alexander William

Publication Date
2018

Peer reviewed|Thesis/dissertation

What Can You Say? Measuring the Expressive Power of Languages

by

Alexander William Kocurek

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Logic and the Methodology of Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Wesley H. Holliday, Co-chair
Associate Professor Seth Yalcin, Co-chair
Professor John G. MacFarlane
Associate Professor Line Mikkelsen

Summer 2018

What Can You Say? Measuring the Expressive Power of Languages

Abstract

What Can You Say? Measuring the Expressive Power of Languages

by

Alexander William Kocurek

Doctor of Philosophy in Logic and the Methodology of Science

University of California, Berkeley

Associate Professor Wesley H. Holliday, Co-chair

Associate Professor Seth Yalcin, Co-chair

There are many different ways to talk about the world. Some ways of talking are more expressive than others—that is, they enable us to say more things about the world. But what exactly does this mean? When is one language able to express more about the world than another? In my dissertation, I systematically investigate different ways of answering this question and develop a formal theory of expressive power. In doing so, I show how these investigations help to clarify the role that expressive power plays within debates in metaphysics, logic, and the philosophy of language.

When we attempt to describe the world, we are trying to distinguish the way things are from all the many ways things could have been—in other words, we are trying to locate ourselves within a region of logical space. According to this picture, languages can be thought of as ways of carving logical space or, more formally, as maps from sentences to classes of models. For example, the language of first-order logic is just a mapping from first-order formulas to model-assignment pairs that satisfy those formulas. Almost all formal languages discussed in metaphysics and logic, as well as many of those discussed in natural language semantics, can be characterized in this way.

Using this picture of language, I analyze two different approaches to defining expressive power, each of which is motivated by different roles a language can play in a debate. One role a language can play is to divide and organize a shared conception of logical space. If two languages share the same conception of logical space (i.e., are defined over the same class of models), then one can compare the expressive power of these languages by comparing how finely they carve logical space. This is the approach commonly employed, for instance, in debates over tense and modality, such as the primitivism-reductionism debate.

But a second role languages can play in a debate is to advance a conception or theory of logical space itself. For example, consider the debate between perdurantism, which claims that objects persist through time by having temporal parts located throughout that time, and endurantism, which claims that objects persist through time by being wholly present at that time. A natural thought about this debate is that perdurantism and endurantism are simply alternative but equally good descriptions of the world rather than competing theories. Whenever the endurantist says, for instance, that an object is red at time t, the perdurantist can say that the object's temporal part at t is red. On this view, one should conceive of perdurantism and endurantism not as theories picking out disjoint regions of logical space, but as theories offering alternative conceptions of logical space: one in which persistence through time is analogous to location in space and one in which it is not. A similar distinction applies to other metaphysical debates, such as the mereological debate between universalism and nihilism.

If two theories propose incommensurable conceptions of logical space, we can still compare their expressive power utilizing the notion of a translation, which acts as a correlation between points in logical space that preserves the language's inferential connections. I build a formal theory of translation that explores different ways of making this notion precise. I then apply this theory to two metaphysical debates, viz., the debate over whether composite objects exist and the debate over how objects persist through time. This allows us to get a clearer picture of the sense in which these debates can be viewed as genuine.

Contents

Acknowledgements

I have many people to thank for their support while writing this dissertation.

First, I would like to thank the members of my committee. Wes Holliday and Seth Yalcin were a major source of inspiration and wisdom while writing this dissertation. I cannot stress how significant their presence was on my philosophical thinking or how fortunate I feel to have had such fantastic advisors throughout my time at UC Berkeley. I am also incredibly grateful to Lara Buchak, Shamik Dasgupta, John MacFarlane, and Line Mikkelsen for giving their helpful feedback both on this dissertation and other work I pursued during my graduate career. Apart from the member of my committee, I also have to thank Johan van Benthem, Hanti Lin, Eric Pacuit, and Kai Wehmeier for helpful discussion and encouragement at various stages of this project. Parts of chapters 2 and 3 were presented at Logica 2017 (in Hejnice, Czechia) and the LORI-VI (in Sapporo, Japan) in a more nascent form, and I am grateful for all the comments and discussion generated at these conferences. A sample of the results from those chapters was published in the LORI-VI conference proceedings as Kocurek 2017. The key results from that article have been included in this dissertation with permission.

Second, I would like to thank the graduate students in the Department of Philosophy. Despite not technically being a member of the department, I never felt more welcome in any other community than this one. I spent countless hours in 301 Moses Hall pitching my ideas, writing on the board, and getting feedback from people, all of which was vital for the development of the work presented here. In particular, I have to thank Mathias Boehm, Sophie Dandelet, Ravit Dotan, Sam Elgin, Ethan Jerzak, Jeff Kaplan, Jackson Kernion, Alex Kerr, Richard Lawrence, Sven Neth, Antonia Peacocke, Emily Perry, Rachel Rudolph, Pia Schneider, Dave Suarez, Alberto Tassoni, and Justin Vlasits for their friendship and their support—both personally and intellectually. Special thanks to Russell Buehler, who was an especially valuable comrade throughout my time at Berkeley. Special thanks also to James Walsh for helpful discussion. I would also like to thank Liz George, Richard Kwant, Anish Mittal, Jan Roden, Rob Smith, and Sebastian Werner for their enduring friendship.

Third, I would like to thank Melissa Fusco for her extraordinary love and affection while pursuing this project. On many occasions I have voiced my thoughts, my ideas, and my struggles to her, and she has always given me her attention, sympathy, and compassion.

Finally, I want to thank my parents, Michelle and Scott, as well as my brothers, Wesley and Zachary, for their unwavering confidence, encouragement, and support throughout every stage of my education. It was a long road to get to this point, and they were always there for me every step of the way. I am blessed to have such a loving and supportive family. I could not have done it without them.

Introduction

There are many different ways one can describe the world. Some are better than others, but many of them are all equally good. What makes one language better than another? Usually, a whole host of factors. Maybe one language is better than another in its simplicity, its efficiency, its flexibility, or its aesthetics. But a fundamental way in which some languages can be better than others is in their expressive capacities—the possibilities they can describe, the distinctions they can draw, the structures they can represent, and so forth. What does this mean, though? What exactly constitutes the expressive power of a language? The main project of this dissertation is to systematize different ways of answering this question.

Expressive power is one of the most fundamental properties of a formal language that one can investigate. Indeed, some of the most foundational results in mathematical logic are results concerning the expressive power of formal languages. The compactness theorem is a case in point. We often summarize its significance by saying things like "The compactness theorem shows, among other things, that first-order logic cannot express the claim that there are infinitely many things." Another prime example is the van Benthem characterization theorem in modal logic. One way to interpret this result is as follows: modal logic is expressively equivalent to the bisimulation-invariant fragment of first-order logic. Results such as these are not only practically useful but also give us insight into the strengths and limitations of a formal system.

In addition, expressive power is philosophically important. A number of disputes in philosophy hinge on whether one side is able to articulate every possibility the other side can articulate. Sometimes, lack of expressive power is raised as a criticism. For instance, it is thought to be a problem for primitivist views about modality that they cannot articulate various quantified modal claims without appealing to irreducible quantification over possible worlds (e.g., "Everything could have failed to exist"). Other times, it is too much expressive power that is objectionable. A classic objection to absolutism about space is that it makes distinctions where none are to be had (e.g., between a universe and its qualitatively identical counterpart where everything is shifted 3 feet in one direction). Other times still, the expressive equivalence of the two sides of the dispute is used as a *reductio* of the whole dispute. If everything that can be said by the mereological universalist can also be said by the mereological nihilist, then one might think there is nothing that really hinges on this dispute (just pick whichever way of talking you prefer!). Getting clear on what exactly are the expressive limitations of a language is crucial, then, for understanding the force of some of these objections.

The notion of expressive power employed in results such as the ones mentioned above can be characterized in the following manner. One language is as expressive as another if for every sentence of the latter, there is a sentence of the former that "means the same thing". To be more precise, think of a language as consisting of three ingredients: (1) a grammar, i.e., a collection of well-formed sentences of the language; (2) a conception of logical space, i.e., a collection of models used to assess the truth of the well-formed sentences; and (3) a semantics, i.e., a function that maps each well-formed sentence to a region of logical space, or its "semantic value".[1] For example, the language of first-order logic consists of the well-formed formulas of the first-order predicate calculus, the class of model-assignment pairs, and the semantics that maps each formula to the class of model-assignments pairs at which the formula is true. On this picture of a language, "means the same thing" can be cashed out in terms of their semantic values: two sentences "mean the same thing" if they pick out the same region of logical space. Thus, two first-order formulas "mean the same thing" if they are satisfied at exactly the same model-assignment pairs. Then we can state the relevant notion of expressive power as follows: a language $\mathbf{L}_1$ is as expressive as another language $\mathbf{L}_2$ if every sentence in $\mathbf{L}_2$ has the same semantic value as some sentence in $\mathbf{L}_1$. So a consequence of the compactness theorem can be stated as follows: first-order logic cannot express the claim that there are infinitely many things, since no first-order formula is true at exactly the models with an infinite domain.

On this simple definition of expressive power, in order to compare the expressive power of two languages, it must be presupposed that the languages in question are defined over the same conception of logical space. In order to say that $\mathbf{L}_1$ is as expressive as $\mathbf{L}_2$, it must be that every model of $\mathbf{L}_2$ is a model of $\mathbf{L}_1$. This is a convenient assumption. It facilitates answering questions about sameness of meaning that would otherwise be difficult to address. For instance, in the study of modal logics, when one wants to show that such-and-such an operator is not expressible in the basic modal language (say), one does so by showing there are Kripke models that can be distinguished by the modal language with such-and-such operator that cannot be distinguished by the basic modal language.[2] One uses one-and-the-same class of models for both languages to determine what can be said in each.

Yet this assumption is not always warranted. In many cases, it is desirable to compare the expressive power of languages with competing conceptions of logical space.

[1] This notion of meaning can be traced back to Wittgenstein [1922, 3.4, 4.022, 4.024]. See also Stalnaker 1976; Lewis 1986.

[2] See, e.g., Blackburn et al. 2001, pp. 64–73.

As a simple example, consider the language of (classical) propositional logic. One way to develop propositional logic is in terms of valuation functions, i.e., functions from atomic sentences to truth values. Another, apparently equivalent, way to develop propositional logic is in terms of valuation sets, i.e., sets of atomic sentences (the ones that are true according to that valuation). Intuitively, there should be no essential difference between these languages in their expressive capacities: each can say exactly what the other can say. And yet, according to the simple notion of expressive power above, these languages are expressively incomparable. Vacuously, *no* sentence of one version is true at "the same class of models" as any sentence of the other. This incomparability of expressivity is, arguably, undesirable. The expressive power of a language does not crucially depend on purely representational features of the logical space over which it is defined. Rather, it depends on the kinds of structures it can articulate on a given logical space.

More serious examples come from the study of nonclassical logics. It is sometimes argued that the double-negation translation from classical logic into intuitionistic logic shows that classical logic is really just a fragment of intuitionistic logic. For instance, Gödel [1933] (reprinted in Gödel [1986, p. 287]) articulates this viewpoint:

> If to the primitive notions of Heyting's propositional calculus we let correspond those notions of the classical propositional calculus that are denoted by the same sign and if to absurdity ($\neg$) we let correspond negation ($\sim$), then the intuitionistic propositional calculus H turns out to be a proper subsystem of the ordinary propositional calculus A. With another correlation (translation) of the notions, however, *the classical propositional calculus is*, conversely, *a subsystem of the intuitionistic one*.

It is natural to conclude from the fact that classical logic can be faithfully translated into intuitionistic logic that the language of intuitionistic logic has greater expressive power than the language of classical logic. A common way of thinking about intuitionistic logic, in light of this result, is that it is essentially an extension of classical logic with special operators (viz., '$\vee$' and '$\rightarrow$'). But the simple notion of expressive power does not yield this result, as the class of Kripke models for intuitionistic logic is disjoint from the class of classical valuations (even though the latter is, in some sense, clearly identifiable as a subclass of the former).

Being able to compare the expressive power of some languages that are defined over distinct conceptions of logical space is also important for philosophical purposes. In many cases, to assume that each side of a philosophical dispute has the same conception of logical space would be to prejudge the dispute in question. We still want to compare the expressive capacities of, say, absolutism and relationalism even if the latter refuses to recognize distinctions the former recognizes. Similarly, we want to be able to compare what can be said by the universalist and the nihilist even if neither leaves open any possibility that the former leaves open.

The main project of this dissertation is to explore different measures of expressive power and compare their strengths and weaknesses. It is *not* to answer the question of what expressive power truly is. I am not sure there is even an answer to this question. As we will see, there are many different notions of expressive power, each with their pros and cons, and there does not seem to be any obvious way of deciding what the one true notion of expressivity amounts to. But even if there is, I take it as an important first step to lay out some options and explore their various properties. That is what I aim to do here.

There are three notable aspects to my approach to this topic. The first is in the level of abstraction with which notions of expressive power are characterized. This aspect of my approach can be traced back to Alfred Tarski's work on logical consequence. While Tarski is known for his conception of logical consequence in terms of truth relative to all interpretations of the non-logical expressions [Tarski, 1983c], he was also keenly interested in more general conceptions of logical consequence, following from the work of Jan Łukasieciz and Stanisław Leśniewski. In Tarski 1983b, p. 30, for instance, logical consequence is thought of as an operator on sentences:

> From the sentences of any set X certain other sentences can be obtained by means of certain operations called *rules of inference*. These sentences are called the *consequences of the set* X. The set of all consequences is denoted by the symbol '$Cn(X)$'.

He stipulates five axioms that an operator Cn over a set S must satisfy to count as a consequence operator, presented below in modern notation [Tarski, 1983b, p. 31]:

Axiom 1. $|S| \leqslant \aleph_0$.

Axiom 2. If $X \subseteq S$, then $X \subseteq Cn(X) \subseteq S$.

Axiom 3. If $X \subseteq S$, then $Cn(Cn(X)) = Cn(X)$.

Axiom 4. If $X \subseteq S$, then $Cn(X) = \bigcup \{Cn(Y) \mid Y \subseteq X \;\&\; |Y| < \aleph_0\}$.

Axiom 5. There exists a sentence $x \in S$ such that $Cn(\{x\}) = S$.

Informally, these axioms state that: (1) there are only countably many formulas in total;[3] (2) consequence is monotonic and reflexive;[4] (3) consequence is transitive; (4) consequence is compact; and (5) there is a falsum sentence that implies everything.[5]

While it can be shown that any logic satisfying Axioms 2–3 can be characterized as the truth-preservation relation of some language (**Proposition 1.1.9**), this characterization of logical consequence is more abstract insofar as it makes *very few*

[3] Tarski [1983a, p. 63] seems to consider this a simplifying assumption.
[4] More accurately, that consequence is "monotonically reflexive". See **Proposition 1.1.9**.
[5] This axiom is dropped in Tarski 1983a.

assumptions about the nature of the syntax, models, or semantics of the language it is defined over. Tarski's aim is to "*make precise the meaning of a series of important metamathematical concepts* which are common to the special metadisciplines, *and to establish the fundamental properties of these concepts*." This abstract approach to logic was also adopted by Adolf Lindenbaum, Jerzy Łoś, and other members of the Warsaw School of Logic around the same time. Much of this work was later summarized (and generalized) by Wójcicki [1988] in an influential introduction to the subject, as well as those working the tradition of abstract algebraic logic.[6] Following in this tradition, this dissertation can be seen as an attempt to establish the fundamental properties of just *one* important metamathematical concept, viz., that of expressive power, without relying on any (or at least, relying on very few) assumptions about the underlying constitution of the language in question.

The second aspect of my approach to expressive power is its emphasis on translations. Given that we are interested in formulating a notion of expressive power that does not require the languages whose expressivity is being compared have commensurable conceptions of logical space, a natural question arises: what does "sameness of meaning" amount to? One promising answer to this question is given in the notion of a translation. Translations, the thought goes, are supposed to be maps from sentences to sentences (or expressions to expressions) that "preserve meaning". If that is so, then investigating what makes a translation adequate will help answer the question of what sameness of meaning amounts to in these cases, thereby giving us a way to measure expressive power.

While translations have always been the subject of much interest for logicians, a formal definition of a translation was (according to Feitosa and D'Ottaviano [2001]) first articulated by Prawitz and Malmnäs [1968], who did so *en route* to discussing the relationship between classical, intuitionistic, and minimal logic. Wójcicki [1988] and Epstein [1990] are among the first to investigate the notion of a translation more thoroughly, both as a general method for studying nonclassical logics and as an independently interesting subject. Much of the work in this dissertation is heavily influenced by these texts. Since then, there has been a flurry of work on the topic of translations as a way of characterizing the deductive strength of logical systems in the literature on abstract logic.[7]

This dissertation differs from the previous work on the subject of translations, however, in two respects. First, while previous work on translations has focused on translations between *logics*, one focus of this dissertation (especially in Chapters 4–5) is on translations between *languages*. It turns out that there is a way of moving back and forth between the two, so previous work on translations between logics can be easily imported to the study of translations between languages.

[6]See Font 2016 for a helpful introduction.

[7]See Carnielli and D'Ottaviano 1997; Carnielli et al. 2009; Feitosa and D'Ottaviano 2001; Pelletier and Urquhart 2003; Caleiro and Gonçalves 2007; Straßburger 2007; Mossakowski et al. 2009; French 2010; Jeřábek 2012; Wigglesworth 2017; Woods 2018.

Second, previous work on translations has tended to largely focus on propositional languages and has thereby ignored languages with quantification.[8] This is entirely reasonable, since the study of translations is already difficult enough without having to worry about the complexities quantifiers introduce. But I think there is much to be learned from the study of quantificational languages. For example, a now well-known result due to Jeřábek [2012] (reviewed in **Theorem 2.6.8**) shows that a great number of languages can be faithfully translated into classical propositional logic. In particular, the result shows that first-order logic can be translated into propositional logic. In fact, it will be shown later in this dissertation (**Theorem 2.6.5** and **Corollary 2.6.7**) that one can even faithfully translate first-order logic into propositional logic bijectively. Moreover, the translational relationship between first-order logic and modal logic has been largely overlooked and so deserves attention (**Proposition 3.3.16**).[9]

The final notable aspect of my approach is in its application to metaphysics. In the study of expressive power, it is hard to ignore the overwhelmingly strong connections to recent discussions in metametaphysics over what makes a dispute "merely verbal." Various claims about the relative expressive power of a particular metaphysical view have been made. As an illustration, consider again the debate between mereological universalism and mereological nihilism. It is often said that whatever the universalist says about tables, the nihilist can say about "atoms arranged table-wise". For some, this point is convincing enough: if everything the universalist can say can also be said by the nihilist, albeit in other terms, then in what sense are the two sides in genuine disagreement? For others, it is hardly persuasive: it matters not what each side *can* say, but what they *do* say.

Part of the difficult in assessing these "metadisputes" is in the fact that the two sides are starting out with radically different assumptions about the debate. The "realist" interpretation of this dispute assumes that we can take for granted that the universalist and the nihilist *mean the same thing* by the words 'part' and 'whole', and are simply disagreeing over the nature of parthood. The "anti-realist" interpretation, by contrast, does not assume that this can be taken for granted, and it takes seriously the idea that each side is simply using the words 'part' and 'whole' differently. And it is difficult to say which interpretation of these sorts of disputes is the right one to take. One goal of the final chapter of this dissertation is to try to use the framework of translations and a more general characterization of expressive power to provide common ground for assessing these sorts of metadisputes. In doing so, I aim to show that even on an anti-realist interpretation, genuine disagreement between the opposing sides in these disputes is still possible.

[8]Some notable exceptions: Barrett and Halvorson 2016a,b.

[9]With that said, I have tried to integrate discussion of quantifiers into a more general framework by thinking of quantifiers are *operators*, rather than as binding devices. It would be worth investigating ways of generalizing the simple pictures of language discussed in this dissertation to fit more neatly with quantificational languages, but the issue is not pursued much here.

Below is a chapter summary of the dissertation.

Chapter 1: Preliminaries. Before we can measure expressive power, we need to get clearer on what a language is. The simplest picture of a language is this: a language is some syntax (a set of well-formed formulas) together with a class of indices (points of evaluation) and a satisfaction relation specifying which indices satisfy which formulas (or, equivalently: an assignment of sentences to semantic values, i.e., sets of indices). Similarly, a logic can be modeled as a syntax paired with a consequence relation (a relation between sets of formulas and formulas).

This chapter introduces a number of important concepts relating to languages and logics that will be used throughout the dissertation. In particular, we define (1) the notion of a Tarskian logic, showing that they are precisely the logics that can be the entailment relations of some language, (2) the notion of a fragment of a language, (3) the notion of a theory, showing that the space of theories in a language form a complete lattice, and (4) the notion of a Lindenbaum-Tarski algebra. We also consider two refinements of the simple picture of language: one which views sentences as composed from atomic sentences together with some operations, and one which views meaning as context-change potential rather than as satisfaction conditions.

Chapter 2: Translation. The notion of expressive power is closely tied to the notion of a translation. Intuitively, one language is as expressive as another just in case the latter is translatable into the former. And two languages are expressively equivalent just in case they are intertranslatable. But what exactly counts as a "translation" from one language to another? One natural idea is that a translation from one language to another must "preserve the meaning" of each sentence of the source language in the target language. However, it is not obvious how to spell out the notion of meaning preservation. In particular, one cannot spell out the notion simply by stipulating that sentences from different languages mean the same thing if they are given the same semantic value. For many languages whose expressive power seem readily comparable do not share the same conception of logical space. As a simple illustration, classical logic and intuitionistic logic are not defined relative to the same class of models. Yet we still want to be able to compare the expressive power of each language (e.g., many theorists want to say intuitionistic logic is more expressive than classical logic).

This chapter explores a relatively simple proposal. The simple proposal is motivated by the idea that at least one necessary criterion for a mapping to count as a translation is that it faithfully preserve inferential connections between the sentences of the language. Thus, if $\mathtt{t}$ is a translation from a language $\mathbf{L}_1$ into $\mathbf{L}_2$, then it ought to at least be the case that an argument is valid in $\mathbf{L}_1$ iff the translation of this argument via $\mathtt{t}$ is valid in $\mathbf{L}_2$. According to the notion of a translation explored in this chapter, the converse is also true: $\mathtt{t}$ is a translation if it faithfully preserves the inferential connections between sentences. While this notion of a translation is fairly minimalistic, it is not uninformative: there are a number of natural and interesting examples of languages not being translatable into others.

Two more tasks are accomplished in this chapter. First, a distinction is drawn between different notions of equivalence between languages (intertranslatability, translational equivalence, and isomorphism). It is argued that the weakest of these notions (intertranslatability) roughly corresponds to expressive equivalence, whereas the middle notion (translational equivalence) roughly corresponds to notational variance. Second, it is shown that this notion of a translation is indeed too weak to formalize expressive power. This is because it can be shown that many seemingly powerful languages can be translated into classical propositional logic. In particular, it is shown that full first-order logic is translationally isomorphic to propositional logic. This shows that we need to place further constraints on translatability in order to achieve an adequate formulation of the intuitive notions of "expressivity" and "notational variance".

Chapter 3: Compositionality. The results from the previous chapter suggest that we need more restricted criteria on translations to get adequate precisifications of expressive power and notational variance. A natural restriction to place on translations is this: translations need to be *compositional*, in the sense that they translate complex formulas as functions of the translations of the parts. This chapter essentially develops different ways of understanding this idea in more detail. Some of these notions of compositionality are too weak insofar as they still render first-order logic and propositional logic equivalent. Others are too strong, as they rule out a number of intuitive examples of expressive equivalence. In between is a notion I call *schematicity*: the translation of a complex formula ought to be a fixed schema of the translation of its parts. This notion of schematicity is refined and systematically characterized in a number of different ways. It is shown, for instance, that schematic equivalence can be characterized as another notion sometimes called "definitional equivalence". The notion is generalized to cover cases where multiple translations are recursively defined in terms of others in a schematic way.

Chapter 4: Logical Space. The previous two chapters focused on notions of translation that were defined in terms of inference-preservation. This chapter looks at several different notions of translation defined in terms of truth-preservation. The idea is this: translations need to not only preserve the validity of arguments, but they must also preserve truth at an index. Of course, if the languages in question have different indices relative to which truth is evaluated, then we need to say more about what preserving truth at an index amounts to. This can be done if translations translate indices as well as sentences. Richard Epstein calls these translations "model-preserving". Model-preserving maps automatically faithfully preserve consequence. Several different characterizations of model-preservation are proven. In particular, it is shown that model-preservation is equivalent to inference-preservation given that the languages involved have a special property (canonically had by classical languages) that I call "opinionation", which states (very roughly) that if a consistent set of formulas fails to imply another formula, then that set can be expanded to a maximally consistent set while still failing to imply that formula. Thus, for classical languages with boolean negation, the minimalist notion of a translation is equivalent to the stronger model-preserving notion. The notions diverge, however, for nonclassical languages.

Chapter 5: Metametaphysics. In this final chapter, the technical apparatus of the previous chapters is applied to metametaphysics—particularly to the debate over which metaphysical disputes are "verbal". A number of metaphysical disputes seem completely irresolvable in the sense that no side can seem to get any footing against the other. There are two common attitudes to have towards apparently irresolvable disputes. On the one hand, there are the ***realists*** about a dispute, who hold that there is an objective fact of the matter as to which side of the dispute (if any) is correct. According to the realist about a metaphysical dispute, the disputants are in genuine disagreement and the irresolvability of the dispute arises simply from the fact that the question under discussion is difficult to answer. On the other hand, there are the ***anti-realists*** about a dispute, who hold that there is no objective fact of the matter as to which side of the dispute is correct. According to the anti-realist about a metaphysical dispute, the disputants are "just talking past one another" and their theories are "notational variants" of one another.

It is commonly thought that anti-realism entails an end to a dispute. Anti-realists about a dispute, it is thought, will hold that there simply is no point in having the dispute in the first place: you can choose to talk however you like. The goal of this chapter is to show that this need not be the case. Even if we adopt an anti-realist framework, according to which the disputants in question ought to be modeled as speaking different languages, there can still be objective disagreements these disputants can have with one another over which language one ought to be using for theorizing.

To do this, I apply the formal framework from the previous chapters to two disputes, both of which are classic examples of purportedly verbal disputes. The first is the dispute in mereology between nihilism (the view that there are no composite objects) and universalism (the view that composition is unrestricted—any set of objects compose a composite object). The second is the dispute over persistence between four-dimensionalism (the view that objects persist through time via temporal parts) and three-dimensionalism (the view that objects persist through time by being "wholly present"). Using the framework of expressive power discussed previously, we can see more clearly in what sense, if any, these disputes are verbal.

In the case of mereology, it is shown that the universalism-nihilism dispute is only verbal if we assume atomism. Otherwise, the universalist has strictly greater expressive power than nihilism, and so the dispute between them can be seen as one over whether to recognize certain possibilities that the universalist recognizes as genuine.

In the case of persistence, while the four-dimensionalist and three-dimensionalist technically have matching expressive powers when matching mereological principles are associated with each view, they seem to differ in expressive power when we assume each view is associated with different mereological principles. In particular, four-dimensionalism is typically associated with mereological principles that are diachronic in nature, whereas three-dimensionalism is associated with mereological principles that are synchronic. This is sufficient to separate the views in their expressivity.

Much can be said about expressive power. But it should be noted that I did not say everything that could be said. This dissertation leaves out some important aspects of the study of expressive power. One lacuna concerns the expressive power of non-Tarskian languages. The notion of expressive power for languages whose entailment relation is not monotonic, contractive, commutative, or reflexive has yet to be investigated. Important examples of such languages include dynamic languages, which may fail to have all of these properties. Though dynamic languages are discussed briefly in § 1.4, it is only to lay out special circumstances when the results of this dissertation can be transferred over to those languages, not to say how to generalize the framework to these languages entirely. Other important examples of non-Tarskian languages include nonclassical languages such as relevance logic, linear logic, and other substructural logics. Much work on the issue of expressive power for these languages still has yet to be done.

Another important aspect of expressive power not discussed in this dissertation is its relation to computational complexity. It is part of the folklore in logic that there is a trade-off between expressive power and computational complexity: in general, the more expressive your language is, the more computationally complex it becomes. It would be interesting to formalize this folklore in a general way. I suspect the framework developed in this dissertation could shed some light on this issue; but I leave this task for future work. A few remarks about using computational complexity in the definition of expressive power are made in § 2.6.3; but otherwise, the topic is mostly untouched here.

Connections between expressive power and other important metamathematical concepts are also lacking. For example, nothing is said about Lindström's theorems, which states that first-order logic is the most expressive logic with compactness and downward Löwenheim-Skolem properties. Nor is anything said about interpolation theorems and their relation to definability [Hoogland, 2001]. There is always a tradeoff between breadth and depth: it is difficult to see what can be said about such connections at the level of abstraction this dissertation employs. So I have omitted discussion of them, if only for lack of things to say.

Finally, the philosophical applications of this framework to metaphysical disputes is far from complete, even on its own terms. The results presented in Chapter 5 are more of a proof of concept rather than a decisive demonstration that the framework of translations and expressive power can clarify the status of these metaphysical disputes. While I do think these results help illustrate the ways in which these disputes can be viewed as substantive, and while I think some initial conclusions can be safely drawn from them, I do not think the project can be thought of as completed at this point. More must be done before we can say for certain what exactly the relationship is between these different metaphysical views.

Chapter 1

Language

Before we can explore how to measure the expressive power of a language, we must first get clear on what a language *is*. In this chapter, we lay out a variety of ways one can go about defining a language in more precise terms.

We begin with a very simple conception of a language—a conception that is sometimes referred to as "model-theoretic semantics" and goes at least as far back as Tarski.[1] On this conception, a language consists of three parts. First, there is a syntax, which tells you what the well-formed formulas are in the language. Second, there is a class of interpretations relative to which these formulas can be evaluated. And third, there is a semantics, which tells you under what interpretations any given formula is satisfied.

Here is the intuitive idea behind this picture of language. One starts with a notion of *logical space*—the space consisting of every possibility, i.e., every way things could have been. Logical space acts as the object of interpretation or evaluation for a language. A language uses formulas of a language to "carve" logical space. Each formula carves logical space into two regions: the region of logical space where the satisfaction conditions of the formula are met and the region where they are not. Thus, a language can be thought of as a way of *partitioning* logical space.

Logical space consists of all the "possibilities". The flavor of possibility here is deliberatively left open. Different applications will require different interpretations. For some purposes (e.g., when engaging in metaphysical theorizing), it will be natural to interpret logical space as consisting of the metaphysical possibilities. For other purposes (e.g., when modeling natural language discourse), it will be natural to interpret logical space as consisting of another kind of possibility (e.g., epistemic). Some of these purposes will be our purposes. But at least one of our purposes is more general. The goal is to characterize different notions of expressive power in abstract terms and to systematically relate them. So for *our* purposes, we can often interpret logical space as consisting of the *logical* or *conceptual* pos-

[1]See Tarski 1944, 1983c.

sibilities. These are the possibilities that are not ruled out solely in virtue of the conventions governing one's language. Maybe these possibilities are epistemically ruled out, or metaphysically impossible, or otherwise unpleasant. But that will not concern us as long as the possibilities described are logically permissible.

In the chapters to follow, we will often use this simple conception of a language as a starting point for understanding different notions of expressive power. This is for two reasons. First, the different metrics of expressive power are easier to introduce using simple languages, and the main ideas motivating these various metrics can already be stated at the level of simple languages. Second, it is instructive to see how far we can go without complicating our picture of language. That way, when we do complicate that picture, we can see exactly what additional benefits are gained in doing so. So to begin, we will develop this simple conception of a language in § 1.1.

Nevertheless, language is obviously not so simple. While the simple conception of language makes stating the different metrics of expressive power easier, there are a number of respects in which such metrics will be deficient without making our picture of language more sophisticated. One way in which the picture can be made more sophisticated is by adding compositional structure. Thus, not only do we have formulas that carve logical space, but also we have *operators* that transform some ways of carving logical space into others. We will develop a compositional conception of language in § 1.3.

Another way of complicating the picture that has gained traction in the formal semantics literature is to make languages *dynamic* rather than *static*. On static conceptions of a language, the meaning of a formula is given by its *satisfaction conditions* (or *truth conditions*). That is, formulas are interpreted as subregions of logical space. On a dynamic conception of language, however, the meaning of a sentence is determined by how it affects the conversational context, i.e., its *context change potential*. Formulas, then, are interpreted as functions from contexts to contexts. (The dynamic conception of language, of course, can be combined with the compositional conception.) This idea is developed further in § 1.4.

§ 1.1 Simple Languages

We start by sketching a simple picture of languages. Here is a brief outline of this section. § 1.1.1 simply defines the notion of a (simple) language and a logic. It is observed that the logics that are the entailment relation of some language are exactly those that satisfy some natural constraints and are sometimes referred to as "Tarskian" logics. § 1.1.2 briefly introduces the notion of a fragment/extension of a language/logic and draws connections between these notions. § 1.1.3 discusses the notion of a theory relative to a language or logic. It is shown that the theories of a language/logic form a nice structure.

§ 1.1.1 Languages and Logics

The simple picture of language is, again, one on which languages consist of a syntax, a logical space, and a semantics. We make this more precise as follows:

Definition 1.1.1 *(Language)*. A ***language*** is triple $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash \rangle$, where:

- $\mathcal{L}$ is a class (the ***syntax***)
- C is a class (the ***evaluation space***)
- $\Vdash \subseteq \mathsf{C} \times \mathcal{L}$ is a relation (the ***satisfaction relation***).

The members of $\mathcal{L}$ are called the ***formulas*** in $\mathcal{L}$ or the ***$\mathcal{L}$-formulas***. The members of C are called ***points of evaluation***, ***indices***, or ***states***.

Convention: If $\mathbf{L}$ is a language, we may use "$\mathcal{L}_{\mathbf{L}}$", "$\mathsf{C}_{\mathbf{L}}$", and "$\Vdash_{\mathbf{L}}$" to denote respectively its syntax, evaluation space, and satisfaction relation. Often, we replace a language's label in subscripts and superscripts with that label's salient subscripts or superscripts. For instance, if $\mathbf{L}_1$ is a language, we often replace "$\mathcal{L}_{\mathbf{L}_1}$" with "$\mathcal{L}_1$", $\mathsf{C}_{\mathbf{L}_1}$" with "$\mathsf{C}_1$", and "$\Vdash_{\mathbf{L}_1}$" with "$\Vdash_1$". The same goes for other notation that is relativized to a language.

A couple of remarks about **Definition 1.1.1** are in order. First, I am allowing that $\mathcal{L}$ and C be proper classes, not just sets. Often times, syntaxes are set-sized, but in some languages (e.g., in infinitary languages), there are class-many formulas. And almost by default, the evaluation space of a language is a proper class. But for the sake of readability, the notation I use for classes will be identical to the notation used for set. Thus, we may write "$\mathsf{X} = \{x \in \mathsf{C} \mid \dots\}$", even if X is a proper class. Similarly, I use "$\subseteq$" for "subclass", "$\in$" for membership, "$\cap$" for class intersection, and so forth. The distinction between classes and sets will not be that important here, so there is no harm in reading this notation ambiguously in this way.

Second, **Definition 1.1.1** says nothing yet about the nature of the syntax of a language. We do not require, from the outset, that the syntax of a language be generated in some uniquely recursive manner: the syntax of a language can be non-well-founded, non-recursively defined, non-arithmetical, and so forth. In effect, the syntax of a language is just a useful way of labeling certain regions of the evaluation space. While most of our focus will be on languages whose syntax is well-behaved and cleanly generated, we want to first see how far we can go with just this simple general characterization of a language.

Third, **Definition 1.1.1** says nothing yet about the nature of the evaluation space. One can interpret the evaluation space either as states in a fixed model or as a space of models. Points of evaluation can be thought of either representationally or interpretationally, in the sense of Etchemendy [1990]. They could be possible states of the world, as in possible world semantics. They could be pairs of models and variable assignments, as in first-order logic. They could be contexts of use in the sense of Kaplan [1977]. They could be pairs of worlds in a model, as in two-dimensional modal logic. They could be information states, as in expressivist semantics for epistemic modals. What you fill in for the evaluation space is completely flexible.

Finally, **Definition 1.1.1** says nothing yet about the nature of the satisfaction relation. In particular, the satisfaction relation does not have to be defined recursively or compositionally. And even if it is, it does not have to be defined recursively in terms of truth-at-a-point in the evaluation space. One can, for instance, define truth-at-a-context in terms truth-at-an-index, and then stipulate that the evaluation space consists only of contexts. Similarly, one might define truth-at-a-pair of worlds, and then restrict the evaluation space to diagonal pairs of worlds. So we can still model languages where consequence is meant to be defined as preservation of truth-at-a-context or at-a-diagonal point.

We said earlier that we can think of the meaning of a formula as a subregion of logical space. We could have defined languages in these terms from the start; that is, we could have replaced the satisfaction relation in **Definition 1.1.1** with a *semantic value* function that maps each formula to a subregion of logical space.

Definition 1.1.2 *(Semantic Value)*. Let $\mathbf{L}$ be a language and let $\phi \in \mathcal{L}_\mathbf{L}$. The ***L*-semantic value** of ϕ (written "$[\![\phi]\!]_\mathbf{L}$") is defined as the class of points of evaluation x such that $x \Vdash_\mathbf{L} \phi$, i.e., $[\![\phi]\!]_\mathbf{L} = \{x \in \mathsf{C}_\mathbf{L} \mid x \Vdash_\mathbf{L} \phi\}$. If $\Gamma \subseteq \mathcal{L}_\mathbf{L}$, the ***L*-semantic value** of Γ is defined as $[\![\Gamma]\!]_\mathbf{L} = \bigcap_{\psi\in\Gamma} [\![\psi]\!]_\mathbf{L}$. Notice that if $\Gamma = \varnothing$, then $[\![\Gamma]\!]_\mathbf{L} = \mathsf{C}_\mathbf{L}$.

Notation: Where $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash\rangle$ is a language, we write "$x \Vdash \Gamma$" to mean "$x \Vdash \psi$ for all $\psi \in \Gamma$". Notice that $[\![\Gamma]\!] = \{x \in \mathsf{C} \mid x \Vdash \Gamma\}$.

Each language naturally gives rise to a notion of consequence:

Definition 1.1.3 *(Entailment)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash\rangle$ be a language. Then where $\Gamma \subseteq \mathcal{L}$ and $\phi, \psi \in \mathcal{L}$, we say that:

- Γ ***L*-entails** ϕ (written "$\Gamma \vDash_\mathbf{L} \phi$") if for all $x \in \mathsf{C}$: if $x \Vdash \Gamma$, then $x \Vdash \phi$.
- ϕ is ***L*-valid** (written "$\vDash_\mathbf{L} \phi$") if $\varnothing \vDash_\mathbf{L} \phi$.
- ϕ is ***L*-equivalent to** ψ (written "$\phi \equiv_\mathbf{L} \psi$") if $\phi \vDash_\mathbf{L} \psi$ and $\psi \vDash_\mathbf{L} \phi$.

Where $\Gamma, \Delta \subseteq \mathcal{L}$, we write "$\Gamma \vDash_\mathbf{L} \Delta$" to mean "$\Gamma \vDash_\mathbf{L} \phi$ for all $\phi \in \Delta$". Likewise, we write "$\Gamma \equiv_\mathbf{L} \Delta$" to mean "$\Gamma \vDash_\mathbf{L} \Delta$ and $\Delta \vDash_\mathbf{L} \Gamma$". We may drop set brackets for readability (e.g., writing "$\Gamma, \phi \vDash \psi$" instead of "$\Gamma \cup \{\phi\} \vDash \psi$").

Note that "$\Gamma \vDash \Delta$" is interpreted as "if every member of Γ is true, then *every* member of Δ is true". This is contrary to how it is interpreted in proof theory, where $\Gamma \vDash \Delta$ says "if every member of Γ is true, then *some* member of Δ is true". This break from standard conventions is notationally more convenient for our purposes.

Fact 1.1.4 *(Equivalent Definition of Entailment)*. Let $\mathbf{L}$ be a language, let $\Gamma \subseteq \mathcal{L}_{\mathbf{L}}$, and let $\phi \in \mathcal{L}_{\mathbf{L}}$. Then $\Gamma \vDash_{\mathbf{L}} \phi$ iff $[\![\Gamma]\!]_{\mathbf{L}} \subseteq [\![\phi]\!]_{\mathbf{L}}$.

The notion of entailment from **Definition 1.1.3** is the familiar notion of truth-preservation: Γ entails ϕ if every point that satisfies all of Γ satisfies ϕ. As is well-known, there notion of entailment has a convenient but equivalent definition. To explain, we should first define the notion of a logic.

Definition 1.1.5 *(Logic)*. A ***logic*** is a pair $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$, where:

- $\mathcal{L}$ is a class of formulas
- $\vdash \subseteq \wp(\mathcal{L}) \times \mathcal{L}$ (the ***consequence relation***).

If $\Gamma \subseteq \mathcal{L}$ and $\phi \in \mathcal{L}$, we say Γ ***L-proves*** ϕ if $\Gamma \vdash \phi$. We write "$\Gamma \vdash \Delta$" to mean "$\Gamma \vdash \phi$ for all $\phi \in \Delta$". We write "$\phi \dashv\vdash \psi$" to mean "$\phi \vdash \psi$ and $\psi \vdash \phi$".

In one sense, this definition of a logic is somewhat restricted. Consequence relations automatically obey a number of structural properties that are rejected by a variety of logics. For instance, consequence relations are commutative: the order of premises does not matter. They are also automatically contractive: repetitions of a premise do not affect the validity of an argument. Many logics rejects one or both of these features. However, since our focus is not on such logics, we need not be troubled by the restrictions on consequence relations imposed by **Definition 1.1.5**.[2]

In another sense, however, this definition of a logic is quite broad. Consequence relations, as defined in **Definition 1.1.5**, need not be reflexive, transitive, or have many of the other structural properties that most logics are assumed to have. However, we will be primarily focused on logics that have many of these nice properties.

Definition 1.1.6 *(Tarskian Logic)*. A logic $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ is ***Tarskian*** if the following three conditions hold for all $\phi \in \mathcal{L}$ and all $\Gamma, \Gamma', \Delta \subseteq \mathcal{L}$:

(CR1) $\phi \vdash \phi$ (***Reflexivity***)

(CR2) if $\Gamma \subseteq \Gamma'$ and $\Gamma \vdash \Delta$, then $\Gamma' \vdash \Delta$ (***Monotonicity***)

(CR3) if $\Gamma \vdash \Gamma'$ and $\Gamma' \vdash \Delta$, then $\Gamma \vdash \Delta$ (***Transitivity***)

[2] These restrictions can be avoided if we treat $\mathcal{L}$ not as a class of *formulas* but rather as a class of *formula-structures* [Font, 2016, pp. 14–15].

Tarskian logics have some nice properties. Chief among them is the following:

Fact 1.1.7 *(Replacement of Equivalents)*. Let $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ be a Tarskian logic, let $\Gamma \subseteq \mathcal{L}$ and let $\phi, \psi, \theta \in \mathcal{L}$.

(a) If $\phi \dashv\vdash \psi$, then $\Gamma, \phi \vdash \theta$ iff $\Gamma, \psi \vdash \theta$. (***Replacement of Equivalent Premises***)

(b) If $\phi \dashv\vdash \psi$, then $\Gamma \vdash \phi$ iff $\Gamma \vdash \psi$. (***Replacement of Equivalent Conclusions***)

The name "Tarskian" is in deference to Tarski [1983b,a]. However, Tarski [1983b, p. 31] defines a logic as a pair of a set of formulas $\mathcal{L}$ together with a consequence *operator*, which is interdefinable with Tarskian consequence *relations*.[3]

Definition 1.1.8 *(Consequence Operator)*. Let $\mathcal{L}$ be some class of formulas. A ***consequence operator*** for $\mathcal{L}$ is a function $\mathtt{Cn}\colon \wp(\mathcal{L}) \to \wp(\mathcal{L})$ such that the following three conditions hold for all $\Gamma, \Delta \subseteq \mathcal{L}$:

(CO1) $\Gamma \subseteq \mathtt{Cn}(\Gamma)$ (***Reflexivity***)

(CO2) If $\Gamma \subseteq \Delta$, then $\mathtt{Cn}(\Gamma) \subseteq \mathtt{Cn}(\Delta)$ (***Monotonicity***)

(CO3) $\mathtt{Cn}(\Gamma) = \mathtt{Cn}(\mathtt{Cn}(\Gamma))$ (***Transitivity***)

We write "$\mathtt{Cn}(\phi_1, \ldots, \phi_n)$" in place of "$\mathtt{Cn}(\{\phi_1, \ldots, \phi_n\})$".

Proposition 1.1.9 *(Alternative Definitions of Tarskian Logic)*. Let $\mathcal{L}$ be a class of formulas, let $\vdash \subseteq \wp(\mathcal{L}) \times \mathcal{L}$, and let $\mathtt{Cn}(\Gamma) := \{\phi \in \mathcal{L} \mid \Gamma \vdash \phi\}$. The following are equivalent:

(a) L is Tarskian.

(b) The following two properties hold for all $\Gamma, \Gamma', \Delta \subseteq \mathcal{L}$:

(CR4) $\Gamma \vdash \Gamma$ (***Monotonic Reflexivity***)

(CR3) if $\Gamma \vdash \Gamma'$ and $\Gamma' \vdash \Delta$, then $\Gamma \vdash \Delta$ (***Transitivity***)

(c) $\mathtt{Cn}$ is a consequence operator for $\mathcal{L}$.

(d) There is a language $\mathbf{L}$ with syntax $\mathcal{L}$ such that $\vdash\ =\ \vDash_{\mathbf{L}}$.

[3]The axioms (CO1)–(CO3) are not the only axioms that he required consequence operators to satisfy. In particular, he also includes (CO4) below as well as an axiom stating that $|\mathcal{L}| \leqslant \aleph_0$.

Proof:

(a) ⇒ (b). (CR4) follows from (CR1) and (CR2).

(b) ⇒ (c). (CO1) follows from (CR4). For (CO2), if $\Gamma \subseteq \Delta$, then $\Delta \vdash \Gamma$ by (CR4). Hence, if $\Gamma \vdash \Sigma$, then $\Delta \vdash \Sigma$ by (CR3). Finally, for (CO3), it suffices to show the right-to-left direction since the left-to-right direction follows from (CO1) and (CO2). Suppose $\phi \in \mathtt{Cn}(\mathtt{Cn}(\Gamma))$. That means $\mathtt{Cn}(\Gamma) \vdash \phi$. But $\Gamma \vdash \mathtt{Cn}(\Gamma)$ by definition. So $\Gamma \vdash \phi$ by (CR3), i.e., $\phi \in \mathtt{Cn}(\Gamma)$.

(c) ⇒ (d). Let $\mathsf{C} = \{\mathtt{Cn}(\Gamma) \mid \Gamma \subseteq \mathcal{L}\}$. Define $\Vdash \subseteq \mathsf{C} \times \mathcal{L}$ so that $\mathtt{Cn}(\Gamma) \Vdash \phi$ iff $\phi \in \mathtt{Cn}(\Gamma)$. We show that $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash \rangle$ is such that $\vdash \, = \, \vDash_{\mathbf{L}}$, i.e., that $\phi \in \mathtt{Cn}(\Gamma)$ iff for all $\Delta \subseteq \mathcal{L}$, if $\mathtt{Cn}(\Delta) \Vdash \Gamma$, then $\mathtt{Cn}(\Delta) \Vdash \phi$.

First, suppose $\phi \in \mathtt{Cn}(\Gamma)$. Let $\Delta \subseteq \mathcal{L}$ be such that $\mathtt{Cn}(\Delta) \Vdash \Gamma$. Thus, $\Gamma \subseteq \mathtt{Cn}(\Delta)$. By (CO2) and (CO3), $\mathtt{Cn}(\Gamma) \subseteq \mathtt{Cn}(\mathtt{Cn}(\Delta)) = \mathtt{Cn}(\Delta)$. So $\phi \in \mathtt{Cn}(\Delta)$, and thus $\mathtt{Cn}(\Delta) \Vdash \phi$.

Next, suppose $\phi \notin \mathtt{Cn}(\Gamma)$. By (CO1), $\Gamma \subseteq \mathtt{Cn}(\Gamma)$, and so $\mathtt{Cn}(\Gamma) \Vdash \Gamma$. But since $\phi \notin \mathtt{Cn}(\Gamma)$, we have $\mathtt{Cn}(\Gamma) \nVdash \phi$, which completes the proof.

(d) ⇒ (a). Trivial, since any entailment relation satisfies (CR1)–(CR3). ■

The construction used in proving the implication from (c) to (d) will be discussed in more detail in § 1.1.3.

Tarski [1983b, p. 31] also required consequence operators to satisfy the following additional axiom:

(CO4) There is an $\mathcal{L}$-formula $\bot$ such that $\mathtt{Cn}(\bot) = \mathcal{L}$ (***Absurdity***)

Although we often do have such a formula present in our languages, we would prefer to frame our results concerning expressive power so as to cover languages without such a formula. Indeed, this axiom is later dropped from the definition of consequence operators in Tarski [1983a]. Even so, the entailment relation for a language does satisfy explosion in the following sense:

Fact 1.1.10 *(Explosion)*. Let $\mathbf{L}$ be a language and let $\Gamma \subseteq \mathcal{L}_{\mathbf{L}}$. If there is no $x \in \mathsf{C}_{\mathbf{L}}$ such that $x \Vdash_{\mathbf{L}} \Gamma$, then $\Gamma \vDash_{\mathbf{L}} \mathcal{L}_{\mathbf{L}}$.

It should be noted that satisfying "explosion" in this sense is much weaker than obeying *ex contradictione quodlibet*, i.e., $\phi, \neg\phi \vDash \psi$ for all ϕ and ψ. Many paraconsistent logics are explosive in the sense of **Fact 1.1.10**, even if only trivially as no set of formulas is unsatisfiable (e.g., see **Example 2.1.12** for an example of a Tarskian paraconsistent logic).

As noted above, a fair number of interesting logics are not Tarskian. But in the present context, Tarskian logics have a number of nice properties (especially **Fact 1.1.7**) that make them easier to work with than their non-Tarskian counterparts. So for now, we adopt the following convention:

> **Convention**: Henceforth, when I say "logic", I mean "Tarskian logic" unless stated otherwise.

By **Proposition 1.1.9**, every language $\mathbf{L}$ determines a unique (Tarskian) logic $\langle \mathcal{L}, \vDash_{\mathbf{L}} \rangle$. Hence, we introduce the following notation:

> **Convention**: Where $\mathbf{X}$ is a language, we write "$\mathsf{L}_{\mathbf{X}}$" or just "X" for the logic $\langle \mathcal{L}_{\mathbf{X}}, \vDash_{\mathbf{X}} \rangle$. We write "$\mathtt{Cn}_{\mathbf{L}}$" for the consequence operator for the language $\mathbf{L}$ and "$\mathtt{Cn}_{\mathsf{L}}$" for the consequence operator for the logic L.

§ 1.1.2 Fragments

Often enough, we will be interested in exploring the expressive power of various fragments of the languages we are interested in. This requires a more precise definition of a fragment.

> **Convention**: In general, "$\upharpoonright_X$" is used to mean "restricted to X." So for instance, "$\Vdash\upharpoonright_{\mathsf{C}' \times \mathcal{L}'}$" denotes the relation $\Vdash \cap\, (\mathsf{C}' \times \mathcal{L}')$. We also more compactly write "$\Vdash\upharpoonright_{\mathsf{C}'}$" for $\Vdash\upharpoonright_{\mathsf{C}' \times \mathcal{L}}$ and "$\Vdash\upharpoonright_{\mathcal{L}'}$" for $\Vdash\upharpoonright_{\mathsf{C} \times \mathcal{L}'}$ (likewise for "$\vDash$" and other notation).

> **Definition 1.1.11** *(Expansions and Restrictions)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages. We say that $\mathbf{L}_1$ is a ***restriction*** of $\mathbf{L}_2$ or that $\mathbf{L}_2$ is an ***expansion*** of $\mathbf{L}_1$ (written "$\mathbf{L}_1 \Subset \mathbf{L}_2$"), if:
>
> (i) $\mathcal{L}_1 \subseteq \mathcal{L}_2$
>
> (ii) $\mathsf{C}_1 \subseteq \mathsf{C}_2$
>
> (iii) $\Vdash_1 = \Vdash_2\upharpoonright_{\mathsf{C}_1 \times \mathcal{L}_1}$—that is, for all $\phi \in \mathcal{L}_1$, $[\![\phi]\!]_1 = [\![\phi]\!]_2 \cap \mathsf{C}_1$.
>
> We say that $\mathbf{L}_1$ is a ***conservative restriction*** of $\mathbf{L}_2$ or that $\mathbf{L}_2$ is a ***preservative expansion*** of $\mathbf{L}_1$ (written "$\mathbf{L}_1 \underline{\Subset} \mathbf{L}_2$"), if $\mathbf{L}_1$ is a restriction of $\mathbf{L}_2$ and in addition:
>
> (iv) $\vDash_1 = \vDash_2\upharpoonright_{\mathcal{L}_1}$—that is, for all $\Gamma \subseteq \mathcal{L}_1$ and $\phi \in \mathcal{L}_1$, $\Gamma \vDash_1 \phi$ iff $\Gamma \vDash_2 \phi$.

To restrict a language, one removes either points of evaluations from the evaluation space or formulas from the syntax (or both) while keeping the semantic values for the remaining formulas relative to the remaining evaluation space fixed.

To conservatively restrict a language, one must also take care not to remove every counterexample to any particular inference over the restricted syntax.

Fact 1.1.12 *(Expansion is a Partial Order)*. The relations $\Subset$ and $\underline{\Subset}$ are partial orders, i.e., reflexive, antisymmetric, and transitive.

Fact 1.1.13 *(Expansions Reflect Entailment)*. Let $\mathbf{L}_1 \Subset \mathbf{L}_2$. Then $\vDash_1 \supseteq \vDash_2\restriction_{\mathcal{L}_1}$, i.e., for all $\Gamma \subseteq \mathcal{L}_1$ and $\phi \in \mathcal{L}_1$, if $\Gamma \vDash_2 \phi$, then $\Gamma \vDash_1 \phi$.

Fact 1.1.13 is prone to lead to confusion. The logic of an expansion of $\mathbf{L}$ is *not* an "expansion" of the logic of $\mathbf{L}$. Rather, when one expands a language, one usually *weakens* the logic (at least over the original syntax). To avoid confusion, we adopt different vocabulary for talking about "expansions" of logics:

Definition 1.1.14 *(Sublogic)*. Let L_1 and L_2 be logics. We say that L_1 is a ***sublogic*** of L_2 (written "$\mathsf{L}_1 \sqsubseteq \mathsf{L}_2$") if $\mathcal{L}_1 \subseteq \mathcal{L}_2$ and $\vdash_1 \subseteq \vdash_2\restriction_{\mathcal{L}_1}$.

This definition of a sublogic is consistent with its usage in the literature. For instance, some logicians talk as though intuitionistic logic is a sublogic of classical logic, or as propositional logic as a sublogic of full first-order logic. But from a model-theoretic perspective, the classical propositional language can be viewed as a restriction of the intuitionistic language (to the class of single-pointed Kripke models) or of the first-order language (to the quantifier-free formulas).

Not every restriction is conservative. This is simply because expansions can contain counterexamples to inferences not present in their restrictions. However, in some special circumstances, conservativity can be guaranteed.

Definition 1.1.15 *(Extensions and Fragments)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages. We say that $\mathbf{L}_2$ is an ***extension*** of $\mathbf{L}_1$, or that $\mathbf{L}_1$ is a ***fragment*** of $\mathbf{L}_2$ (written "$\mathbf{L}_1 \subseteq \mathbf{L}_2$"), if $\mathbf{L}_1 \Subset \mathbf{L}_2$ and $\mathsf{C}_1 = \mathsf{C}_2$. We say that $\mathbf{L}_2$ is a ***proper extension*** of $\mathbf{L}_1$, or that $\mathbf{L}_1$ is a ***proper fragment*** of $\mathbf{L}_2$ (written "$\mathbf{L}_1 \subset \mathbf{L}_2$"), if $\mathbf{L}_1 \subseteq \mathbf{L}_2$ but $\mathbf{L}_1 \neq \mathbf{L}_2$. Where $\mathcal{L}_1 \subseteq \mathcal{L}_2$, we write "$\mathbf{L}_2 \restriction_{\mathcal{L}_1}$" for the fragment of $\mathbf{L}_2$ obtained by restricting the syntax to $\mathcal{L}_1$.

Similarly, let L_1 and L_2 be logics. We say that L_1 is a ***fragment*** of L_2 (written "$\mathsf{L}_1 \subseteq \mathsf{L}_2$") if $\mathsf{L}_1 \sqsubseteq \mathsf{L}_2$ and $\vdash_1 = \vdash_2\restriction_{\mathcal{L}_1}$. We denote the fragment of L_2 obtained by restricting the syntax to L_1 as "$\mathsf{L}_2 \restriction_{\mathcal{L}_1}$".

This terminology is again consistent with how many logicians use the term "fragment" and "extension" in the literature. For instance, it is often said that modal logic is (equivalent to) a fragment of first-order logic, viz., the bisimulation-invariant fragment. By "the bisimulation-invariant fragment", they usually mean the language of first-order logic but restricted to the bisimulation-invariant first-order formulas. Similarly, it is often said that classical logic is the fragment of intuitionistic logic with just negation and conjunction.

Fact 1.1.16 *(Fragments are Conservative Restrictions)*. If $\mathbf{L}_1 \subseteq \mathbf{L}_2$, then $\mathbf{L}_1 \Subset \mathbf{L}_2$.

Fact 1.1.17 *(The Logic of a Restriction Contains a Fragment of the Logic)*. If $\mathbf{L}_1 \Subset \mathbf{L}_2$, then $\mathsf{L}_2 \upharpoonright_{\mathcal{L}_1} \sqsubseteq \mathsf{L}_1$.

Fact 1.1.18 *(The Logic of a Conservative Restriction is a Fragment of the Logic)*. If $\mathbf{L}_1 \Subset \mathbf{L}_2$, then $\mathsf{L}_1 = \mathsf{L}_2 \upharpoonright_{\mathcal{L}_1}$.

Put in other terms: **Fact 1.1.17** says that the logic of an expansion is an extension of a sublogic. And **Fact 1.1.18** says that the logic of a preservative expansion is an extension of the logic.

The converse of **Fact 1.1.17** generally does not hold: an extension of a sublogic need not be the logic of an expansion. A simple way to see this is to note that even if $\mathsf{L}_2 \upharpoonright_{\mathcal{L}_1} \sqsubseteq \mathsf{L}_1$, it need not be the case that $\mathcal{L}_2 \supseteq \mathcal{L}_1$, as would be required to be the logic of an expansion of $\mathbf{L}_1$. However, we do get a limited version of the converse of **Fact 1.1.17**: all sublogics are fragments of the logic of an expansion.

Proposition 1.1.19 *(A Sublogic is a Fragment of the Logic of an Expansion)*. Let $\mathbf{L}_1$ be a language and let $\mathsf{L}_2 \sqsubseteq \mathsf{L}_1$. Then there is a $\mathbf{L}_2$ such that $\mathbf{L}_1 \Subset \mathbf{L}_2$ and $\mathsf{L}_2 = \mathsf{L}_{\mathbf{L}_2} \upharpoonright_{\mathcal{L}_2}$.

Proof: Let $\mathtt{Th}(\mathsf{L}_2) = \{\mathtt{Cn}_2(\Gamma) \mid \Gamma \subseteq \mathcal{L}_2\}$ (for the class of "L_2-theories"; this notation will be introduced in § 1.1.3). Without loss of generality, assume that $\mathsf{C} \cap \mathtt{Th}(\mathsf{L}_2) = \varnothing$. Let $\mathbf{L}_2 = \langle \mathcal{L}_1, \mathsf{C} \cup \mathtt{Th}(\mathsf{L}_2), \Vdash_2 \rangle$ where $\Vdash_2 \upharpoonright_{\mathsf{C}} = \Vdash_1$ and where $\mathtt{Cn}_2(\Gamma) \Vdash_2 \phi$ iff $\phi \in \mathtt{Cn}_2(\Gamma)$. Clearly $\mathbf{L}_1 \Subset \mathbf{L}_2$. So we just need to show that $\mathsf{L}_2 = \mathsf{L}_{\mathbf{L}_2} \upharpoonright_{\mathcal{L}_2}$.

Let $\Gamma \cup \{\phi\} \subseteq \mathcal{L}_2$. Suppose $\Gamma \vdash_2 \phi$. Then $\Gamma \vdash_1 \phi$, so for all $x \in \mathsf{C}$, if $x \Vdash_2 \Gamma$, then $x \Vdash_2 \phi$. Moreover, $\phi \in \mathtt{Cn}_2(\Gamma)$, so if $\mathtt{Cn}_2(\Delta) \Vdash_2 \Gamma$, that means $\mathtt{Cn}_2(\Delta) \supseteq \mathtt{Cn}_2(\Gamma) \ni \phi$. Hence, $\Gamma \vDash_{\mathbf{L}_2} \phi$. Conversely, suppose $\Gamma \nvdash_2 \phi$. Then $\mathtt{Cn}_2(\Gamma) \Vdash_2 \Gamma$ even though $\mathtt{Cn}_2(\Gamma) \nVdash_2 \phi$. So $\Gamma \nvDash_{\mathbf{L}_2} \phi$. ∎

The converse of **Fact 1.1.18** also holds. One should take care, however, as there are two things that one could state the converse of **Fact 1.1.18** (one in terms of fragments and the other in terms of extensions). One way would be this: a fragment of the logic of a language **L** is the logic of some conservative restriction of **L**. In fact, we can even show that it is the logic of a fragment of **L**:

Proposition 1.1.20 *(A Fragment of the Logic is the Logic of a Fragment)*. Let $\mathbf{L}_1$ be a language and let $\mathsf{L}_0 \subseteq \mathsf{L}_1$. Then there is a $\mathbf{L}_0 \subseteq \mathbf{L}_1$ such that $\mathsf{L}_0 = \mathsf{L}_{\mathbf{L}_0}$.

Proof: Let $\mathsf{L}_0 \subseteq \mathsf{L}_1$. Let $\mathbf{L}_0 = \mathbf{L}_1 \restriction_{\mathcal{L}_0}$. It is easy to check that $\mathsf{L}_0 = \mathsf{L}_{\mathbf{L}_0}$. ■

Another way to state the converse of **Fact 1.1.18** is this: an extension of the logic is the logic of some preservative expansion. However, we cannot generally take this to be the logic of some extension, since extending the logic may require adding more points of evaluation.

Example 1.1.21 *(An Extension of a Logic that is Not the Logic of Any Extension)*. Let $\mathbf{L}_1 = \langle \{p\}, \{x\}, \Vdash_1 \rangle$ where $x \Vdash_1 p$. Note that if $\mathbf{L} \supseteq \mathbf{L}_1$, then $\vdash_1 \phi$ for any $\phi \in \mathcal{L}_{\mathbf{L}}$. But clearly there are preservative expansions of $\mathbf{L}_1$ that do not validate every formula. For instance, let $\mathcal{L}_2 = \{p, q\}$ and let L_2 be so that $\vdash_2 p$ and $\nvdash_2 q$. Then $\mathsf{L}_2 \supseteq \mathsf{L}_1$.

Proposition 1.1.22 *(An Extension of the Logic is the Logic of a Preservative Expansion)*. Let $\mathbf{L}_1$ be a language and let $\mathsf{L}_2 \supseteq \mathsf{L}_1$. Then there is an $\mathbf{L}_2 \mathrel{\underline{\supset}} \mathbf{L}_1$ such that $\mathsf{L}_2 = \mathsf{L}_{\mathbf{L}_2}$.

Proof: Without loss of generality, assume that $\mathsf{C} \cap \mathtt{Th}(\mathsf{L}_2) = \varnothing$. Let $\mathtt{Diag}_1(x) := \{\phi \in \mathcal{L}_1 \mid x \Vdash_1 \phi\}$ and let $\mathbf{L}_2 = \langle \mathcal{L}_2, \mathsf{C} \cup \mathtt{Th}(\mathsf{L}_2), \Vdash_2 \rangle$ where:

(i) if $x \in \mathsf{C}$ and $\phi \in \mathcal{L}_1$, then $x \Vdash_2 \phi$ iff $x \Vdash_1 \phi$

(ii) if $x \in \mathsf{C}$ and $\phi \in \mathcal{L}_2 - \mathcal{L}_1$, then $x \Vdash_2 \phi$ iff $\mathtt{Diag}_1(x) \vdash_2 \phi$

(iii) if $x \in \mathtt{Th}(\mathsf{L}_2)$ and $\phi \in \mathcal{L}_2$, then $x \Vdash_2 \phi$ iff $\phi \in x$.

Clearly, $\mathbf{L}_1 \Subset \mathbf{L}_2$. We show two things: that $\mathbf{L}_2$ is preservative over $\mathbf{L}_1$ and that $\mathsf{L}_2 = \mathsf{L}_{\mathbf{L}_2}$.

First, let $\Gamma \cup \{\phi\} \subseteq \mathcal{L}_1$. By **Fact 1.1.13**, we just need to show that $\mathbf{L}_2$ preserves $\mathbf{L}_1$-entailments. Suppose $\Gamma \vDash_1 \phi$. Then every $x \in \mathsf{C}$ is such that $x \Vdash_2 \Gamma$ only if $x \Vdash_2 \phi$. So every $x \in \mathsf{C}$ is such that $x \Vdash_2 \Gamma$ only if $x \Vdash_2 \phi$ by (i). In addition, if $x \in \mathtt{Th}(\mathsf{L}_2)$ and $x \Vdash_2 \Gamma$, then $\Gamma \subseteq x$. But since $\mathsf{L}_2 \supseteq \mathsf{L}_1$, $\phi \in x$, so $x \Vdash_2 \phi$. Hence, $\Gamma \vDash_2 \phi$.

Next, let $\Gamma \cup \{\phi\} \subseteq \mathcal{L}_2$. Suppose that $\Gamma \vdash_2 \phi$. Let $x \in \mathsf{C}$ be such that $x \Vdash_2 \Gamma$. By (ii), $\mathtt{Diag}_1(x) \vdash_2 \Gamma$. Hence, $\mathtt{Diag}_1(x) \vdash_2 \phi$. If $\phi \in \mathcal{L}_1$, then $x \Vdash_2 \phi$ by (i). If $\phi \in \mathcal{L}_2 - \mathcal{L}_1$, then $x \Vdash_2 \phi$ by (ii). So either way, $x \Vdash_2 \phi$. Now let $x \in \mathtt{Th}(\mathsf{L}_2)$ be such that $x \Vdash_2 \Gamma$. Thus, $\Gamma \subseteq x$. But since $\Gamma \vdash_2 \phi$, $\phi \in x$, in which case $x \Vdash_2 \phi$ by (iii). Hence, $\Gamma \vDash_{\mathbf{L}_2} \phi$. Conversely, suppose $\Gamma \nvDash_2 \phi$. Then $\phi \notin \mathtt{Cn}_2(\Gamma)$. Hence, $\mathtt{Cn}_2(\Gamma) \nVdash_2 \phi$, and thus $\Gamma \nvDash_{\mathbf{L}_2} \phi$. ■

A particularly important kind of fragment of a language is the kind obtained under the *image* of a mapping from another language.

Definition 1.1.23 *(Image Language)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. The ***t-image language*** with respect to $\mathbf{L}_2$ is defined to be $\mathtt{t}[\mathbf{L}_1] :=$ $\langle \mathtt{t}[\mathcal{L}_1], \mathsf{C}_2, \Vdash_{\mathtt{t}[1]} \rangle$, where $\Vdash_{\mathtt{t}[1]} = \Vdash_2 \restriction_{\mathtt{t}[\mathcal{L}_1]}$. Image logics are likewise defined.

Fact 1.1.24 *(Image Languages are Fragments)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}[\mathbf{L}_1] \subseteq \mathbf{L}_2$.

An alternative way to guarantee conservativity would be to require that every new point in C_2 be "closed" under $\vDash_1$. This can be made more precise using the notion of a diagram.

Definition 1.1.25 *(Diagram)*. Let $\mathbf{L}$ be a language. The ***L-diagram*** of $x \in \mathsf{C}_{\mathbf{L}}$ is defined as $\mathtt{Diag}_{\mathbf{L}}(x) := \{\phi \in \mathcal{L}_{\mathbf{L}} \mid x \Vdash_{\mathbf{L}} \phi\}$. We write "$x \equiv_{\mathbf{L}} y$" to mean "$\mathtt{Diag}_1(x) = \mathtt{Diag}_1(y)$".

Proposition 1.1.26 *(Alternative Characterization of Preservative Expansions)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages where $\mathbf{L}_2$ is an expansion of $\mathbf{L}_1$. Then $\mathbf{L}_1 \trianglelefteq \mathbf{L}_2$ iff for all $y \in \mathsf{C}_2$, $\mathtt{Diag}_2(y) \cap \mathcal{L}_1 = \mathtt{Cn}_1(\mathtt{Diag}_2(y) \cap \mathcal{L}_1)$.

Proof:

(⇒) Suppose that $\mathbf{L}_1 \trianglelefteq \mathbf{L}_2$. Let $y \in \mathsf{C}_2$. Since $\mathtt{Cn}_1$ is a consequence operator, $\mathtt{Diag}_2(y) \cap \mathcal{L}_1 \subseteq \mathtt{Cn}_1(\mathtt{Diag}_2(y) \cap \mathcal{L}_1)$. Conversely, suppose $\phi \in \mathtt{Cn}_1(\mathtt{Diag}_2(y) \cap \mathcal{L}_1)$. That means $\mathtt{Diag}_2(y) \cap \mathcal{L}_1 \vDash_1 \phi$. But since $\mathbf{L}_1 \trianglelefteq \mathbf{L}_2$, $\mathtt{Diag}_2(y) \cap \mathcal{L}_1 \vDash_2 \phi$. And since $y \Vdash_2 \mathtt{Diag}_2(y) \cap \mathcal{L}_1$ by definition, we have $y \Vdash_2 \phi$. So $\phi \in \mathtt{Diag}_2(y) \cap \mathcal{L}_1$.

(⇐) Suppose for all $y \in \mathsf{C}_2$, $\mathtt{Diag}_2(y) \cap \mathcal{L}_1 = \mathtt{Cn}_1(\mathtt{Diag}_2(y) \cap \mathcal{L}_1)$. Let $\Gamma \subseteq \mathcal{L}_1$ and $\phi \in \mathcal{L}_1$. We want to show that $\Gamma \vDash_1 \phi$ iff $\Gamma \vDash_2 \phi$. For the left-to-right direction, suppose $\Gamma \vDash_1 \phi$. Let $y \in \mathsf{C}_2$ where $y \Vdash_2 \Gamma$. Thus, $\Gamma \subseteq \mathtt{Diag}_2(y) \cap \mathcal{L}_1 = \mathtt{Cn}_1(\mathtt{Diag}_2(y) \cap \mathcal{L}_1)$. So $\phi \in \mathtt{Cn}_1(\mathtt{Diag}_2(y) \cap \mathcal{L}_1) =$ $\mathtt{Diag}_2(y) \cap \mathcal{L}_1$. Hence, $y \Vdash_2 \phi$. Since y was arbitrary, $\Gamma \vDash_2 \phi$. For the right-to-left direction, suppose $\Gamma \nvDash_1 \phi$. That means for some $x \in \mathsf{C}_1$,

$x \Vdash_1 \Gamma$ but $x \nVdash_1 \phi$. Now, since $\Vdash_1 = \Vdash_2 \cap (\mathsf{C}_1 \times \mathcal{L}_1)$, $x \Vdash_2 \Gamma$ and $x \nVdash_2 \phi$. Hence, $\Gamma \nvDash_2 \phi$. ■

§ 1.1.3 Theories

While we are primarily focused on *language*, we could use this framework as a way of modeling *theories* more generally. Often, theories are defined to be sets of sentences. But this way of enumerating theories is well-known to be problematic. On the one hand, it seems plausible that two sets of sentences can be used to represent the same theory. On the other hand, one and the same set of sentences can be used to represent different theories depending on how those sentences are interpreted. We could avoid both of these problems by thinking of a theory as a kind of language that is obtained by restricting the evaluation space via some class of sentences.

Definition 1.1.27 *(Theory)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash \rangle$ be a language. An ***L*-*theory*** is a $\Gamma \subseteq \mathcal{L}$ such that $\Gamma = \mathtt{Cn}_{\mathbf{L}}(\Gamma)$. We let $\mathtt{Th}(\mathbf{L})$ be the class of all $\mathbf{L}$-theories. Similarly, where $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ is a logic, an ***L*-*theory*** is a $\Gamma \subseteq \mathcal{L}$ such that $\Gamma = \mathtt{Cn}_{\mathsf{L}}(\Gamma)$. We let $\mathtt{Th}(\mathsf{L})$ be the class of all L-theories.

Fact 1.1.28 *(The Theories of a Language are the Theories of its Logic)*. Let $\mathbf{L}$ be a language. Then $\mathtt{Th}(\mathbf{L}) = \mathtt{Th}(\mathsf{L}_{\mathbf{L}})$.

Fact 1.1.29 *(Intersections of Theories are Theories)*. Let L be a logic and let $\Sigma \subseteq \mathtt{Th}(\mathsf{L})$. Then $\bigcap \Sigma \in \mathtt{Th}(\mathsf{L})$.

Thus, two sets of sentences can represent the same theory in a language as long as they are semantically equivalent; and a single set of sentences can represent multiple theories relative to different languages.

Theories induce restrictions on languages. To adopt a theory is, in a sense, to restrict one's language so that the statements of the theory are valid. This idea is formalized by the notion of a reduction, which will be useful in later chapters.

Definition 1.1.30 *(Reduction)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash \rangle$ be a language and let $\Gamma \subseteq \mathcal{L}$. The **Γ-*reduction*** of $\mathbf{L}$ is the language $\mathbf{L}_\Gamma = \langle \mathcal{L}, [\![\Gamma]\!], \Vdash\restriction_{[\![\Gamma]\!] \times \mathcal{L}} \rangle$. We write "$\Vdash_\Gamma$" in place of "$\Vdash_{\mathbf{L}_\Gamma}$" (likewise for the other parameters of a language). Similarly, the **Γ-*reduction*** of L is the logic $\mathsf{L}_\Gamma = \langle \mathcal{L}, \vdash_\Gamma \rangle$ where $\Delta \vdash_\Gamma \phi$ iff $\Gamma, \Delta \vdash \phi$.

Fact 1.1.31 *(The Logic of a Reduction is the Reduction of the Logic)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash \rangle$ be a language and let $\Gamma \subseteq \mathcal{L}$. Then $\mathsf{L}_{\mathbf{L}_\Gamma} = \mathsf{L}_\Gamma$.

Reductions are kinds of restrictions to a language, whereby we reduce the space of possibilities to those left open by the theory. Except in the trivial case, such reductions will not be conservative restrictions of the language.

Fact 1.1.32 *(Reductions are Almost Never Conservative Restrictions)*. Let $\mathbf{L}$ be a language and let $\Gamma \subseteq \mathcal{L}$. Then $\mathbf{L}_\Gamma \trianglelefteq \mathbf{L}$ iff $\mathbf{L}_\Gamma = \mathbf{L}$.

This is just to say that a conservative theory is uninformative: one only genuinely reduces logical space by being non-conservative with respect to what is valid in one's original language.

The class of theories relative to a single language has a nice well-behaved structure when ordered by entailment.

Definition 1.1.33 *(Theory Space)*. Let $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ be a logic. The ***theory space*** of L is the pair $\mathbb{T}_\mathsf{L} := \langle \mathtt{Th}(\mathsf{L}), \vdash \rangle$. Similarly, let $\mathbf{L}$ be a language. The ***theory space*** of $\mathbf{L}$ is defined as $\mathbb{T}_\mathbf{L} := \mathbb{T}_{\mathsf{L}_\mathbf{L}}$.

Proposition 1.1.34 *(Structure of Theory Spaces)*. Let L be a logic. Then $\mathbb{T}_\mathsf{L}$ is a complete lattice.

Proof: Where $\Sigma \subseteq \mathtt{Th}(\mathsf{L})$, define:

- $\bigwedge \Sigma := \mathtt{Cn}_\mathsf{L}(\bigcup \Sigma)$
- $\bigvee \Sigma := \bigcap \Sigma$.

It is straightforward to verify that $\bigwedge \Sigma$ and $\bigvee \Sigma$ are greatest lower bound and least upper bound of Σ respectively. ■

Notation: Let $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ be a logic. Where $\mathbb{T}_\mathsf{L}$ is a theory space for L and where $\Sigma \subseteq \mathtt{Th}(\mathsf{L})$, the greatest lower bound and least upper bound of Σ in $\mathbb{T}_\mathsf{L}$, respectively, are defined as follows:

$$\bigwedge \Sigma := \mathtt{Cn}_\mathsf{L}\Big(\bigcup \Sigma\Big)$$

$$\bigvee \Sigma \coloneqq \bigcap \Sigma.$$

Every complete lattice is bounded. In the case of theory spaces, the top element is $\mathtt{Cn}_{\mathsf{L}}(\varnothing)$ while the bottom element is $\mathcal{L}$. If we reversed the direction of the order, of course, $\bigwedge$ would correspond to intersection and $\bigvee$ would correspond to the closure of the union.

One may wonder whether we can say more about the algebraic structure of a theory space in general. The answer is negative:

Proposition 1.1.35 *(Every Complete Lattice is a Theory Space).* Let $\langle L, \leqslant \rangle$ be a complete lattice. Then there is a logic L such that $\mathbb{T}_{\mathsf{L}}$ is lattice-isomorphic to $\langle L, \leqslant \rangle$.

Proof: Define $\mathsf{L} \coloneqq \langle L, \vdash \rangle$, where for all $\Gamma \subseteq L$ and $\phi \in L$, $\Gamma \vdash \phi$ iff $\bigwedge \Gamma \leqslant \phi$. (Note that L is Tarskian.) Let $f : L \to \mathtt{Th}(\mathsf{L})$ be such that $f(\phi) = \mathtt{Cn}_{\mathsf{L}}(\phi)$. We show that f is a lattice-isomorphism.

First, bijectivity. f is injective: If $f(\phi) = f(\psi)$, then $\phi \dashv\vdash \psi$. By definition, $\phi \leqslant \psi \leqslant \phi$, so $\phi = \psi$. f is surjective: Let $\Gamma \in \mathtt{Th}(\mathsf{L})$. Since $\langle L, \leqslant \rangle$ is complete, $\bigwedge \Gamma$ is defined. Hence, $\Gamma \dashv\vdash \bigwedge \Gamma$, and so $\Gamma = \mathtt{Cn}_{\mathsf{L}}(\bigwedge \Gamma) = f(\bigwedge \Gamma)$.

Now we show f is a homomorphism. Let $\Gamma \subseteq L$. Then:

$$\begin{aligned} f(\bigwedge \Gamma) &= \mathtt{Cn}_{\mathsf{L}}(\bigwedge \Gamma) \\ &= \mathtt{Cn}_{\mathsf{L}}(\Gamma) \\ &= \mathtt{Cn}_{\mathsf{L}}(\bigcup \{\mathtt{Cn}_{\mathsf{L}}(\phi) \mid \phi \in \Gamma\}) \\ &= \bigwedge \{\mathtt{Cn}_{\mathsf{L}}(\phi) \mid \phi \in \Gamma\} \\ &= \bigwedge f[\Gamma]. \end{aligned}$$

Moreover:

$$\begin{aligned} f(\bigvee \Gamma) &= \mathtt{Cn}_{\mathsf{L}}(\bigvee \Gamma) \\ &= \{\phi \in L \mid \bigvee \Gamma \vdash \phi\} \end{aligned}$$

$$
\begin{aligned}
&= \{\phi \in L \mid \bigvee \Gamma \leqslant \phi\} \\
&= \{\phi \in L \mid \forall \psi \in \Gamma\colon\ \psi \leqslant \phi\} \\
&= \bigcap_{\psi \in \Gamma} \{\phi \in L \mid \psi \leqslant \phi\} \\
&= \bigcap_{\psi \in \Gamma} \mathtt{Cn}_{\mathsf{L}}(\psi) \\
&= \bigvee \{\mathtt{Cn}_{\mathsf{L}}(\psi) \mid \psi \in \Gamma\} \\
&= \bigvee f[\Gamma].
\end{aligned}
$$

This completes the proof. ■

The proof of **Proposition 1.1.35** shows a stronger result: every complete lattice is isomorphic to the theory space of some *completely conjunctive* logic (in the sense of **Definition 1.2.24** below).

We can use the lattice-theoretic properties of a theory space to determine what the conservative restrictions of a language are.

Proposition 1.1.36 *(Expansions and Theory Spaces)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages such that $\mathbf{L}_1 \Subset \mathbf{L}_2$. Then $\mathbf{L}_1 \trianglelefteq \mathbf{L}_2$ iff $\mathbb{T}_1$ is (meet-)semilattice-embeddable in $\mathbb{T}_2$ via the map $f\colon \mathtt{Cn}_1(\Gamma) \mapsto \mathtt{Cn}_2(\Gamma)$.

Proof:

(⇒) Suppose $\mathbf{L}_1 \trianglelefteq \mathbf{L}_2$. f is a well-defined injective function: $\mathtt{Cn}_1(\Gamma) = \mathtt{Cn}_1(\Delta)$ iff $\Gamma \equiv_1 \Delta$ iff $\Gamma \equiv_2 \Delta$ iff $\mathtt{Cn}_2(\Gamma) = \mathtt{Cn}_2(\Delta)$ iff $f(\mathtt{Cn}_1(\Gamma)) = f(\mathtt{Cn}_1(\Delta))$. Now we show it is a semilattice-embedding. Let $\Sigma \subseteq \mathtt{Th}(\mathbf{L}_1)$. Then $f(\bigwedge_{\Gamma\in\Sigma} \mathtt{Cn}_1(\Gamma)) = f(\mathtt{Cn}_1(\bigcup \Sigma)) = \mathtt{Cn}_2(\bigcup \Sigma) = \bigwedge f[\Sigma]$.

(⇐) Suppose f is a semilattice-embedding of $\mathbb{T}(\mathbf{L}_1)$ into $\mathbb{T}(\mathbf{L}_2)$. Then $\Gamma \vDash_1 \phi$ iff $\mathtt{Cn}_1(\Gamma) = \mathtt{Cn}_1(\Gamma \cup \{\phi\}) = \mathtt{Cn}_1(\mathtt{Cn}_1(\Gamma) \cup \mathtt{Cn}_1(\phi)) = \mathtt{Cn}_1(\Gamma) \wedge \mathtt{Cn}_1(\phi)$ iff $\mathtt{Cn}_2(\Gamma) = f(\mathtt{Cn}_1(\Gamma)) = f(\mathtt{Cn}_1(\Gamma) \wedge \mathtt{Cn}_1(\phi)) = f(\mathtt{Cn}_1(\Gamma)) \wedge f(\mathtt{Cn}_1(\phi)) = \mathtt{Cn}_2(\Gamma) \wedge \mathtt{Cn}_2(\phi) = \mathtt{Cn}_2(\mathtt{Cn}_2(\Gamma) \cup \mathtt{Cn}_2(\phi)) = \mathtt{Cn}_2(\Gamma \cup \{\phi\})$ iff $\Gamma \vDash_2 \phi$. ■

Note that $\mathbb{T}(\mathbf{L}_1)$ need not be a full sublattice of $\mathbb{T}(\mathbf{L}_2)$ via f if $\mathbf{L}_1 \trianglelefteq \mathbf{L}_2$. In particular, the image of a join under f need not be the join of the images under f.

Example 1.1.37 *(The Theory Space of a Conservative Restriction Need not be a Sublattice)*. Define $\mathbf{L}_1$ and $\mathbf{L}_2$ as follows:

- $\mathcal{L}_1 = \{p, q\}$ and $\mathcal{L}_2 = \{p, q, r\}$
- $\mathsf{C}_1 = \{p, q\}$ and $\mathsf{C}_2 = \{p, q, \top\}$
- for all $x \in \mathsf{C}_1$ and $\phi \in \mathcal{L}_1$, $x \Vdash_1 \phi$ iff $x = \phi$
- for all $x \in \mathsf{C}_2$ and $\phi \in \mathcal{L}_2$, $x \Vdash_2 \phi$ iff $x \neq \top$ and either $\phi = x$ or $\phi = r$.

Notice that $\mathtt{Cn}_1(\Gamma) = \Gamma$ for $\Gamma \subseteq \mathcal{L}_1$, that $\mathtt{Cn}_2(\varnothing) = \varnothing$, and that $\mathtt{Cn}_2(\Gamma) = \Gamma \cup \{r\}$ for $\varnothing \neq \Gamma \subseteq \mathcal{L}_2$. Hence, $\mathbf{L}_1 \ntrianglelefteq \mathbf{L}_2$. $\mathtt{Th}(\mathbf{L}_1)$ and $\mathtt{Th}(\mathbf{L}_2)$ are represented diagrammatically below:

$$\begin{tikzcd}
& \varnothing & \\
\{p\} \arrow[ur] & & \{q\} \arrow[ul] \\
& \{p, q\} \arrow[ul] \arrow[ur] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& \varnothing & \\
& \{r\} \arrow[u] & \\
\{p, r\} \arrow[ur] & & \{q, r\} \arrow[ul] \\
& \{p, q, r\} \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

$\mathtt{Th}(\mathbf{L}_1)$ $\qquad$ $\mathtt{Th}(\mathbf{L}_2)$

Thus, $f(\{p\} \vee \{q\}) = f(\varnothing) = \varnothing$, while $f(\{p\}) \vee f(\{q\}) = \{p, r\} \vee \{q, r\} = \{r\}$. So f does not preserve joins.

An important substructure of the theory space of a logic is its Lindenbaum-Tarski algebra. This substructure encodes the entailments between formulas-up-to-equivalence (Lindenbaum-Tarski algebras will play a central role in § 2.6.1).

Definition 1.1.38 *(Lindenbaum-Tarski Algebra)*. Let L be a logic. We define the ***Lindenbaum-Tarski algebra*** of L to be the pair $\mathbb{L}_{\mathsf{L}} := \langle \mathcal{L}/\dashv\vdash, \leqslant \rangle$, where:

- $[\phi]_{\mathsf{L}} := \{\psi \in \mathcal{L} \mid \phi \dashv\vdash \psi\}$
- $\mathcal{L}/\dashv\vdash \; := \{[\phi]_{\mathsf{L}} \mid \phi \in \mathcal{L}\}$
- $[\phi]_{\mathsf{L}} \leqslant [\psi]_{\mathsf{L}}$ iff $\phi \vdash \psi$.

The ***Lindenbaum-Tarski algebra*** of a language $\mathbf{L}$ is defined as $\mathbb{L}_{\mathbf{L}} := \mathbb{L}_{\mathsf{L}_{\mathbf{L}}}$.

Fact 1.1.39 *(Lindenbaum-Tarski Algebras are Partial Orders)*. $\mathbb{L}_{\mathsf{L}}$ is a partial order for any logic **L**.

Unlike theory spaces, Lindenbaum-Tarski algebras can have significantly less structure than that of a lattice.

Proposition 1.1.40 *(Every Partial Order is a Lindenbaum-Tarski Algebra)*. Let $\mathbb{P} = \langle P, \leqslant \rangle$ be a partial order. Then there is a logic L such that $\mathbb{L}_{\mathsf{L}}$ is order-isomorphic to $\mathbb{P}$.

Proof: Define $\mathsf{L} := \langle P, \vdash \rangle$, where for all $\Gamma \subseteq P$ and $\phi \in P$, $\Gamma \vdash \phi$ iff there is a $\psi \in \Gamma$ such that $\psi \leqslant \phi$. (Note that L is Tarskian.) Then it is easy check that the map $f : \phi \mapsto \{\phi\}$ is an isomorphism from $\mathbb{P}$ to $\mathbb{L}_{\mathsf{L}}$. ∎

We can think of Lindenbaum-Tarski algebras as substructures of a theory space, viz., the ones consisting of only the *principal* theories.

Definition 1.1.41 *(Principal Theory)*. Let L be a logic. An L-theory $\Gamma \in \mathtt{Th}(\mathsf{L})$ is ***principal*** if there is a $\phi \in \mathcal{L}$ such that $\Gamma = \mathtt{Cn}_{\mathsf{L}}(\phi)$. We let $\mathtt{PTh}(\mathsf{L})$ be the class of all principal L-theories. The ***principal substructure*** of the theory space $\mathtt{Th}(\mathsf{L})$ of L is the pair $\langle \mathtt{PTh}(\mathsf{L}), \vdash \rangle$.

Proposition 1.1.42 *(Lindenbaum-Tarski Algebras are Isomorphic to the Principal Substructure of Theory Spaces)*. Let L be a logic. Then $\mathbb{L}_{\mathsf{L}}$ is isomorphic to the principal substructure of $\mathbb{T}_{\mathsf{L}}$.

Proof: It is easily verified that $f : [\phi]_{\mathsf{L}} \mapsto \mathtt{Cn}_{\mathsf{L}}(\phi)$ is an order-isomorphism from $\mathbb{L}_{\mathsf{L}}$ to $\langle \mathtt{PTh}(\mathsf{L}), \vdash \rangle$. ∎

We briefly introduce one more concept that will be helpful later on. Recall that **Proposition 1.1.9** states that a logic is Tarskian just in case it is the logic of some language with the same syntax. In showing the implication from (c) to (d) of **Proposition 1.1.9**, we appealed to a structure that essentially used the class of theories as the evaluation space. This will be a useful construction later on:

Definition 1.1.43 *(Canonical Languages)*. Let $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ be a logic. We define the ***canonical language*** for L to be the language $\mathbf{Can}(\mathsf{L}) = \langle \mathcal{L}, \mathtt{Th}(\mathsf{L}), \ni \rangle$—that is, $\Gamma \Vdash_{\mathbf{Can}(\mathsf{L})} \phi$ iff $\phi \in \Gamma$.

The canonical language is, in a sense, the most compact language a logic can have. For one thing, the logic of the canonical language of L is just L.

Fact 1.1.44 *(Logics are the Logics of Their Canonical Language).* Let $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$. Then $\mathsf{L}_{\mathbf{Can}(\mathsf{L})} = \mathsf{L}$.

Moreover, any language whose logic is L is isomorphic (in the sense of **Definition 2.3.1**, introduced in the next chapter) to a preservative expansion of $\mathbf{Can}(\mathsf{L})$.

Fact 1.1.45 *(Canonical Languages are Isomorphic to Conservative Restrictions).* Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash \rangle$ be a language. Then there is a language $\mathbf{L}^* = \langle \mathcal{L}, \mathsf{C}^*, \Vdash^* \rangle$ such that:

(i) $\mathbf{Can}(\mathsf{L}) \trianglelefteq \mathbf{L}^*$

(ii) there is a bijective map $f : \mathsf{C} \to \mathsf{C}^*$ where for all $x \in \mathsf{C}$ and all $\phi \in \mathcal{L}$, $x \Vdash \phi$ iff $f(x) \Vdash^* \phi$.

This is easy to establish, since each point $x \in \mathsf{C}$ can be replaced with $\mathtt{Cn}_{\mathsf{L}}(\mathtt{Diag}_{\mathbf{L}}(x))$ without loss.

§ 1.2 Properties of Languages

In this section, we review some properties of a language that are of philosophical and theoretical interest. We will refer back to these properties throughout the rest of the dissertation, but for the purposes of this chapter, this section may be skimmed or skipped and returned to when necessary.

§ 1.2.1 Opinionated Languages

In defining different measures of expressive power, it turns out we can often simplify the measure when our languages have the following special property:

Definition 1.2.1 *(Maximal States and Opinionated Languages).* A ***maximal state*** in a language $\mathbf{L}$ is an $x \in \mathsf{C}_{\mathbf{L}}$ such that for no $x' \in \mathsf{C}_{\mathbf{L}}$ is $\mathtt{Diag}_{\mathbf{L}}(x) \subset \mathtt{Diag}_{\mathbf{L}}(x')$. A language $\mathbf{L}$ is ***opinionated*** if all of its states are maximal.

Opinionated languages are those in which the points of evaluation are "maximally opinionated". In an opinionated language, each state can be thought of as a maximal consistent subclass of formulas. We now characterize exactly when a logic is the logic of an opinionated language. First, a definition:

Definition 1.2.2 *(Absurd State)*. An ***absurd state*** in a language $\mathbf{L}$ is an $x \in \mathsf{C}_\mathbf{L}$ such that $x \Vdash_\mathbf{L} \phi$ for all $\phi \in \mathcal{L}_\mathbf{L}$.

Absurd states never affect a consequence relation, since they are never counterexamples to an inference.

Fact 1.2.3 *(Absurd States Never Affect Entailment)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash \rangle$ and let $\mathsf{C}^+ \supseteq \mathsf{C}$. Let $\mathbf{L}^+ = \langle \mathcal{L}, \mathsf{C}^+, \Vdash^+ \rangle$ be the expansion of $\mathbf{L}$ such that if $x \in \mathsf{C}^+ - \mathsf{C}$, then $x \Vdash^+ \mathcal{L}$. Then $\mathbf{L} \stackrel{\subseteq}{=} \mathbf{L}^+$.

A simple but important observation about absurd states is this: in non-trivial opinionated languages, absurd states are not possible.

Lemma 1.2.4 *(Opinionated Languages and Absurd States)*. If $\mathbf{L}$ is opinionated, then either every $x \in \mathsf{C}_\mathbf{L}$ is an absurd state in $\mathbf{L}$ or none is.

Proof: Let $\mathbf{L}$ be opinionated and let $x \in \mathsf{C}_\mathbf{L}$ be absurd in $\mathbf{L}$. Then for any $x' \in \mathsf{C}_\mathbf{L}$, $\mathtt{Diag}_\mathbf{L}(x') \not\subset \mathtt{Diag}_\mathbf{L}(x)$. But $\mathtt{Diag}_\mathbf{L}(x') \subseteq \mathcal{L}_\mathbf{L} = \mathtt{Diag}_\mathbf{L}(x)$, so $\mathtt{Diag}_\mathbf{L}(x') = \mathcal{L}_\mathbf{L}$, i.e., x' is absurd too. ∎

The opposite of an absurd state is an "empty" state, i.e., one that satisfies no formula:

Definition 1.2.5 *(Empty State)*. An ***empty state*** in a language $\mathbf{L}$ is an $x \in \mathsf{C}_\mathbf{L}$ such that $x \nVdash_\mathbf{L} \phi$ for all $\phi \in \mathcal{L}_\mathbf{L}$.

Unlike absurd states, empty states can affect the logic of a language, albeit in only one respect, viz., they eliminate validities. But like absurd states, only trivial opinionated languages have them:

Fact 1.2.6 *(Opinionated Logics and Empty States)*. If $\mathbf{L}$ is opinionated, then either every $x \in \mathsf{C}_\mathbf{L}$ is an empty state in $\mathbf{L}$ or none is.

Classical languages are generally opinionated. Nonclassical languages (e.g., **IPL**, **K3**, **LP**, etc.) are generally not opinionated. In terms of logics, almost any logic can be viewed as the logic of some non-opinionated language, while only very special logics can be viewed as the logic of an opinionated language. To clarify, we introduce the following definition:

Definition 1.2.7 *(Symmetric Logic)*. A logic $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ is ***symmetric*** if its consequence relation is symmetric—that is, for all $\Gamma, \Delta \subseteq \mathcal{L}$, if $\Gamma \vdash \Delta$, then $\Delta \vdash \Gamma$.

Proposition 1.2.8 *(Almost All Logics are the Logic of a Non-Opinionated Language)*. Let $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ be a logic. Then L is the logic of a non-opinionated language iff L is not symmetric.

Proof: Suppose $\mathsf{L} = \mathsf{L}_{\mathbf{L}}$ where $\mathbf{L}$ is not opinionated. Let $\mathtt{Diag}_{\mathbf{L}}(x) \subset \mathtt{Diag}_{\mathbf{L}}(y)$. Then $\mathtt{Diag}_{\mathbf{L}}(x) \nvdash \mathtt{Diag}_{\mathbf{L}}(y)$, even though clearly $\mathtt{Diag}_{\mathbf{L}}(y) \vdash \mathtt{Diag}_{\mathbf{L}}(x)$. So L is not symmetric. Conversely, suppose L is not symmetric. Consider the canonical language $\mathbf{Can}(\mathsf{L})$ of L. Let $\Gamma \vdash \Delta$ while $\Delta \nvdash \Gamma$. Then $\mathtt{Cn}(\Delta) \subset \mathtt{Cn}(\Gamma)$, in which case $\mathtt{Diag}_{\mathbf{Can}(\mathsf{L})}(\mathtt{Cn}(\Delta)) \subset \mathtt{Diag}_{\mathbf{Can}(\mathsf{L})}(\mathtt{Cn}(\Gamma))$. Therefore, $\mathbf{Can}(\mathsf{L})$ is not opinionated. ■

Thus, almost any interesting logic will be the logic of some non-opinionated language. This applies even to classical languages.[4] On the other hand, as we now show, only very special logics are the logics of an opinionated language.

Definition 1.2.9 *(Consistent, Maximal, and Maximal Consistent Theories)*. Let $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ be a logic. Where $\Gamma \in \mathtt{Th}(\mathsf{L})$, we say Γ is ***L-consistent*** if $\Gamma \nvdash \mathcal{L}$; otherwise, we say Γ is ***L-inconsistent***. We say L is ***trivial*** if $\varnothing$ is L-inconsistent. Γ is ***L-maximal*** if for all $\Delta \in \mathtt{Th}(\mathsf{L})$, if $\Gamma \subset \Delta$, then $\Delta = \mathcal{L}$ (i.e., Δ is L-inconsistent). Γ is ***L-maximally consistent*** if Γ is L-maximal and L-consistent. Where L is a logic, we let $\mathtt{Con}(\mathsf{L})$ be the class of L-consistent theories, $\mathtt{Max}(\mathsf{L})$ be the class of L-maximal theories, and $\mathtt{MaxCon}(\mathsf{L})$ be the class of L-maximally consistent theories. These notions lift to the corresponding notions for languages in the obvious way.

In short, the result we show below roughly states that the logics that are the logic of some opinionated language are exactly those where every consistent theory can be extended to a maximally consistent theory. This is reminiscent of the Lindenbaum construction used to prove the completeness of classical logic.

Definition 1.2.10 *(Lindenbaum Logic)*. A logic $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ is ***Lindenbaum*** if for all $\Gamma \in \mathtt{Th}(\mathsf{L})$ and $\phi \in \mathcal{L}$ such that $\phi \notin \Gamma$, there is a L-maximally consistent theory $\Delta \supseteq \Gamma$ such that $\phi \notin \Delta$.

[4] See, e.g., Humberstone [1981]; Holliday [2014, 2018] for work on possibility semantics (a non-opinionated language for classical propositional modal logic).

Proposition 1.2.11 *(Only Lindenbaum Logics are the Logic of an Opinionated Language)*. A logic L is the logic of an opinionated language iff it is Lindenbaum.

Proof:

($\Rightarrow$) Suppose $\mathsf{L} = \mathsf{L}_{\mathbf{L}}$ where $\mathbf{L}$ is opinionated. If $\mathbf{L}$ contains absurd states, then $\vDash_{\mathbf{L}} \mathcal{L}$, and so $\mathtt{Th}(\mathsf{L}) = \mathtt{Th}(\mathbf{L}) = \{\mathcal{L}\}$, i.e., $\mathtt{Con}(\mathsf{L}) = \varnothing$. In that case, L is trivially Lindenbaum. So suppose $\mathbf{L}$ contains no absurd states. Let $\Gamma \in \mathtt{Th}(\mathsf{L})$ and $\phi \notin \Gamma$. Then $\Gamma \nvdash \phi$, since Γ is a theory. So there is a $x \in \mathsf{C}_{\mathbf{L}}$ such that $\Gamma \subseteq \mathtt{Diag}_{\mathbf{L}}(x)$ but $\phi \notin \mathtt{Diag}_{\mathbf{L}}(x)$. Since $\mathbf{L}$ has no absurd states, $\mathtt{Diag}_{\mathbf{L}}(x)$ is L-consistent. Let $\Delta \in \mathtt{Th}(\mathbf{L})$ be such that $\mathtt{Diag}_{\mathbf{L}}(x) \subset \Delta$. If $\psi \notin \Delta$, then $\Delta \nvdash \psi$, so there is a $y \in \mathsf{C}_{\mathbf{L}}$ such that $\Delta \subseteq \mathtt{Diag}_{\mathbf{L}}(y)$ but $\psi \notin \mathtt{Diag}_{\mathbf{L}}(y)$. But then $\mathtt{Diag}_{\mathbf{L}}(x) \subset \Delta \subseteq \mathtt{Diag}_{\mathbf{L}}(y)$, $\lightning$. Hence, $\Delta = \mathcal{L}$, i.e., $\mathtt{Diag}_{\mathbf{L}}(x)$ is L-maximal.

($\Leftarrow$) If $\mathtt{Con}(\mathsf{L}) = \varnothing$, let $\mathbf{L} = \langle \mathcal{L}, \{x\}, \Vdash \rangle$ where $x \Vdash \mathcal{L}$. Then trivially $\mathbf{L}$ is opinionated and also $\mathsf{L}_{\mathbf{L}} = \mathsf{L}$. So suppose $\mathtt{Con}(\mathsf{L}) \neq \varnothing$. Let $\mathbf{L} = \langle \mathcal{L}, \mathtt{MaxCon}(\mathsf{L}), \ni \rangle$. Clearly, $\mathbf{L}$ is opinionated. We now show that $\mathsf{L}_{\mathbf{L}} = \mathsf{L}$.

Suppose $\Gamma \vdash \phi$. If Γ is L-inconsistent, then no $\Delta \in \mathtt{MaxCon}(\mathsf{L})$ contains Γ. Thus, vacuously $\Gamma \vDash_{\mathbf{L}} \phi$. If Γ is L-consistent, then for any $\Delta \in \mathtt{MaxCon}(\mathsf{L})$ such that $\Gamma \subseteq \Delta$, $\phi \in \mathtt{Cn}_{\mathsf{L}}(\Gamma) \subseteq \Delta$. Hence, $\Gamma \vDash_{\mathbf{L}} \phi$. Suppose instead that $\Gamma \nvdash \phi$. Then $\phi \notin \mathtt{Cn}_{\mathsf{L}}(\Gamma)$. By supposition, there is a $\Delta \in \mathtt{MaxCon}(\mathsf{L})$ such that $\Delta \supseteq \mathtt{Cn}_{\mathsf{L}}(\Gamma)$ and $\phi \notin \Delta$. Hence, $\Gamma \nvDash_{\mathbf{L}} \phi$. ■

§ 1.2.2 Compactness

Recall that first-order predicate logic is *compact*, meaning that if a set of formulas is finitely satisfiable, then so is the whole set. It is a standard exercise to show how, for first-order logic, this is equivalent to the claim that if a set of formulas Γ implies some formula ϕ, then already some finite subset of Γ implies ϕ. But in a more general setting, these two notions can come apart.

Definition 1.2.12 *(Compactness)*. A logic L is ***compact*** if whenever $\Gamma \vdash \phi$, there is some finite $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vdash \phi$. A language $\mathbf{L}$ is ***compact*** if $\mathsf{L}_{\mathbf{L}}$ is.

Definition 1.2.13 *((Finite) Satisfiability)*. Let $\mathbf{L}$ be a language and let $\Gamma \subseteq \mathcal{L}$. Γ is ***L-satisfiable*** if there is a non-absurd $x \in \mathsf{C}$ such that $x \Vdash \Gamma$. Otherwise, Γ is ***L-unsatisfiable***. Γ is ***finitely L-satisfiable*** if for each finite $\Gamma_0 \subseteq \Gamma$, there is a non-absurd $x \in \mathsf{C}$ such that $x \Vdash \Gamma_0$. Otherwise, Γ is ***finitely L-unsatisfiable***.

Definition 1.2.14 *(Finitarity)*. A language **L** is ***finitary*** if for all $\Gamma \subseteq \mathcal{L}$, if Γ is finitely **L**-satisfiable, then Γ is **L**-satisfiable.

Proposition 1.2.15 *(Compactness Implies Finitarity)*. If **L** is compact, then it is finitary.

Proof: Suppose Γ is not **L**-satisfiable. Thus, $\Gamma \vDash \mathcal{L}$. Hence, by compactness, there is a finite $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vDash \mathcal{L}$, in which case Γ_0 is not **L**-satisfiable. So Γ is not finitely **L**-satisfiable. ■

The converse does not hold, however.

Example 1.2.16 *(Finitarity Without Compactness)*. Define **L** as follows:

- $\mathcal{L} = \{p_1, p_2, p_3, \ldots, q, r\}$
- $\mathsf{C} = \{X \mid X \subseteq \{p_1, p_2, p_3, \ldots\} \text{ and } |X| < \aleph_0\} \cup \{\{q, p_1, p_2, p_3, \ldots\}\}$
- $\Vdash = \ni$.

If $\Gamma \subseteq \mathcal{L}$ is finitely **L**-satisfiable, that means $r \notin \Gamma$, so $\{q, p_1, p_2, p_3, \ldots\} \Vdash \Gamma$. Hence, **L** is finitary. But $p_1, p_2, p_3, \ldots \vDash_{\mathbf{L}} q$, while for no $i_1, \ldots, i_m$ does $p_{i_1}, \ldots, p_{i_m} \vDash_{\mathbf{L}} q$.

The reason that these two formulations of compactness are equivalent in the classical setting is that classical languages have negation. In the standard proof that finitarity implies compactness, one moves from the fact that $\Gamma \vDash \phi$ to the conclusion that $\Gamma \cup \{\neg \phi\}$ is not satisfiable, which is only possible if each formula of the language has a negation. We can simulate this idea by generalizing the notion of (finite) satisfiability as follows:

Definition 1.2.17 *((Finite) Satisfiability with Omission)*. Let **L** be a language and let $\Gamma, \Delta \subseteq \mathcal{L}$. Γ is ***L-satisfiable while omitting*** Δ if there is a non-absurd $x \in \mathsf{C}$ such that $x \Vdash \Gamma$ while $x \nVdash \phi$ for all $\phi \in \Delta$. Γ is ***finitely L-satisfiable while omitting*** Δ if for each finite $\Gamma_0 \subseteq \Gamma$, there is a non-absurd $x \in \mathsf{C}$ such that $x \Vdash \Gamma_0$ while $x \nVdash \phi$ for all $\phi \in \Delta$. When $\Delta = \{\phi_1, \ldots, \phi_n\}$ is finite, we may drop set brackets around $\phi_1, \ldots, \phi_n$ and say "omitting $\phi_1, \ldots, \phi_n$" rather than "omitting $\{\phi_1, \ldots, \phi_n\}$".

Thus, to say Γ is (finitely) **L**-satisfiable is to say Γ is (finitely) **L**-satisfiable while omitting $\varnothing$.

Fact 1.2.18 *(Relation Between Consequence and Unsatisfiability)*. Let $\mathbf{L}$ be a language and let $\Gamma \subseteq \mathcal{L}$ and $\phi \in \mathcal{L}$. Then $\Gamma \vDash \phi$ iff Γ is not $\mathbf{L}$-satisfiable while omitting ϕ.

Corollary 1.2.19 *(Alternative Formulation of Compactness)*. Let $\mathbf{L}$ be a language. Then $\mathbf{L}$ is compact iff for all $\Gamma \subseteq \mathcal{L}$ and $\phi \in \mathcal{L}$, if Γ is finitely $\mathbf{L}$-satisfiable while omitting ϕ, then Γ is $\mathbf{L}$-satisfiable while omitting ϕ.

One might ask whether there is anything we can infer about the structure of a Lindenbaum-Tarski algebra if L is compact. The answer is negative. Recall that every partial order is isomorphic to the Lindenbaum-Tarski algebra of some logic (**Proposition 1.1.40**). That construction yielded not just any logic, but a compact logic. Hence:

Corollary 1.2.20 *(Every Partial Order is a Lindenbaum-Tarski Algebra for a Compact Logic)*. Let $\mathbb{P} = \langle P, \leqslant \rangle$ be a partial order. Then there is a *compact* logic L such that $\mathbb{L}_{\mathsf{L}}$ is order-isomorphic to $\mathbb{P}$.

So compactness does not, by itself, place any constraints on the structure of the partial order obtained via the Lindenbaum-Tarski construction. However, when combined with other properties (such as the one discussed next), compactness can be quite informative.

§ 1.2.3 Conjunctivity

Almost all of the major connectives are subject to philosophical controversy. Intuitionists disagree with classical logicians over whether double negation elimination is valid; relevant logicians disagree with classical logicians over whether disjunctive syllogism is valid; many nonclassical logics reject one or more of the standard rules for the connectives. But arguably the least controversial connective is conjunction. Although relevant logics sometimes abandon conjunction introduction, since adding conjunction introduction allows one to derive irrelevant consequences,[5] very few logics abandon conjunction elimination. So while we will return to the topic of connectives in § 1.3.4, it is natural to inquire into logics which have a conjunction for any two formulas (or any set of formulas). These concepts will be useful in proving the main triviality results of § 2.6.1.

[5] For instance, if $p, q \vDash p \wedge q$, then $p, q \vDash p$ by conjunction elimination and transitivity of $\vDash$.

Definition 1.2.21 *(Semiconjunction)*. Let L be a logic and let $\Gamma \subseteq \mathcal{L}$ and $\theta \in \mathcal{L}$. We say that θ is a ***semiconjunction*** of Γ in L if $[\theta]_{\mathsf{L}} = \bigwedge_{\phi\in\Gamma} [\phi]_{\mathsf{L}}$. We write "$\bigwedge \Gamma$" for any $\mathcal{L}$-formula such that $[\bigwedge \Gamma]_{\mathsf{L}} = \bigwedge_{\phi\in\Gamma} [\phi]_{\mathsf{L}}$. If $\Gamma = \{\phi, \psi\}$, we write "$\phi \wedge \psi$" instead of "$\bigwedge \{\phi, \psi\}$". L is ***semiconjunctive*** if any $\phi, \psi \in \mathcal{L}$ have a semiconjunction. (We could equivalently state this in terms of finitely many $\phi_1, \ldots, \phi_n \in \mathcal{L}$.) L is ***completely semiconjunctive*** if any $\Gamma \subseteq \mathcal{L}$ has a conjunction.

Fact 1.2.22 *(Semiconjunctivity and Semilattices)*. A logic L is (completely) semiconjunctive iff $\mathbb{L}_{\mathsf{L}}$ is a (complete) meet-semilattice.

Semiconjunctions automatically validate conjunction elimination. But even when defined, semiconjunctions need not validate conjunction introduction—that is, even if $\bigwedge \Gamma$ is defined, we are not automatically guaranteed that $\Gamma \vdash \bigwedge \Gamma$.

Example 1.2.23 *(Semiconjunctive Logics without Conjunction Introduction)*. Let $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ where $\mathcal{L} = \{p, q, p \wedge q\}$ and where $\Gamma \vdash \phi$ iff $\phi \in \Gamma$ or $p \wedge q \in \Gamma$. (Note L is Tarskian.) Then L is semiconjunctive (if $\phi \neq \psi$, then $\phi \wedge \psi = p \wedge q$). But L does not obey conjunction introduction since $p, q \nvdash p \wedge q$.

This justifies the prefix 'semi' in 'semiconjunctive': such logics are only guaranteed to have half of a usual conjunction. Still, almost all semiconjunctive logics of interest do obey conjunction introduction. This motivates the following definition.

Definition 1.2.24 *(Adjunctive Logic)*. A logic L is ***adjunctive*** if for any $\Gamma \subseteq \mathcal{L}$ and any semiconjunction $\theta \in \mathcal{L}$ of Γ in L, then $\Gamma \vdash \theta$. We say L is ***(completely) conjunctive*** if it is (completely) semiconjunctive and adjunctive.

Unlike semiconjunctivity, adjunctivity does not correspond exactly to any condition on Lindenbaum-Tarski algebras, since we can make a non-adjunctive logic adjunctive (even conjunctive) without any changes to the Lindenbaum-Tarski algebra (e.g., **Example 1.2.23**). We can say something of interest, however, viz., that conjunctive logics are logics where finitary consequence reduces to the order relation in Lindenbaum-Tarski algebras. That is:

Fact 1.2.25 *(Reducing (Finite) Consequence to the Lindenbaum-Tarski Algebra)*. Let L be a logic. Then L is (completely) conjunctive iff for all finite (and all infinite)

$\Gamma \subseteq \mathcal{L}$, $\bigwedge_{\phi\in\Gamma} [\phi]_{\mathsf{L}}$ is defined and for all $\psi \in \mathcal{L}$:

$$\Gamma \vdash \psi \quad \Leftrightarrow \quad \bigwedge_{\phi\in\Gamma} [\phi]_{\mathsf{L}} \leqslant [\psi]_{\mathsf{L}} .$$

Corollary 1.2.26 *(Reducing Compact Consequence to the Lindenbaum-Tarski Algebra).* Let L be a conjunctive logic. Then L is compact iff for all $\Gamma \subseteq \mathcal{L}$ and $\psi \in \mathcal{L}$:

$$\Gamma \vdash \psi \quad \Leftrightarrow \quad \exists \Gamma_0 \subseteq \Gamma\colon\ |\Gamma_0| < \aleph_0 \text{ and } \bigwedge_{\phi\in\Gamma_0} [\phi]_{\mathsf{L}} \leqslant [\psi]_{\mathsf{L}} .$$

In sum: conjunctivity says that all finite-premise inferences can be reduced to single-premise inferences; complete conjunctivity says that all inferences can be reduced to single-premise inferences; and compactness says that all inferences can be reduced to finite-premise inferences. The Lindenbaum-Tarski algebra is a summary of a logic's single-premise inferences. Thus, a compact conjunctive logic can be completely summarized by its Lindenbaum-Tarski algebra.

§ 1.3 Compositional Languages

The simple picture of a language discussed in the previous section will be a useful starting point for our investigations into different measures of expressive power. Nevertheless, for many purposes, it is too spartan. For in defining the expressive power of a language, we are not simply interested in whether everything that one language can say can be said in another language. We are also interested in whether the latter can interpret the former in a more-or-less uniform way. That is, we are interested in whether languages can interpret other languages in a way that more-or-less preserves their underlying syntactic structure. To determine whether this is the case, we need to be more explicit about what a language's underlying syntactic structure consists in. Making this precise is the aim of this section.

§ 1.3.1 Signatures and Subformulas

To start, we will lay out the syntactic foundations for compositional languages. The primary difference between compositional languages and simple languages is that the formulas of a compositional language are "built up" in a fairly systematic way. To make this precise, we introduce the notion of a "signature", which is essentially a kind of lexicon for formal languages.

Definition 1.3.1 *(Signature)*. A ***signature*** is a pair $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$, where $\mathtt{At}$ is a class, and $\mathtt{Op}$ is a class (disjoint from $\mathtt{At}$) of functions $\triangle$ of any arity. Where γ is an ordinal, we let $\mathtt{Op}^\gamma$ denote the class of γ-ary functions in $\mathtt{Op}$. The members of $\mathtt{At}$ are called the ***atomic*** Σ***-formulas*** and the members of $\mathtt{Op}$ are called the Σ***-operators***.

Definition 1.3.2 *(Generated Syntax)*. Where $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ is a signature, the Σ***-syntax*** is the smallest class $\mathcal{L}_\Sigma$ such that:

(i) $\mathtt{At} \subseteq \mathcal{L}_\Sigma$

(ii) for all $\triangle \in \mathtt{Op}^\gamma$ and all $\rho \in \mathcal{L}_\Sigma^\gamma$, if $\triangle(\rho)$ is defined, then $\triangle(\rho) \in \mathcal{L}_\Sigma$.

The members of $\mathcal{L}_\Sigma$ are called the Σ***-formulas*** or the $\mathcal{L}_\Sigma$***-formulas***. A Σ-formula ϕ is ***complex*** if $\phi = \triangle(\rho)$ for some $\triangle \in \mathtt{Op}^\gamma$ and some $\rho \in \mathcal{L}_\Sigma^\gamma$; otherwise, it is ***simple***.

Definition 1.3.3 *(Languages and Logics with Signatures)*. A Σ***-language*** is an ordered tuple of the form $\mathbf{L} = \langle \Sigma, \mathsf{C}, \Vdash \rangle$ where $\langle \mathcal{L}_\Sigma, \mathsf{C}, \Vdash \rangle$ is a language in the sense of **Definition 1.1.1**. A Σ***-logic*** is an ordered pair $\mathsf{L} = \langle \Sigma, \vdash \rangle$ where $\langle \mathcal{L}_\Sigma, \vdash \rangle$ is a logic in the sense of **Definition 1.1.5**.

As it stands, the definition of a language with signature is still quite minimal. For instance, formulas in a signature Σ may not uniquely decompose, i.e., we might have $\triangle_1(\rho_1) = \triangle_2(\rho_2)$ even if $\triangle_1 \neq \triangle_2$. Formulas might even contain themselves as proper Σ-subformulas; for example, nothing in **Definition 1.3.2** rules out there being a $\triangle \in \mathtt{Op}$ and a ϕ such that $\triangle(\phi) = \phi$. Likewise, our definition does not yet rule out the possibility that an atomic Σ-formula is also complex. But generally, we will focus on signatures that do not have these strange properties. More precisely, our focus is on languages where formulas are syntactically *unambiguous*.

Definition 1.3.4 *(Unambiguity)*. Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be a signature. A Σ-formula ϕ is ***unambiguous*** in Σ if either:

(i) $\phi \in \mathtt{At}$ and for no $\triangle \in \mathtt{Op}^\gamma$ and no $\rho \in \mathcal{L}_\Sigma^\gamma$ is $\phi = \triangle(\rho)$, or

(ii) $\phi \notin \mathtt{At}$ and for all $\triangle_1 \in \mathtt{Op}^{\gamma_1}$, all $\triangle_2 \in \mathtt{Op}^{\gamma_2}$, all $\rho_1 \in \mathcal{L}_\Sigma^{\gamma_1}$, and all $\rho \in \mathcal{L}_\Sigma^{\gamma_2}$, if $\triangle_1(\rho_1) = \phi = \triangle_2(\rho_2)$, then $\triangle_1 = \triangle_2$ and $\rho_1 = \rho_2$.

We say Σ is ***unambiguous*** if every $\phi \in \mathcal{L}_\Sigma$ is unambiguous in Σ. A Σ-language is ***unambiguous*** if Σ is unambiguous.

To say that a language is unambiguous is to say that one cannot "construct" the same formula via two distinct methods. That is, one can decompose each formula uniquely in the following sense:

Fact 1.3.5 *(Unique Decomposition).* If $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ is unambiguous, then for all $\phi \in \mathcal{L}_\Sigma$, either:

(i) $\phi \in \mathtt{At}$ and for no $\triangle \in \mathtt{Op}^\gamma$ and no $\rho \in \mathcal{L}_\Sigma^\gamma$ is $\phi = \triangle(\rho)$, or

(ii) $\phi \notin \mathtt{At}$ and there are unique $\triangle \in \mathtt{Op}^\gamma$ and $\rho \in \mathcal{L}_\Sigma^\gamma$ such that $\phi = \triangle(\rho)$.

Part (ii) in **Fact 1.3.5** applies equally to 0-ary operators. Thus, if $\triangle_0 \in \mathtt{Op}^0$, then for no distinct $\triangle_1 \in \mathtt{Op}^\gamma$ and $\rho \in \mathcal{L}_\Sigma^\gamma$ is $\triangle_0() = \triangle_1(\rho)$. In unambiguous signatures, 0-ary operators behave like atomic formulas in that they have no constituents. With this in mind, if $\triangle_0$ is 0-ary, we often write "$\triangle_0$" in place of "$\triangle_0()$" for brevity.

Unambiguous signatures have the nice feature that we can build the entire signature from the bottom up by applying the operators to the atomic formulas a finite number of times. To make this more precise, we need to define the notion of a subformula.

Definition 1.3.6 *(Subformula).* Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be a signature and let $\phi, \psi \in \mathcal{L}_\Sigma$ where $\phi = \triangle(\rho)$ for some $\gamma \neq 0$, some $\triangle \in \mathtt{Op}^\gamma$, and some $\rho \in \mathcal{L}_\Sigma^\gamma$. We say ψ is a ***Σ-constituent*** of ϕ if for some $\beta < \gamma$, $\rho(\beta) = \psi$. We say ψ is a ***proper Σ-subformula*** of ϕ if there are some $\theta_0, \ldots, \theta_n \in \mathcal{L}_\Sigma$ such that $\psi = \theta_0$, $\phi = \theta_n$, and for $i < n$, θ_i is a constituent of θ_{i+1}. We say ψ is an ***Σ-subformula*** of ϕ if either $\psi = \phi$ or ψ is a proper subformula of ϕ. We define:

$$\mathtt{Sub}_\Sigma(\phi) = \{\psi \mid \psi \text{ is a } \Sigma\text{-subformula of } \phi\}$$
$$\mathtt{PSub}_\Sigma(\phi) = \{\psi \mid \psi \text{ is a proper } \Sigma\text{-subformula of } \phi\}\,.$$

We generally drop mention of Σ when the Σ in question is obvious.

Definition 1.3.7 *(Well-Founded).* Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be a signature. A Σ-formula $\phi \in \mathcal{L}_\Sigma$ is ***well-founded*** in Σ if there are no $\psi_1, \psi_2, \psi_3, \ldots \in \mathcal{L}_\Sigma$ such that:

$$\mathtt{Sub}(\phi) \supset \mathtt{Sub}(\psi_1) \supset \mathtt{Sub}(\psi_2) \supset \mathtt{Sub}(\psi_3) \supset \cdots\,.$$

We say Σ is ***well-founded*** if every Σ-formula is well-founded in Σ.

The appendix to this chapter (§ 1.5) contains a proof that unambiguous signatures are well-founded. What this means is that, for unambiguous signatures, the relation $\sqsubseteq$ defined by $\phi \sqsubseteq \psi := \phi \in \mathtt{Sub}(\psi)$ is a weak well-founded partial order (i.e., there are no infinite $\sqsubseteq$-descending chains). And given that each formula is uniquely decomposable in such languages, we can think of a formula as a tree of formulas whose root is the formula itself and each node branches into that node's constituents. We will be primarily focused on such signatures throughout, since otherwise we would lose the ability to talk coherently about a "schema", which will be important in what follows.

Convention: Henceforth, when I say "signature", I mean "unambiguous signature" unless otherwise specified.

Furthermore, as we have defined things, operators need not be defined on every sequence of formulas. But for our purposes, this level of generality is not entirely necessary, and it is a hassle to have to worry about whether or not $\triangle(\rho)$ is defined all the time. Hence, we introduce the following definition and convention for convenience:

Definition 1.3.8 *(Partial Operators)*. Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be some signature. A Σ-operator $\triangle \in \mathtt{Op}^{\gamma}$ is ***partial*** if for some $\rho \in \mathcal{L}_{\Sigma}^{\gamma}$, $\triangle(\rho)$ is undefined (i.e., $\mathcal{L}_{\Sigma}^{\gamma} \nsubseteq \mathrm{dom}\,(\triangle)$). Otherwise, it is ***total***. We say Σ is ***total*** if each $\triangle \in \mathtt{Op}^{\gamma}$ is total.

Convention: Henceforth, when I say "signature", I mean "total signature" unless otherwise specified.

The definitions in what follows could be modified in straightforward ways to allow for partial signatures.

A word of caution before moving on. These conventions to take "signature" to mean "total unambiguous signature" should not hide the fact that when we define a new language from an old one, we must be careful to check that the new language is total and unambiguous. This is not something that is automatically given to us, and so we cannot be complacent: we will need to generally check that the newly constructed language has an unambiguous and total signature before applying any of results to it (just as we need to check that a logic is Tarskian when we construct it from some other mathematical object).

§ 1.3.2 Schemas and Substitutions

We now move on to defining the notion of a *schema*, which is crucial for many of the compositional approaches to defining expressive power in later chapters. Intuitively, a schema is a method for constructing a formula from basic constituents. One can "fill in" this construction with arbitrary formulas to construct a new formula. Or, more briefly, it is a syntax tree with placeholders for arbitrary formulas.

Definition 1.3.9 *(Schema)*. Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be a signature. Let Π be a class disjoint from $\mathcal{L}_\Sigma$ and $\mathtt{Op}$. The notion of a Σ-***schema*** with parameters in Π is defined recursively as follows:

- Each $\phi \in \mathcal{L}_{\mathbf{L}}$ is a Σ-schema with parameters in Π
- Each $\xi \in \Pi$ is a Σ-schema with parameters in Π
- If $\triangle \in \mathtt{Op}^\gamma$ and if σ is a γ-sequence of Σ-schemas with parameters in Π, then $\langle \triangle, \sigma \rangle$ is a Σ-schema with parameters in Π
- Nothing else is a Σ-schema with parameters in Π.

We let $\mathtt{Sch}_\Sigma(\Pi)$ be the class of Σ-schemas with parameters Π. If $\Theta \in \mathtt{Sch}_\Sigma(\Pi)$, the members of Π that occur somewhere in Θ are called the ***parameters*** of Θ. Where π is a sequence listing the members of Π exactly once, we write "$\Theta(\pi)$" to indicate that the parameters of Θ are all among the members of π. We usually use "π" for a sequence of parameters and "ξ" for a single parameter. We may write "$\mathtt{Sch}_\Sigma(\pi)$" for "$\mathtt{Sch}_\Sigma(\Pi)$" and "$\mathtt{Sch}_\Sigma(\xi)$" for "$\mathtt{Sch}_\Sigma(\{\xi\})$".

Let $\iota \colon \Pi \to \mathcal{L}_\Sigma$. The ***instantiation*** of a Σ-schema Θ with ι is the Σ-formula $\Theta[\iota]$ defined recursively as follows:

- If $\Theta = \phi \in \mathcal{L}_\Sigma$, then $\Theta[\iota] = \phi$.
- If $\Theta = \xi \in \Pi$, then $\Theta[\iota] = \iota(\xi)$.
- If $\Theta = \langle \triangle, \sigma \rangle$ where σ is a γ-sequence of Σ-schemas with parameters in Π, then $\Theta[\iota] = \triangle(\sigma[\iota])$, where $\sigma[\iota](\beta) = \sigma(\beta)[\iota]$ for $\beta < \gamma$.

If $\Theta(\pi)$ is a Σ-schema where π is a γ-sequence of distinct parameters and if $\rho \in \mathcal{L}_\Sigma^\gamma$, we write "$\Theta(\rho)$" for the instantiation of Θ with $\iota \colon \pi(\beta) \mapsto \rho(\beta)$.

If we think of a Σ-syntax as an algebra of formulas with operators $\mathtt{Op}_\Sigma$, then a schema is just a polynomial (or rather, a polynomial "symbol"[6]) over that algebra. The concept of a schema will be crucial in defining an adequate notion of a translation in later chapters.

The last syntactic concept we introduce is that of a substitution, which (in unambiguous signatures) can be used to provide an alternative definition of the concept of a subformula.

Definition 1.3.10 *(Substitution)*. Where $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ is a signature, a map $\sigma \colon \mathcal{L}_\Sigma \to \mathcal{L}_\Sigma$ is a Σ-***substitution*** if for all $\triangle \in \mathtt{Op}^\gamma$ and all $\rho \in \mathcal{L}_\Sigma^\gamma$:

$$\sigma(\triangle(\rho)) = \triangle(\sigma \circ \rho).$$

[6] Grätzer [2008, p. 39].

Fact 1.3.11 *(Substitutions are Determined by What They Do to Atomics)*. Let σ_1 and σ_2 be substitutions for a Σ-language **L** such that for each $\phi \in \mathtt{At}$, $\sigma_1(\phi) = \sigma_2(\phi)$. Then $\sigma_1 = \sigma_2$.

Substitutions in a signature $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ are essentially just endomorphisms on the language algebra $\langle \mathcal{L}_\Sigma, \mathtt{Op} \rangle$.

In what follows, we adopt the following notation:

Notation: Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$, let $\pi \in \mathtt{At}^\gamma$ where $\pi(\beta) = \pi(\beta')$ implies $\beta = \beta'$, and let $\rho \in \mathcal{L}_\Sigma^\gamma$. We write "$[\pi/\rho]$" (in postfix notation) for the substitution σ such that $\sigma(\pi(\beta)) = \rho(\beta)$ for all $\beta < \gamma$ and $\sigma(\chi) = \chi$ for all $\chi \in \mathtt{At}$ that do not occur in π. If $\gamma = n + 1 \in \mathbb{N}$, we may write "$[\pi(0)/\rho(0), \ldots, \pi(n)/\rho(n)]$" instead of $[\pi/\rho]$.

So for instance, $(p \wedge q)[p/(q \wedge r)] = ((q \wedge r) \wedge q)$. As usual with such notation, one should be warned that simultaneous substitution is not the same as iterated substitution—that is, in general, $[\pi_1/\phi_1, \ldots, \pi_n/\phi_n] \neq [\pi_1/\phi_1] \circ \cdots \circ [\pi_n/\phi_n]$. For instance, $p[p/q, q/r] = q$ whereas $p[p/q][q/r] = r$.

Using substitutions, we can actually introduce an alternative definition of the notion of a subformula as follows:[7]

Fact 1.3.12 *(Alternative Definition of Subformula)*. Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be a signature, let $\phi, \psi \in \mathcal{L}_\Sigma$, and let $\chi \in \mathtt{At}$.

(a) $\chi \in \mathtt{Sub}(\phi)$ iff for every Σ-substitution σ, if $\sigma(\phi) = \phi$, then $\sigma(\chi) = \chi$.

(b) If $\chi \notin \mathtt{Sub}(\phi)$, then $\psi \in \mathtt{Sub}(\phi)$ iff there is a formula ϕ' and a Σ-substitution σ where $\sigma(\chi) = \psi$ such that $\sigma(\phi') = \phi$.

(c) If $\chi \notin \mathtt{Sub}(\phi)$, then $\psi \in \mathtt{PSub}(\phi)$ iff there is a formula $\phi' \neq \chi$ and a Σ-substitution σ where $\sigma(\chi) = \psi$ such that $\sigma(\phi') = \phi$.

[7] This way of defining subformulas comes from Wójcicki [1988, pp. 17–18]. Actually, Wójcicki defines ψ to be a subformula of ϕ just in case there is a formula ϕ' and an atomic ξ such that $\phi = \phi'[\xi/\psi]$. This works fine for languages where each subformula has finitely many subformulas, and where one has an infinite supply of atomic formulas. Wójcicki was primarily working with such languages, so the definition was fitting for those purposes. But such a definition is inadequate for languages with formulas containing every atomic formula. For instance, consider the propositional formula $\phi = (p \wedge q) \wedge (p \wedge p) \wedge (q \wedge q)$. Clearly, $p \wedge q$ is a subformula of ϕ. But there is no formula ϕ' in a propositional logic whose only atomics are p and q such that $\phi = \phi'[p/(p \wedge q)]$ or $\phi'[q/(p \wedge q)]$, as both of these substitutions contain unwanted substitutions. Thus, only the more general definition of subformulas given in **Definition 1.3.6** works for these languages as well.

§ 1.3.3 Intensionality

Now that we have the notion of a compositional syntax, we can address the question of what it takes for the *semantics* of a language to be compositional. Intuitively, a semantics for a language is compositional if the meaning of a complex formula is completely determined by the meaning of its immediate constituents together with its syntactic structure. This can be precisified in a number of ways, but the standard way of articulating this idea is to invoke meaning algebras. Here, our meaning algebra is the collection of subclasses of our evaluation space. Thus, we can present a simplified notion of compositionality as follows:

Notation: Where **L** is a Σ-language, and where $\rho \in \mathcal{L}_\Sigma^\gamma$, we define $[\![\rho]\!] := [\![\cdot]\!] \circ \rho$, i.e., $[\![\rho]\!](\beta) = [\![\rho(\beta)]\!]$ for $\beta < \gamma$. In the finite case, we define:

$$[\![\phi_1, \ldots, \phi_n]\!] := \langle [\![\phi_1]\!], \ldots, [\![\phi_n]\!] \rangle.$$

In addition, where $\rho_1, \rho_2 \in \mathcal{L}_\Sigma^\gamma$, we write "$\rho_1 \equiv_{\mathbf{L}} \rho_2$" in place of "$\rho_1(\beta) \equiv_{\mathbf{L}} \rho_2(\beta)$ for all $\beta < \gamma$".

Definition 1.3.13 *(Intensionality)*. Let $\mathbf{L} = \langle \Sigma, \mathsf{C}, \Vdash \rangle$ be a Σ-language. We say that $\triangle \in \mathtt{Op}^\gamma$ is ***L-intensional*** if there is a function $[\![\triangle]\!] : \mathsf{C}^\gamma \to \mathsf{C}$ such that for all $\rho \in \mathcal{L}_\Sigma^\gamma$, if $\triangle(\rho)$ is defined, then:

$$[\![\triangle(\rho)]\!] = [\![\triangle]\!]([\![\rho]\!]).$$

Otherwise, $\triangle$ is ***L-hyperintensional***. If $\triangle$ is **L**-intensional, then $[\![\triangle]\!]$ is called the ***L-intension*** of $\triangle$. We say that **L** is ***intensional*** if each Σ-operator is **L**-intensional. Otherwise, **L** is ***hyperintensional***.

We can lift intensions up to schemas by induction:

Fact 1.3.14 *(Lifting Intensionality to Schemas)*. Let **L** be a intensional Σ-language. Then for any Σ-schema $\Theta(\pi)$, there is a function $[\![\Theta]\!]$ such that whenever $\Theta(\rho)$ is defined, $[\![\Theta(\rho)]\!] = [\![\Theta]\!]([\![\rho]\!])$.

Proposition 1.3.15 *(Intensionality Validates Substitution)*. Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be a signature and let **L** be a Σ-language. Then $\triangle$ is **L**-intensional iff $\triangle$ obeys the substitution of logical equivalents, i.e., for all $\rho_1, \rho_2 \in \mathcal{L}_\Sigma^\gamma$, if $\rho_1 \equiv \rho_2$, then $\triangle(\rho_1) \equiv \triangle(\rho_2)$.

Proof: The left-to-right direction is easy, given that $\rho_1 \equiv \rho_2$ iff $[\![\rho_1]\!] = [\![\rho_2]\!]$. For the right-to-left direction, it suffices to show that the definition of $[\![\triangle]\!]$ from **Definition 1.3.13** is a function (at least for definable members of C^γ; it does not matter what $\triangle$ does to a member of C^γ that is not picked out by some $\rho \in \mathcal{L}^\gamma_\Sigma$). But the functionality of $[\![\triangle]\!]$ obviously follows by the substitution of equivalents. ■

Hence, we can say an operator in a *logic* is intensional just in case it validates the substitution of logical equivalents. Likewise, we can say a logic is intensional if each of its operators are intensional.

Hyperintensional logics are gaining much attention at the moment.[8] A number of philosophers have argued that a variety of natural language operators—in particular, attitude verbs and counterfactuals—are hyperintensional. Most of the languages we will be focusing on here are intensional. But that does not mean that this framework does not have anything to say about hyperintensional languages.

Often, fans of hyperintensionality face some difficulty in specifying how hyperintensional an operator can get. The issue is that many of the standard reasons for accepting that an operator is hyperintensional threaten to undermine almost any non-trivial inference involving that operator: if you cannot even substitute logical equivalents in the scope of the operator, what other substitution principle could possibly survive? One way to resolve this problem is to revise one's language by allowing for more fine-grained states in the evaluation space. By doing so, one can weaken the logic and thereby convert the hyperintensional operator into an intensional one. While this may be undesirable, the effects can be mitigated by showing how the old notion of consequence can be interpreted in the new logic. So even if a logic is hyperintensional, one might find suitable intensional logics in the near vicinity. For the most part, this dissertation focuses on intensional languages.

§ 1.3.4 Connectives

We will generally be interested not just in charting the landscape of notions of expressive power from an abstract perspective but also in saying what follows for particular philosophically important languages. Many of these languages have familiar kinds of operators, such as negation, conjunction, and so forth. Since these connectives will be discussed later on, it will be useful to tabulate some of the more common connectives and note their properties.

There are broadly two ways to define the connectives. On the one hand, we can define them *proof-theoretically* in terms of how they interact with the underlying logic of a language. On the other hand, we can define them *semantically* in terms of how their semantic values are defined.

[8]See Mares 1997; Nolan 1997; Restall 1997; Vander Laan 1997; Brogaard and Salerno 2013; Jago 2014; Kment 2014; Priest 2016; Berto et al. 2017.

Proof-Theoretic Connectives

Definition 1.3.16 *(Proof-Theoretic Operators)*. Let $\mathsf{L} = \langle \mathcal{L}, \vdash \rangle$ be a logic and let $\theta \in \mathcal{L}$. We say θ is a ***proof-theoretic***…

- …***falsum*** if $\theta \vdash \mathcal{L}$.
- …***verum*** if $\vdash \theta$.
- …***negation*** of $\phi \in \mathcal{L}$ if for all $\Gamma \subseteq \mathcal{L}$: $\Gamma \vdash \theta$ iff $\Gamma, \phi \vdash \mathcal{L}$.
- …***(finitary) conjunction*** of $\phi, \psi \in \mathcal{L}$ if for all $\Gamma \subseteq \mathcal{L}$: $\Gamma \vdash \theta$ iff $\Gamma \vdash \phi$ and $\Gamma \vdash \psi$.
- …***(finitary) disjunction*** of $\phi, \psi \in \mathcal{L}$ if for all $\Gamma \subseteq \mathcal{L}$ and all $\chi \in \mathcal{L}$: $\Gamma, \theta \vdash \chi$ iff $\Gamma, \phi \vdash \chi$ and $\Gamma, \psi \vdash \chi$.
- …***infinitary conjunction*** of $\Delta \subseteq \mathcal{L}$ if for all $\Gamma \subseteq \mathcal{L}$: $\Gamma \vdash \theta$ iff $\Gamma \vdash \phi$ for all $\phi \in \Delta$.
- …***infinitary disjunction*** of $\Delta \subseteq \mathcal{L}$ if for all $\Gamma \subseteq \mathcal{L}$ and all $\psi \in \mathcal{L}$: $\Gamma, \theta \vdash \psi$ iff $\Gamma, \phi \vdash \psi$ for all $\phi \in \Delta$.
- …***conditional*** of $\phi, \psi \in \mathcal{L}$ if for all $\Gamma \subseteq \mathcal{L}$: $\Gamma \vdash \theta$ iff $\Gamma, \phi \vdash \psi$.
- …***biconditional*** of $\phi, \psi \in \mathcal{L}$ if for all $\Gamma \subseteq \mathcal{L}$: $\Gamma \vdash \theta$ iff $\Gamma, \phi \vdash \psi$ and $\Gamma, \psi \vdash \phi$.

We will talk of L as having a proof-theoretic operator *simpliciter* if it has that proof-theoretic operator for all formulas. For example, L has proof-theoretic conjunction if for each $\phi, \psi \in \mathcal{L}$, there is a proof-theoretic conjunction of ϕ and ψ in L.

Let L now be a Σ-logic where $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$. We say an operator $\triangle \in \mathtt{Op}$ is a ***proof-theoretic operator*** if it always outputs a particular proof-theoretic operator for some formulas. For example, $\triangle$ is a proof-theoretic (finitary) conjunction if for all $\phi, \psi \in \mathcal{L}$, $\triangle(\phi, \psi)$ is a proof-theoretic (finitary) conjunction of ϕ and ψ.

All of these notions lift to languages in the usual way.

Notation: We denote proof-theoretic operators as follows:

- $\bot$ for falsum
- $\top$ for verum
- $\neg$ for negation
- $\wedge$ for finitary conjunction
- $\vee$ for finitary disjunction
- $\bigwedge$ for infinitary conjunction
- $\bigvee$ for infinitary disjunction
- $\rightarrow$ for conditional
- $\leftrightarrow$ for biconditional

As usual, we write the binary operators using infix notation. Note that we are not assuming '$\neg$', '$\wedge$', etc. pick out some operators in a signature—even though when our logic has proof-theoretic negation, conjunction, etc., it will usually have them as operators. Context should make clear when '$\neg$' is used to pick out an operator in a signature and when it is simply used as a convenient shorthand in the metalanguage.

It is easy to check that any two proof-theoretic operators under the same classification are equivalent. That is:

Fact 1.3.17 *(Uniqueness of Proof-Theoretic Operators)*. Let L be a logic and let $\theta_1, \theta_2 \in \mathcal{L}$.

(a) If there is a $\phi \in \mathcal{L}$ such that θ_1 and θ_2 are both proof-theoretic negations of ϕ, then $\theta_1 \dashv\vdash \theta_2$.

(b) If there are some $\phi, \psi \in \mathcal{L}$ such that θ_1 and θ_2 are both proof-theoretic conjunctions (disjunctions, etc.) of ϕ and ψ, then $\theta_1 \dashv\vdash \theta_2$.

(c) If there is a $\Gamma \subseteq \mathcal{L}$ such that θ_1 and θ_2 are both proof-theoretic infinite conjunctions (disjunctions) of Γ, then $\theta_1 \dashv\vdash \theta_2$.

Thus, if $\triangle_1, \triangle_2 \in \mathtt{Op}$ are both proof-theoretic negations (conjunctions, etc.), then they are provably equivalent. For instance, if $\neg_1, \neg_2 \in \mathtt{Op}_{\mathsf{L}}$ are both proof-theoretic negations for a logic L, then $\neg_1 \phi \dashv\vdash \neg_2 \phi$. So we can write "$\neg$" ambiguously for any proof-theoretic negation without confusion. An interesting corollary of this is the following:

Corollary 1.3.18 *(Intuitionistic Logic and Classical Negation)*. There is no extension of intuitionistic logic that has a "classical negation", i.e., a proof-theoretic negation satisfying double negation elimination.

Proof (Sketch): If we add a classical negation $\neg_c$ to intuitionistic logic, then where $\neg_i$ is the original intuitionistic negation, we have $\neg_c \phi \dashv\vdash \neg_i \phi$ by **Fact 1.3.17**. But since $\neg_c \neg_c \phi \dashv\vdash \phi$, it follows that $\neg_i \neg_i \phi \dashv\vdash \phi$. ∎

Corollary 1.3.18 deserves some comment. The result does not state that one could not have a language containing two non-equivalent negations when one has an intuitionistic semantics and the other has a classical semantics. Indeed, that is entirely possible. What the result entails about such a language is that one of those negations would not be a *proof-theoretic* negation in the sense of **Definition 1.3.16**. As an illustration, suppose we add a classical negation to the standard Kripke semantics for intuitionistic logic, i.e., an operator $\neg_c$ such that for all pointed Kripke models $\langle \mathcal{M}, x\rangle$ and all ϕ:

$$\mathcal{M}, x \Vdash \neg_c \phi \quad \Leftrightarrow \quad \mathcal{M}, x \nVdash \phi.$$

Then the intuitionistic negation $\neg_i$ is no longer proof-theoretic: while $\Gamma \vdash \neg_i \phi$ implies $\Gamma, \phi \vdash \bot$, the converse is not true (e.g., $\neg_c p, p \vdash \bot$ but $\neg_c p \nvdash \neg_i p$). By contrast, the classical negation $\neg_c$ is a proof-theoretic negation here.

Since proof-theoretic connectives are completely decomposable into implications without mention of connectives, it immediately follows by **Fact 1.1.7** that:

Fact 1.3.19 *(Intensionality of Proof-Theoretic Operators)*. Each proof-theoretic operator is intensional.

Finally, the following are easy to verify:

Fact 1.3.20 *(Derived Proof-Theoretic Inferences)*. Assuming L has the appropriate proof-theoretic operators:

(i) $\phi, \neg\phi \vdash \mathcal{L}$.

(ii) If $\phi \vdash \psi$, then $\neg\psi \vdash \neg\phi$.

(iii) $\phi \vdash \neg\neg\phi$.

(iv) $\phi, \psi \vdash \phi \wedge \psi$.

(v) $\phi \wedge \psi \vdash \phi$, and $\phi \wedge \psi \vdash \psi$.

(vi) $\phi \vdash \phi \vee \psi$ and $\psi \vdash \phi \vee \psi$.

(vii) $\phi \vee \psi, \neg\phi \vdash \psi$ and $\phi \vee \psi, \neg\psi \vdash \phi$.

(viii) $\neg(\phi \vee \psi) \dashv\vdash \neg\phi \wedge \neg\psi$.

(ix) $\neg\phi \vee \neg\psi \vdash \neg(\phi \wedge \psi)$.

(x) $\vdash \phi \to \phi$.

(xi) $\phi \to \psi, \phi \vdash \psi$ and $\phi \to \psi, \neg\psi \vdash \neg\phi$.

(xii) $\phi \to \psi \vdash \neg\psi \to \neg\phi$.

(xiii) $\phi \to \psi, \psi \to \theta \vdash \phi \to \theta$.

(xiv) $\phi \to \psi, \psi \to \phi \vdash \phi \leftrightarrow \psi$.

(xv) $\phi \leftrightarrow \psi \vdash \phi \to \psi$ and $\phi \leftrightarrow \psi \vdash \psi \to \phi$.

Observe that **Fact 1.3.20**(iv) implies that L is adjunctive. Thus, if L contains a proof-theoretic conjunction, then L is conjunctive.[9] Observe also that we do not have any of the following inferences in general:

- $\neg\neg\phi \vdash \phi$
- $\vdash \phi \vee \neg\phi$
- $\neg(\phi \wedge \psi) \vdash \neg\phi \vee \neg\psi$.

Indeed, the reader might have observed that this list of derived proof-theoretic inferences include valid inferences from propositional intuitionistic logic and exclude classically valid but intuitionistically invalid inferences. That is because these proof-theoretic operators are all exactly how the rules governing the connectives are defined in intuitionistic logic. Thus, one can use intuitionistic logic to reason with proof-theoretic connectives.

Semantic Connectives

Definition 1.3.21 *(Truth-Functional Operators)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, \Vdash \rangle$ be a language and let $\theta \in \mathcal{L}$. We say θ is a ***truth-functional***...

- ...***falsum*** if $[\![\theta]\!] = \varnothing$.
- ...***verum*** if $[\![\theta]\!] = \mathsf{C}$.
- ...***negation*** of $\phi \in \mathcal{L}$ if $[\![\theta]\!] = \mathsf{C} - [\![\phi]\!]$.
- ...***(finitary) conjunction*** of $\phi, \psi \in \mathcal{L}$ if $[\![\theta]\!] = [\![\phi]\!] \cap [\![\psi]\!]$.
- ...***(finitary) disjunction*** of $\phi, \psi \in \mathcal{L}$ if $[\![\theta]\!] = [\![\phi]\!] \cup [\![\psi]\!]$.
- ...***infinite conjunction*** of $\Delta \subseteq \mathcal{L}$ if $[\![\theta]\!] = [\![\Delta]\!]$.
- ...***infinite disjunction*** of $\Delta \subseteq \mathcal{L}$ if $[\![\theta]\!] = \bigcup_{\phi\in\Delta} [\![\phi]\!]$.

As in **Definition 1.3.16**, we will talk of L having a certain truth-functional operator, or of an operator in a signature being a certain truth-functional operator, in the same way as before.

Fact 1.3.22 *(Truth-Functional Operators are Proof-Theoretic)*. Any truth-functional negation (conjunction, etc.) of some formula(s) is also a proof-theoretic negation (conjunction, etc.) of the same formula(s).

In some cases, we have the converse:

[9] For L to be semiconjunctive, one would need to add a ***semiconjunction***, where a semiconjunction of $\phi, \psi \in \mathcal{L}$ is a $\theta \in \mathcal{L}$ such that for all $\eta \in \mathcal{L}$: $\eta \vdash \theta$ iff $\eta \vdash \phi$ and $\eta \vdash \psi$. The difference between semiconjunctions and conjunctions, then, is whether we are allowed to appeal to multiple premises in proving ϕ and ψ.

Fact 1.3.23 *(Some Proof-Theoretic Operators are Truth-Functional)*. Any proof-theoretic verum is a truth-functional verum. Likewise, any proof-theoretic (infinitary) conjunction is a truth-functional (infinitary) conjunction.

But not always. In particular, proof-theoretic negation need not be a truth-functional negation. Likewise for disjunction. For example, the possibility semantics for classical propositional logic developed by Humberstone [1981] has a proof-theoretic disjunction and negation (since it preserves classical logic) but neither of their truth-functional versions.

We now record some simple but useful facts about truth-functional negation and how they relate to the properties of languages discussed in § 1.2.

Fact 1.3.24 *(Negation Implies Opinionation)*. Suppose L has truth-functional negation. Then L is opinionated.

Fact 1.3.25 *(Negation Collapses Finitarity and Compactness)*. Suppose $\mathbf{L}$ is finitary and has truth-functional negation. Then $\mathbf{L}$ is compact.

§ 1.4 Dynamic Languages

In the previous sections, we laid the foundations for thinking about language in a rather static way. A language, on this picture, is just a way of carving logical space. But while this may be adequate for thinking about theoretical or formal languages, many linguists and philosophers of language have argued that natural languages need a more dynamic picture.[10] According to dynamic semantics, the meaning of a sentence is not, in general, the class of possibilities in logical space where it is satisfied, but rather the dynamic effect that sentence has on a conversational context, i.e., its context-change potential. In this section, we lay out this alternative picture of language and explain its connection with the picture sketched in previous sections.

Definition 1.4.1 *(Dynamic Language)*. A ***dynamic language*** is a triple of the form $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$, where:

- $\mathcal{L}$ is a class (of formulas)
- C is a class (of ***contexts***)
- $[\cdot] : \mathcal{L} \to (\mathsf{C} \to \mathsf{C})$ is a map (the ***context change potential***).

[10] See Karttunen 1976; Kamp 1981; Heim 1983; Groenendijk and Stokhof 1991; Groenendijk et al. 1996; Stalnaker 1999. For an overview, see Rothschild and Yalcin 2015.

Notation: Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a dynamic language. We write $[\cdot]$ in postfix notation. Thus, where $s \in \mathsf{C}$ and $\phi \in \mathcal{L}$, we write "$s\,[\phi]$" in place of "$[\phi]\,(s)$". Where $\rho \in \mathcal{L}^{<\omega}$ and $|\rho| = n$, we write "$s\,[\rho]$" for "$s\,[\rho(0)] \cdots [\rho(n-1)]$". (In the special case where $|\rho| = 0$, we set $s\,[\rho] := s$.)

Moreover, we write "$s \Vdash \phi_1, \ldots, \phi_n$" in place of "$s\,[\phi_1] \cdots [\phi_n] = s$". As before, we write "$s \Vdash \Gamma$" in place of "$s \Vdash \phi$ for all $\phi \in \Gamma$".

Consequence for dynamic languages can be defined in multiple ways. One way to define consequence is in terms of the ordinary Tarksian notion of consequence articulated in **Definition 1.1.3**.

Definition 1.4.2 *(Static Entailment)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a dynamic language and let $\rho, \sigma \in \mathcal{L}^{<\omega}$. We say that ρ ***statically L-entails*** σ (written "$\rho \vDash_{\mathbf{L}} \sigma$") if for all $s \in \mathsf{C}$, if $s \Vdash \rho(i)$ for each $i < |\rho|$, then $s \Vdash \sigma(j)$ for each $j < |\sigma|$. The usual definitions and abbreviations for entailment apply.

Obviously, no confusion should arise from using "$\vDash_{\mathbf{L}}$" here for static entailment, since this corresponds to the ordinary notion of entailment for "static" languages (**Definition 1.1.1**). It is easy to check that static entailment is Tarskian.

There are also more dynamic ways of defining consequence. These definitions take a little more effort to articulate, however. If we are only dealing with a finite number of premises, it is fairly easy.

Definition 1.4.3 *(Dynamic Entailment)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a dynamic language and let $\rho, \sigma \in \mathcal{L}^{<\omega}$. We say that ρ ***dynamically L-entails*** σ (written "$\rho \Rrightarrow_{\mathbf{L}} \sigma$") if for all $s \in \mathsf{C}$, $s\,[\rho] \Vdash \sigma$—that is:

$$s\,[\rho]\,[\sigma] = s\,[\rho]\,.$$

Defining this notion of entailment when there are infinite sequences of premises, however, is a bit more complex. Suppose we want to articulate something like the thought that $\phi_1, \phi_2, \phi_3, \ldots$ dynamically entail ψ. We are not always guaranteed, however, that something like "$s\,[\phi_1]\,[\phi_2]\,[\phi_3] \cdots$" is going to be well-defined (e.g., what happens if $s\,[\phi_{2i+1}] \neq s$ but $s\,[\phi_{2i+1}]\,[\phi_{2i+2}] = s$ for $i \geqslant 0$?). What is more, it seems we do want something like infinite-premise inferences for modeling natural language inferences. Let L_n be a sentence saying "there are at least n-many things", and let *Inf* be a sentence saying "there are infinitely many things". Then arguably, we want to capture the fact the $L_1, L_2, L_3, \ldots$ entails *Inf*.[11] For our purposes, though,

[11] There are several ways this might be done. One way is to define the result of an infinite sequence of updates as a class of contexts that recur infinitely often along the update. Thus, where $\rho \in \mathcal{L}^{\omega}$, we would say $s' \in s\,[\rho]$ if there are infinitely many $n \in \mathbb{N}$ such that $s\,[\rho(0), \ldots, \rho(n)] = s'$. A more generalizable method might be obtained if we place a metric space on the class of contexts (which

we do not need to explore such generalizations of the dynamic framework, since finite-premise arguments will be the main focus.

Unlike static entailment, dynamic entailment is far from Tarskian. The only property it has reminiscent of Tarskian logics is transitivity:

Fact 1.4.4 *(Dynamic Entailment is Transitive).* Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a dynamic language. Then for all $\rho, \sigma, \tau \in \mathcal{L}^{<\omega}$, if $\rho \models\!\!\Rightarrow \sigma \models\!\!\Rightarrow \tau$, then $\rho \models\!\!\Rightarrow \tau$.

Dynamic entailment is not generally reflexive, commutative, contractive, or (reflexively) monotonic.

There is also a kind of "mixed" notion of entailment, which combines static and dynamic elements.

Definition 1.4.5 *(Kinematic Entailment).* Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a dynamic language and let $\rho, \sigma \in \mathcal{L}^{<\omega}$. We say that ρ ***kinematically L-entails*** σ (written "$\rho \models\!\!\Rightarrow_{\mathbf{L}} \sigma$") if for all $s \in \mathsf{C}$, if $s \Vdash \rho$, then $s \Vdash \sigma$—that is:

$$s[\rho] = s \quad \Rightarrow \quad s[\sigma] = s$$

Kinematic entailment, like dynamic entailment, is not Tarskian. But it is more Tarskian than dynamic entailment in that kinematic entailment is also reflexive.

Fact 1.4.6 *(Kinematic Entailment is Reflexive and Transitive).* Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a dynamic language. Then for all $\rho, \sigma, \tau \in \mathcal{L}^{<\omega}$:

(i) $\rho \models\!\!\Rightarrow \rho$

(ii) if $\rho \models\!\!\Rightarrow \sigma \models\!\!\Rightarrow \tau$, then $\rho \models\!\!\Rightarrow \tau$.

Again, however, kinematic entailment in general lacks commutativity, contraction, and (reflexive) monotonicity.

Dynamic entailment is more strict than kinematic entailment, which in turn is more strict than static entailment.

Example 1.4.7 *(Kinematic Entailment without Dynamic Entailment).* Let $\mathcal{L} = \{+\}$, let $\mathsf{C} = \mathbb{N}$, and let $n+ := n[+] = n + 1$ (see diagram below). Then for no n do we have that $n + + = n$. So vacuously, $+, + \models\!\!\Rightarrow +$. But $+ \not\models\!\!\Rightarrow +$, since for no n do we have that $n + + = n+$.

would informally represent information-theoretic distances amongst the contexts). Then we could take $s[\rho]$ to be the limit set of s for ρ—that is, $s' \in s[\rho]$ if, roughly, as we apply more and more of the sequence of updates in ρ, we approach s' "in the limit".

$$\begin{tikzcd} 0 \arrow[r, "+"] & 1 \arrow[r, "+"] & 2 \arrow[r, "+"] & 3 \arrow[r, "+"] & \cdots \end{tikzcd}$$

For a more natural example, see **Example 1.4.12** below.

Example 1.4.8 *(Static Entailment without Kinematic Entailment)*. Let $\mathcal{L} = \{*\}$, let $\mathsf{C} = \{0, 1\}$, and let $n* := n\,[*] = 1 - n$ (see diagram below). Vacuously, $*, * \models *$, but $*, * \not\mid\!\Rightarrow *$, since $0 * * = 0$, but $0* = 1$.

$$\begin{tikzcd} 0 \arrow[r, bend left, "*"] & 1 \arrow[l, bend left, "*"] \end{tikzcd}$$

Proposition 1.4.9 *(Hierarchy of Entailments for Dynamic Languages)*. Let $\mathbf{L}$ be a dynamic language, let $\rho, \sigma \in \mathcal{L}^{<\omega}$, and let $\psi \in \mathcal{L}$.

(a) If $\rho \mid\!\Rrightarrow_{\mathbf{L}} \sigma$, then $\rho \mid\!\Rightarrow_{\mathbf{L}} \sigma$.

(b) If $\rho \mid\!\Rightarrow_{\mathbf{L}} \psi$, then $\{\rho(k) \mid k < |\rho|\} \models_{\mathbf{L}} \psi$.

Proof:

(a) Suppose $\rho \mid\!\Rrightarrow \sigma$. Let $s\,[\rho] = s$. Then:

$$\begin{aligned} s\,[\sigma] &= s\,[\rho]\,[\sigma] \\ &= s\,[\rho] \\ &= s. \end{aligned}$$

(b) Suppose $\rho \mid\!\Rightarrow \psi$. Without loss of generality, let $\rho = \langle \phi_1, \ldots, \phi_n \rangle$. Let $s\,[\phi_i] = s$ for each $i \leqslant n$. Then:

$$\begin{aligned} s\,[\phi_1] \cdots [\phi_n] &= s\,[\phi_2] \cdots [\phi_n] \\ &= s\,[\phi_3] \cdots [\phi_n] \\ &\;\vdots \\ &= s. \end{aligned}$$

Hence, $s\,[\psi] = s$. ∎

Note that we cannot generalize part (b) of **Proposition 1.4.9** to allow for sequences of formulas in the conclusion of an entailment. That is, we cannot say in general that if $\rho \mid\!\Rightarrow_{\mathbf{L}} \sigma$, then $\{\rho(k) \mid k < |\rho|\} \models_{\mathbf{L}} \{\sigma(k) \mid k < |\sigma|\}$.

Example 1.4.10 *(Kinematic Entailment without Static Entailment)*. Let $\mathcal{L} = \langle p, q \rangle$, let $\mathsf{C} = \{0, 1\}$, and let $[\cdot]$ be defined as follows:

- $0\,[p] = 1\,[p] = 1$
- $0\,[q] = 1\,[q] = 0$.

Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ (see diagram below). Then $p \models\!\!\Rightarrow\; q, p$, since $s\,[p] = s$ implies $s = 1$, and $1\,[q]\,[p] = 0\,[p] = 1$. But $1\,[p] = 1$, while $1\,[q] = 0$. So $p \not\models q, p$.

$$\begin{tikzcd} 0 \arrow[loop left, "q"] \arrow[r, bend left, "p"] & 1 \arrow[l, bend left, "q"] \arrow[loop right, "p"] \end{tikzcd}$$

These different notions of entailment do coincide under special circumstances. We now characterize exactly when such collapses will occur. We start with the collapse of kinematic and dynamic entailments to static entailment.

Definition 1.4.11 *(Monotonicity)*. A dynamic language $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ is called ***kinematically monotonic*** if for all $\phi_1, \ldots, \phi_n \in \mathcal{L}$, $\phi_1, \ldots, \phi_n \models\!\!\Rightarrow\; \phi_i$—that is, for all $s \in \mathsf{C}$ and $1 \leqslant i \leqslant n$:

$$s\,[\phi_1] \cdots [\phi_n] = s \quad \Rightarrow \quad s\,[\phi_i] = s.$$

$\mathbf{L}$ is ***dynamically monotonic*** if for all $\phi_1, \ldots, \phi_n \in \mathcal{L}$, $\phi_1, \ldots, \phi_n \mathrel{|\!\!\Rrightarrow}\; \phi_i$—that is, for all $s \in \mathsf{C}$ and $1 \leqslant i \leqslant n$:

$$s\,[\phi_1] \cdots [\phi_n]\,[\phi_i] = s\,[\phi_1] \cdots [\phi_n]\,.$$

Example 1.4.12 *(Domain Semantics)*. Yalcin [2007, 2012] put forward the following semantics (the "domain semantics") for the epistemic modal 'might' (we ignore expanding the language with attitude verbs for simplicity). Let $\mathtt{Prop} = \{p_1, p_2, p_3, \ldots\}$ and let $\mathcal{L}_{\mathbf{D}}$ be recursively defined as follows:

$$\phi ::= p \mid \neg\,\phi \mid (\phi \wedge \phi) \mid \Diamond\phi.$$

An ***information model*** is a pair $\mathcal{M} = \langle W, V \rangle$ where W is a nonempty set and $V : \mathtt{Prop} \to \wp(W)$. Let D be the class of pairs of the form $\langle \mathcal{M}, s \rangle$ where $s \subseteq W$. Let $\mathbf{D} := \{\mathcal{L}_{\mathbf{D}}, \mathsf{D}, [\cdot]\}$, where (writing "$s\,[\phi]$" instead of "$\langle \mathcal{M}, s \rangle\,[\phi]$"):

$$\begin{array}{rcl} s\,[p] & = & \{w \in s \mid w \in V(p)\} \\ s\,[\neg\,\phi] & = & s - s\,[\phi] \end{array}$$

$$s[\phi \wedge \psi] = s[\phi][\psi]$$
$$s[\Diamond\phi] = \{w \in s \mid s[\phi] \neq \varnothing\}.$$

D is not dynamically monotonic. For example, $s[\Diamond p][\neg p][\Diamond p] = \varnothing$, while $s[\Diamond p][\neg p]$ does not have to be empty. But it is kinematically monotonic. Note first that (by an easy induction) $s[\phi] \subseteq s$ for all $\phi \in \mathcal{L}_{\mathbf{D}}$ and all s. So if $s[\phi_1]\cdots[\phi_n] = s$, then:

$$s = s[\phi_1]\cdots[\phi_n] \subseteq s[\phi_1] \subseteq s.$$

Hence, $s[\phi_1] = s$. By induction, then $s[\phi_i] = s$.

These versions of monotonicity are quite strong. They already suffice to restore the normal Tarskian properties (e.g., contraction, commutativity, etc.). But rather than verify these properties by hand, one can establish them all at once with the following result:

Proposition 1.4.13 *(Collapsing to Static Entailment)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a dynamic language.

(a) **L** is kinematically monotonic iff static entailment implies kinematic entailment, i.e., for all $\phi_1, \ldots, \phi_n, \psi_1, \ldots, \psi_m \in \mathcal{L}$:

$$\phi_1, \ldots, \phi_n \vDash \psi_1, \ldots, \psi_m \quad \Rightarrow \quad \phi_1, \ldots, \phi_n \mathrel{|\!\!\Rightarrow} \psi_1, \ldots, \psi_m.$$

(b) **L** is dynamically monotonic iff static entailment implies dynamic entailment, i.e., for all $\phi_1, \ldots, \phi_n, \psi_1, \ldots, \psi_m \in \mathcal{L}$:

$$\phi_1, \ldots, \phi_n \vDash \psi_1, \ldots, \psi_m \quad \Rightarrow \quad \phi_1, \ldots, \phi_n \mathrel{|\!\!\!\Rrightarrow} \psi_1, \ldots, \psi_m.$$

Proof:

(a) The right-to-left direction is immediate given that $\phi_1, \ldots, \phi_n \vDash \phi_i$. For the left-to-right direction, let $\phi_1, \ldots, \phi_n \vDash \psi_1, \ldots, \psi_m$ and suppose that $s \in \mathsf{C}$ is such that $s[\phi_1]\cdots[\phi_n] = s$. By kinematic monotonicity, $s[\phi_i] = s$ for $1 \leqslant i \leqslant n$. So by static entailment, $s[\psi_j] = s$ for $1 \leqslant j \leqslant m$. But then by repeated application, it follows that $s[\psi_1]\cdots[\psi_m] = s$. So $\phi_1, \ldots, \phi_n \mathrel{|\!\!\Rightarrow} \psi_1, \ldots, \psi_m$.

(b) Again, the right-to-left direction is immediate. For the left-to-right direction, let $\phi_1, \ldots, \phi_n \vDash \psi_1, \ldots, \psi_m$ and let $s \in \mathsf{C}$. By dynamic monotonicity, we know that $s\,[\phi_1]\cdots[\phi_n] \Vdash \phi_i$ where $1 \leqslant i \leqslant n$. Hence, $s\,[\phi_1]\cdots[\phi_n] \Vdash \psi_j$ where $1 \leqslant j \leqslant m$. But then by repeated functional application, we obtain $s\,[\phi_1]\cdots[\phi_n]\,[\psi_1]\cdots[\psi_m] = s\,[\phi_1]\cdots[\phi_n]$. So $\phi_1,\ldots,\phi_n \Rrightarrow \psi_1,\ldots,\psi_m$. ∎

Very roughly, **Proposition 1.4.13** states that the difference between kinematic/dynamic entailment and static entailment is that static entailment is reflexively monotonic while the former notions of entailment are not.

It turns out the kinematic monotonicity is equivalent to another important property, viz., what Rothschild and Yalcin [2015, p. 13] call "antisymmetry":

Definition 1.4.14 *(Antisymmetry)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot]\rangle$ be a dynamic language. We say $\mathbf{L}$ is ***antisymmetric*** if for all $s, s' \in \mathsf{C}$ and all $\phi_1,\ldots,\phi_n,\psi_1,\ldots,\psi_m \in \mathcal{L}$, if $s\,[\phi_1]\cdots[\phi_n] = s'$ and $s'\,[\psi_1]\cdots[\psi_m] = s$, then $s = s'$.

Proposition 1.4.15 *(Antisymmetry is Kinematic Monotonicity)*. A dynamic language is antisymmetric iff it is kinematically monotonic.

Proof:

(⇐) Suppose $s\,[\phi_1]\cdots[\phi_n] = s'$ and $s'\,[\psi_1]\cdots[\psi_m] = s$. Then we have:

$$s\,[\phi_1]\cdots[\phi_n]\,[\psi_1]\cdots[\psi_m] = s.$$

Hence, $s\,[\phi_i] = s$ for $1 \leqslant i \leqslant n$ (and also $s\,[\psi_j] = s$ for $1 \leqslant j \leqslant m$). So by repeated functional application, $s' = s\,[\phi_1]\cdots[\phi_n] = s$.

(⇒) Suppose $s\,[\phi_1]\cdots[\phi_n] = s$. Let $s_0 = s$ and let $s_{i+1} = s_i\,[\phi_i]$. Suppose for induction that $s = s_i$. Then $s\,[\phi_i] = s_{i+1}$ and $s_{i+1}\,[\phi_{i+1}]\cdots[\phi_n] = s$. Hence, $s = s_{i+1}$. Since $s = s_0$ by definition, we have $s = s_i$ where $1 \leqslant i \leqslant n$. So $s\,[\phi_i] = s_i\,[\phi_i] = s_{i+1} = s$. ∎

Rothschild and Yalcin [2015] show that a conversational system is isomorphic to a "weakly static" system just in case it is antisymmetric. Putting these together, we now have further characterizations of the notion of weakly static system in terms of the collapse of static and kinematic entailment.

This exactly characterizes when the less static notions of entailment collapse to static entailment. But what about kinematic and dynamic entailment? When do they collapse? The answer can be given with the following definition:

Definition 1.4.16 *(Strong Idempotence)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$. We say $\mathbf{L}$ is ***strongly idempotent*** if for all $\phi_1, \ldots, \phi_n \in \mathcal{L}$, $\phi_1, \ldots, \phi_n \models\!\!\!\Rightarrow \phi_1, \ldots, \phi_n$—that is, for all $s \in \mathsf{C}$:

$$s[\phi_1]\cdots[\phi_n][\phi_1]\cdots[\phi_n] = s[\phi_1]\cdots[\phi_n].$$

We say $\mathbf{L}$ is ***(simply) idempotent*** if for all $\phi \in \mathcal{L}$, $\phi \models\!\!\!\Rightarrow \phi$—that is, for all $s \in \mathsf{C}$:

$$s[\phi][\phi] = s[\phi].$$

Proposition 1.4.17 *(Collapsing to Kinematic Entailment)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a dynamic language. Then $\mathbf{L}$ is strongly idempotent iff kinematic entailment implies dynamic entailment, i.e., for all $\phi_1, \ldots, \phi_n, \psi_1, \ldots, \psi_m \in \mathcal{L}$:

$$\phi_1, \ldots, \phi_n \models\!\!\!\Rightarrow \psi_1, \ldots, \psi_m \quad \Rightarrow \quad \phi_1, \ldots, \phi_n \models\!\!\!\Rightarrow \psi_1, \ldots, \psi_m.$$

Proof: The right-to-left direction is immediate from **Fact 1.4.6**. For the converse, suppose $\phi_1, \ldots, \phi_n \models\!\!\!\Rightarrow \psi_1, \ldots, \psi_m$. Let $s \in \mathsf{C}$ and let $t := s[\phi_1]\cdots[\phi_n]$. By strong idempotence, $t[\phi_1]\cdots[\phi_n] = t$. Hence, $t[\psi_1]\cdots[\psi_m] = t$, which means:

$$s[\phi_1]\cdots[\phi_n][\psi_1]\cdots[\psi_m] = s[\phi_1]\cdots[\phi_n].$$

Thus, $\phi_1, \ldots, \phi_n \models\!\!\!\Rightarrow \psi_1, \ldots, \psi_m$. ■

Another important property of dynamic languages is *commutativity* of context shift:

Definition 1.4.18 *(Commutativity)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a dynamic language. We say $\mathbf{L}$ is ***commutative*** if for all $\phi, \psi \in \mathcal{L}$ and for all $s \in \mathsf{C}$:

$$s[\phi][\psi] = s[\psi][\phi].$$

Fact 1.4.19 *(Commutativity of Premises)*. Let $\mathbf{L} = \langle \mathcal{L}, \mathsf{C}, [\cdot] \rangle$ be a commutative language. Then the order of premises in dynamic entailment does not matter, i.e., for any $\phi_1, \ldots, \phi_n, \psi_1, \ldots, \psi_m \in \mathcal{L}$ and any permutations $\pi : n \to n$ and $\sigma : m \to m$:

$$\phi_1, \ldots, \phi_n \models\!\!\!\Rightarrow \psi_1, \ldots, \psi_m \quad \Leftrightarrow \quad \phi_{\pi(1)}, \ldots, \phi_{\pi(n)} \models\!\!\!\Rightarrow \psi_{\sigma(1)}, \ldots, \psi_{\sigma(m)}.$$

By themselves, commutativity and idempotence seem to place relatively weak constraints on a dynamic language. For instance, a commutative language need not be idempotent (see **Example 1.4.7** for an example). Nor must a commutative language be kinematically monotonic (see **Example 1.4.8** for an example). Likewise, idempotent languages need not be commutative or kinematically monotonic. Commutativity simply says that the order of premises does not matter for dynamic entailment, and idempotence simply says that dynamic entailment is reflexive. Nevertheless, these two properties *taken together* place a relatively strong constraint on a dynamic language:

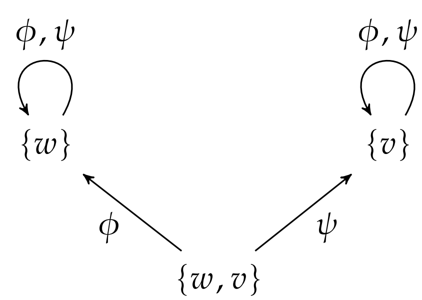
Fact 1.4.20 *(Commutativity and Idempotence)*. If **L** is commutative and idempotent, then it is dynamically monotonic.

Thus, commutative and idempotent languages collapse the distinctions between the different notions of entailment for dynamic languages. What is more, the combination of commutativity and idempotence is of philosophical significance, as it is exactly what characterizes isomorphism to what Rothschild and Yalcin [2015, p. 10] call "strongly static" conversational systems.

Thus, it is worth inquiring into what exactly these combination of properties say about a dynamic language's entailment relation. As noted above, commutative and idempotent languages are dynamically monotonic. Moreover, dynamically monotonic languages are strongly idempotent by **Propositions 1.4.9**, **1.4.13 and 1.4.17**. However, dynamic monotonicity does not imply commutativity, as a simple example illustrates:

Example 1.4.21 *(Non-Commutative Dynamically Non-Monotonic Language)*. Let $\mathcal{L} = \{\phi, \psi\}$, let $\mathsf{C} = \wp(\{w, v\}) - \{\varnothing\}$, and define $[\cdot]$ as follows:

- $\{w, v\}[\phi] = \{w\}$, $\{w, v\}[\psi] = \{v\}$
- $X[\phi] = X[\psi] = X$ where $\varnothing \neq X \subset \{w, v\}$.



Observe that $\{w, v\}[\phi][\psi] = \{w\}$ while $\{w, v\}[\psi][\phi] = \{v\}$. So **L** is not commutative. But static entailment does imply dynamic entailment. No formula (or sequence of formulas) is statically valid or dynamically valid. And ϕ and ψ are both statically and dynamically equivalent to one another.

There is, however, a necessary and sufficient condition for $[\cdot]$ to be commutative and idempotent in terms of collapse of entailments plus one further constraint.

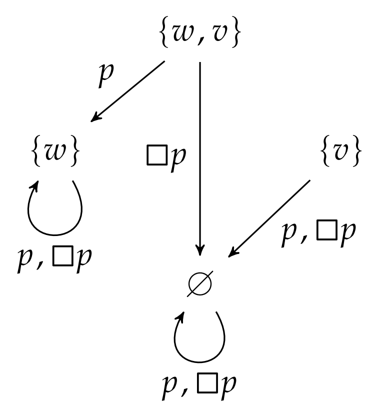
Definition 1.4.22 *(Dynamic Replacement)*. Let $\mathbf{L}$ be a dynamic language. We say $\mathbf{L}$ obeys ***dynamic replacement*** if for all $\rho, \sigma \in \mathcal{L}^{<\omega}$:

$$\rho \Leftarrow\!\mid\!\Rightarrow \sigma \quad \Rightarrow \quad [\rho] = [\sigma] \,.$$

Dynamic replacement means that any two sequences of formulas that accept one another will generate the same update. Example 1.4.21 is an example of a language where dynamic replacement fails. Another example where dynamic replacement fails comes from update semantics.

Example 1.4.23 *(Simple Language with 'Must')*. Let $\mathbf{L} = \langle \{p, \Box p\}, \wp(\{w, v\}), [\cdot] \rangle$, where:

- $\{w, v\}[p] = \{w\}[p] = \{w\}$, $\{v\}[p] = \varnothing[p] = \varnothing$
- $\{w\}[\Box p] = \{w\}$, $\{w, v\}[\Box p] = \{v\}[\Box p] = \varnothing[\Box p] = \varnothing$.



Here, $p \Leftarrow\!\mid\!\Rightarrow \Box p$, but $[p] \neq [\Box p]$.

Lemma 1.4.24 *(Commutativity Implies Dynamic Replacement)*. If $\mathbf{L}$ is a commutative dynamic language, then $\mathbf{L}$ obeys dynamic replacement.

Proof: Suppose $\rho \Leftarrow\!\!|\!\!\Rightarrow \sigma$. Thus, for all $s \in \mathsf{C}$, we have:

$$\begin{aligned} s[\rho] &= s[\rho][\sigma] \\ &= s[\sigma][\rho] \\ &= s[\sigma]. \end{aligned}$$

The second step follows by repeated applications of commutativity. ∎

Example 1.4.25 *(Dynamic Replacement Does Not Imply Commutativity)*. Let $\mathbf{B} = \langle \{L, R\}, 2^{<\omega}, [\cdot] \rangle$, where for all $s \in 2^{<\omega}$:

$$\begin{aligned} s[L] &= s \frown \langle 0 \rangle \\ s[R] &= s \frown \langle 1 \rangle. \end{aligned}$$

In other words, the states in $\mathbf{B}$ are nodes in an infinite binary tree, and L says "go left" and R says "go right".

Vacuously, $\mathbf{B}$ satisfies dynamic replacement. First, notice that for no nonempty $\rho, \sigma \in \{L, R\}^{<\omega}$ is $\rho \Leftarrow\!\!|\!\!\Rightarrow \sigma$, since nonempty sequences necessarily shift to states further up the tree. Moreover, $\rho \Leftarrow\!\!|\!\!\Rightarrow \langle\,\rangle$ only if $\rho = \langle\,\rangle$. So dynamic replacement holds in $\mathbf{B}$. But $[\cdot]$ is not commutative: $[L, R] \neq [R, L]$.

While idempotence is not equivalent to dynamic monotonicity, and while commutativity is not equivalent to dynamic replacement, the conjunction of the former is equivalent to the conjunction of the latter.

Proposition 1.4.26 *(Characterizing Commutativity and Idempotence)*. A dynamic language is commutative and idempotent iff it is dynamically monotonic and obeys dynamic replacement.

Proof: The left-to-right direction has already been shown. For the right-to-left direction, we just need to show that our dynamic language $\mathbf{L}$ is commutative, since idempotence is guaranteed by dynamic monotonicity and **Proposition 1.4.13**. First, observe that for any $s \in \mathsf{C}$ and any $\phi, \psi \in \mathcal{L}$:

$$\begin{aligned} s[\phi][\psi][\psi][\phi] &= s[\phi][\psi][\phi] \\ &= s[\phi][\psi]. \end{aligned}$$

Both steps follow from dynamic monotonicity. Likewise, $s[\psi][\phi][\phi][\psi] = s[\psi][\phi]$. Hence, $\phi, \psi \Leftarrow\!\!|\!\!\Rightarrow \psi, \phi$. By dynamic replacement, it follows that $[\phi][\psi] = [\psi][\phi]$. ∎

Thus, we have another characterization of "strongly static" conversational systems: a strongly static conversational system is one where the dynamic entailment relation is dynamically monotonic and obeys dynamic replacement.

§ 1.5 Appendix

We now provide the proof that unambiguous signatures are well-founded. First, the following properties of $\mathtt{Sub}$ are easy to verify.

Fact 1.5.1 *(Properties of Subformulas)*. Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be a signature.

(a) $\mathtt{Sub}(\phi) = \{\phi\} \cup \mathtt{PSub}(\phi)$.

(b) $\phi \in \mathtt{PSub}(\psi)$ iff $\mathtt{Sub}(\phi) \subseteq \mathtt{PSub}(\psi)$.

(c) $\phi \in \mathtt{Sub}(\psi)$ iff $\mathtt{Sub}(\phi) \subseteq \mathtt{Sub}(\psi)$.

(d) If $\phi \in \mathtt{Sub}(\psi)$ and $\psi \in \mathtt{Sub}(\theta)$, then $\phi \in \mathtt{Sub}(\theta)$.

(e) $\phi \in \mathtt{PSub}(\psi)$ and $\psi \in \mathtt{PSub}(\theta)$, then $\phi \in \mathtt{PSub}(\theta)$.

This next fact follows from the definition of a Σ-constituent (**Definition 1.3.6**):

Fact 1.5.2 *(Unambiguous Atomics Have No Proper Subformulas)*. Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be unambiguous, and let $\phi \in \mathtt{At}$ or $\phi = \triangle^0$. Then $\mathtt{PSub}(\phi) = \varnothing$.

Corollary 1.5.3 *(Nothing is a Proper Subformula of Itself)*. Suppose $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ is unambiguous. Then $\phi \notin \mathtt{PSub}(\phi)$ for all $\phi \in \mathcal{L}_\Sigma$.

Proof: Let $\mathcal{L}' := \mathcal{L}_\Sigma - \{\psi \mid \psi \in \mathtt{PSub}(\psi)\}$. We will show that $\mathcal{L}' = \mathcal{L}_\Sigma$ by showing it contains $\mathtt{At}$ and is closed under $\mathtt{Op}$. We proceed by induction. The base case is covered by **Fact 1.5.2**. Suppose now that $\phi = \triangle(\rho) \in \mathcal{L}_\Sigma$ where $\triangle \in \mathtt{Op}^\gamma$ and for each $\beta < \gamma$, $\rho(\beta) \in \mathcal{L}'$. Suppose for *reductio* that $\phi \notin \mathcal{L}'$. That means that $\phi \in \mathtt{PSub}(\phi)$. So there must be some $\theta_0, \ldots, \theta_n$ such that $\phi = \theta_0 = \theta_n$ and for each $i < n$, θ_i is a Σ-constituent of θ_{i+1}. Since $\theta_n = \phi$, and since θ_{n-1} must be a Σ-constituent of θ_n, it follows that for some $\beta < \gamma$, $\theta_{n-1} = \rho(\beta)$. But then $\rho(\beta), \phi = \theta_0, \theta_1, \ldots, \theta_{n-1} = \rho(\beta)$ is a chain of Σ-constituents. So $\rho(\beta) \in \mathtt{PSub}(\rho(\beta))$, contrary to the assumption that $\rho(\beta) \in \mathcal{L}'$, $\lightning$. Thus, $\phi = \triangle(\rho) \in \mathcal{L}'$. Therefore, $\mathcal{L}' = \mathcal{L}$. ■

This feature of unambiguous signatures crucially assumes that the syntax of a signature is the *smallest* class containing the atomic formulas that is closed under the operators. We could revise **Definition 1.3.2** so that Σ-syntaxes need not be the smallest class as long as it is closed under the Σ-operators. In that case, we could have unambiguous languages containing formulas that were proper subformulas of themselves. But we do not study such languages here.

Lemma 1.5.4 *(Uniqueness of Subformulas).* Let $\Sigma = \langle \mathtt{At}, \mathtt{Op} \rangle$ be unambiguous. Then for all $\phi, \psi \in \mathcal{L}_\Sigma$, if $\mathtt{Sub}(\phi) = \mathtt{Sub}(\psi)$, then $\phi = \psi$.

Proof: By induction on ϕ. The base case is immediate by **Fact 1.5.2**. Now, let $\phi = \triangle(\rho)$ where $\triangle \in \mathtt{Op}^\gamma$ and $\rho \in \mathcal{L}_\Sigma^\gamma$. Suppose that $\mathtt{Sub}(\phi) = \mathtt{Sub}(\psi)$ and (for our inductive hypothesis) that for each $\beta < \gamma$, $\mathtt{Sub}(\rho(\beta)) = \mathtt{Sub}(\psi)$ implies $\rho(\beta) = \psi$. Observe that $\mathtt{PSub}(\phi) = \bigcup_{\beta<\gamma} \mathtt{Sub}(\rho(\beta))$. If $\psi \notin \mathtt{PSub}(\phi)$, then $\phi = \psi$. If $\psi \in \mathtt{PSub}(\phi)$, then $\psi \in \mathtt{Sub}(\rho(\beta))$ for some $\beta < \gamma$. Hence, $\mathtt{Sub}(\rho(\beta)) \supseteq \mathtt{Sub}(\psi) = \mathtt{Sub}(\phi) \supseteq \mathtt{Sub}(\rho(\beta))$. Thus, $\mathtt{Sub}(\rho(\beta)) = \mathtt{Sub}(\psi) = \mathtt{Sub}(\phi)$. So by inductive hypothesis, $\rho(\beta) = \psi = \phi$,[a] and thus $\phi = \psi$. ■

[a] This actually entails a contradiction, since $\psi \in \mathtt{PSub}(\phi)$ but $\phi \notin \mathtt{PSub}(\phi)$. Hence, we cannot have $\psi \in \mathtt{PSub}(\phi)$, which suffices to reach our conclusion anyway.

Corollary 1.5.5 *(Unambiguity Implies Well-Foundedness).* Every unambiguous signature is well-founded.

Proof: Essentially the same proof as **Corollary 1.5.3**. ■

Chapter 2

Translation

One of the key methods for making precise the claim that one language is as expressive as another is to provide a *translation* between them. Informally, a translation maps sentences of one language to "equivalent" sentences—sentences with "the same meaning"—in another language. If everything that can be said by one language can be said by another, one should in principle be able to specify a translation from the former to the latter that preserves meaning.

Spelling out what "equivalent" means in this context is not entirely straightforward. If two languages share the same conception of logical space, then we can spell out the notion in terms of semantic value: each sentence of the weaker language must be mapped to a sentence of the stronger language that has the same semantic value. But when languages fail to share a common conception of logical space, they will generally assign different types of semantic values to their sentences. Thus, this simple characterization of equivalence will make it impossible for there ever to be translations between such languages.

To illustrate, suppose we take two versions of the language of classical propositional logic. In the first, our points of evaluation are valuation functions, i.e., functions from atomic formulas to $\{0, 1\}$. In the second, our points of evaluation are just sets of atomic formulas. Intuitively, these two languages are just notational variants. But for no sentence in one language is there a sentence of the other language with the same semantic value because semantic values in the two languages are different types of objects.

In such cases, we need an alternative characterization of what "equivalence" means. There are several ways one could go at this point. But it is usually agreed that if a translation preserves the meaning of a sentence, it ought to preserve its inferential role. If the translation of a valid argument in the source language is not a valid argument in the target language, then this is a sign that something has gone wrong with the translation. Thus, one necessary condition on a translation is that it faithfully embeds the logic of the source language in the logic of the target language.

Of course, unless the meaning of a sentence is identified by its inferential role, as in inferentialism, preservation of valid inference patterns is not in general sufficient for a mapping to count as an adequate translation between languages. It is easy to preserve the valid inference patterns in a translation without preserving *meaning*. Still, since preservation of inference patterns seems to be at least a necessary condition for a translation to be adequate, it is worth investigating when inference-preserving translations from one language to another exist and what we can infer about the expressive power of a language from the existence of such translations.

In this chapter, we focus on translations that preserve exactly the valid inference patterns of a language. We begin in § 2.1 by providing a minimal definition of a such a translation and some examples illustrating the concept. Alternative formulations of translatability are also explored. Then in § 2.2–2.4, we explain the different notions of equivalence from the perspective of translatability, remarking on the philosophical difference between expressive equivalence and notational variance in § 2.5. However, it turns out that mere preservation of valid inference patterns is a relatively weak constraint. We show in § 2.6 that a great number of languages are translatable into classical propositional logic, including classical *first-order* logic. To avoid such consequences, we will move to a compositional setting and explore the varieties of translations in that framework in Chapter 3.

§ 2.1 Definition

To begin, we start with the definition of a translation between logics as well as some simple examples illustrating the concept. We will then examine alternative equivalent definitions of the concept.

Definition 2.1.1 *(Translation)*. Let L_1 and L_2 be some logics and let $\Lambda \subseteq \mathcal{L}_2$. We say that a function $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is a ***quasitranslation*** from L_1 into L_2 if for all $\Gamma \subseteq \mathcal{L}_1$ and all $\phi \in \mathcal{L}_1$:

$$\Gamma \vdash_1 \phi \quad \Rightarrow \quad \mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi).$$

We say that $\mathtt{t}$ is a ***translation*** (written "$\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$") if the converse is also the case, i.e., for all $\Gamma \subseteq \mathcal{L}_1$ and all $\phi \in \mathcal{L}_1$:

$$\Gamma \vdash_1 \phi \quad \Leftrightarrow \quad \mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi).$$

We call a pair of translations $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ and $\mathtt{s}\colon \mathsf{L}_2 \rightsquigarrow \mathsf{L}_1$ is a ***translation scheme*** (written "$\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \overset{\leftrightsquigarrow}{\rightsquigarrow} \mathsf{L}_2$"). We say L_1 is ***translatable*** into L_2 (written "$\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$") if $\mathtt{t}\colon \mathcal{L}_1 \rightsquigarrow \mathcal{L}_2$ for some $\mathtt{t}$. We say L_1 and L_2 are ***intertranslatable*** (written "$\mathsf{L}_1 \overset{\leftrightsquigarrow}{\rightsquigarrow} \mathsf{L}_2$") if $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \overset{\leftrightsquigarrow}{\rightsquigarrow} \mathsf{L}_2$ for some $\mathtt{t}$ and $\mathtt{s}$. These notions are lifted from logics to languages in the obvious way.

Translations have been studied in a variety of places.[1] Essentially, a quasitranslation is a logic-homomorphism and a translation is a logic-embedding.

Right away, it is worth observing the following obvious facts about translations (which we often invoke without mention):

Fact 2.1.2 *(Simple Observations about Interpretations and Translations).*

(a) If $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ and $\mathtt{s}\colon \mathsf{L}_2 \rightsquigarrow \mathsf{L}_3$, then $\mathtt{s} \circ \mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_3$.

(b) If $\mathsf{L}_1 \subseteq \mathsf{L}_2$, then $\mathtt{id}_1\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, where $\mathtt{id}_1$ is the identity map on L_1.

(c) $\rightsquigarrow$ is a preorder, i.e., it is reflexive and transitive (by (a) and (b)).

(d) $\overset{\leftsquigarrow}{\rightsquigarrow}$ is a congruence relation for $\rightsquigarrow$, i.e., it is an equivalence relation (reflexive, symmetric, transitive) and if $\mathsf{L}_1 \overset{\leftsquigarrow}{\rightsquigarrow} \mathsf{L}_1'$ and $\mathsf{L}_2 \overset{\leftsquigarrow}{\rightsquigarrow} \mathsf{L}_2'$, then $\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ iff $\mathsf{L}_1' \rightsquigarrow \mathsf{L}_2'$.

(e) Translations faithfully preserve equivalence, i.e., if $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, then for all $\phi, \psi \in \mathcal{L}_1$, $\phi \dashv\vdash_1 \psi$ iff $\mathtt{t}(\phi) \dashv\vdash_2 \mathtt{t}(\psi)$.

(f) Translations faithfully preserve validities, i.e., if $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, then for all $\phi \in \mathcal{L}_1$, $\vdash_1 \phi$ iff $\vdash_2 \mathtt{t}(\phi)$.

(g) If $\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, then L_2 is compact only if L_1 is compact.

The existence of quasitranslations is generally not very informative. Almost any logic can be quasitranslated into almost any other logic.

Proposition 2.1.3 *(Quasitranslatability is Trivial).* L_1 is quasitranslatable into L_2 iff it is not the case that L_1 has validities while L_2 does not.

Proof: Suppose first that L_1 has validities and L_2 does not. Let $\phi \in \mathcal{L}_1$ be such that $\vdash_1 \phi$. Then if there were a quasitranslation $\mathtt{t}$ from L_1 to L_2, then $\vdash_2 \mathtt{t}(\phi)$, $\lightning$. So L_1 is not quasitranslatable into L_2.

Next, suppose it is not the case that L_1 has validities while L_2 does not.

[1] See, e.g., Epstein 1990; Carnielli and D'Ottaviano 1997; Feitosa and D'Ottaviano 2001; Pelletier and Urquhart 2003; Humberstone 2005; Caleiro and Gonçalves 2007; Straßburger 2007; Carnielli et al. 2009; Mossakowski et al. 2009; French 2010; Turner 2011b,a; Jeřábek 2012; Barrett and Halvorson 2016a,b; Wigglesworth 2017; Woods 2018. Our terminology here parts ways from the standard conventions: many authors use the term 'translation' for what we call quasitranslations, and use the term 'conservative translation' for what we call translations. Some authors, e.g., Humberstone [2005] and French [2010, Chp. 2] use the term "faithful" instead. Pelletier and Urquhart [2003, p. 266] use "exact" for "conservative", referring to all translations (conservative or not) as "sound" translation schemes.

That is, either there is no $\phi \in \mathcal{L}_1$ such that $\vdash_1 \phi$ or there is a $\psi \in \mathcal{L}_2$ such that $\vdash_2 \psi$. There are two cases to consider:

(i) L_2 **has some validities**. Let $\psi \in \mathcal{L}_2$ be such that $\vdash_2 \psi$. Define $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ such that $\mathtt{t}(\phi) = \psi$ for any $\phi \in \mathcal{L}_1$. Then since $\vdash_2 \mathtt{t}(\phi)$ for any $\phi \in \mathcal{L}_1$, $\mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi)$ for any $\Gamma \subseteq \mathcal{L}_1$. So $\mathtt{t}$ is a quasi-translation vacuously.

(ii) L_1 **and** L_2 **have no validities**. Pick an arbitrary $\psi \in \mathcal{L}_2$ and set $\mathtt{t}(\phi) = \psi$ for all $\phi \in \mathcal{L}_1$. Then if $\Gamma \vdash_1 \phi$, then $\Gamma \neq \varnothing$, so $\mathtt{t}[\Gamma] = \{\psi\}$. So trivially, $\mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi) = \psi$. Hence, $\mathtt{t}$ is a quasitranslation. ■

Hence, quasitranslatability is an uninformative notion. As a result, only very occasionally will we discuss quasitranslations.

Sometimes, logicians focus on validity-preserving maps, rather than full translations.

Definition 2.1.4 *(Validity-Preserving Map)*. Let L_1 and L_2 be some logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. We say $\mathtt{t}$ is a ***validity-preserving map*** if for all $\phi \in \mathcal{L}_1$:

$$\vdash_1 \phi \quad \Leftrightarrow \quad \vdash_2 \mathtt{t}(\phi).$$

The existence of validity-preserving maps, however, is also uninformative. Generally, one can get a validity-preserving map just by mapping all validities to a single validity and all non-validities to a single non-validity.

Fact 2.1.5 *(Triviality of Validity-Preservation)*. Let L_1 and L_2 be logics.

(a) If there are $\psi_1, \psi_2 \in \mathcal{L}_2$ such that $\vdash_2 \psi_1$ but $\nvdash_2 \psi_2$, then there is a validity-preserving map from L_1 to L_2.

(b) If $\vdash_2 \psi$ for all $\psi \in \mathcal{L}_2$, then there is a validity-preserving map from L_1 to L_2 iff $\vdash_1 \phi$ for all $\phi \in \mathcal{L}_1$.

(c) If $\nvdash_2 \psi$ for all $\psi \in \mathcal{L}_2$, then there is a validity-preserving map from L_1 to L_2 iff $\nvdash_2 \phi$ for all $\phi \in \mathcal{L}_1$.

Of course, validity-preservation becomes more interesting when we place syntactic constraints on the map in question. For instance, it is non-trivial to show that there exists a validity-preserving map from, say, ZFC into ZF that commutes with the booleans and maps quantifiers to bounded quantifiers. But since our focus at the moment is on languages without a specified syntactic structure, we set aside mere validity-preservation for now.

§ 2.1.1 Examples

We now examine some classic examples of translations to illustrate its ubiquity. These examples should be quite familiar, but we lay them out in full in order to introduce notation we use throughout.

Example 2.1.6 *(Sheffer Stroke)*. Let $\mathtt{Prop} = \{p_1, p_2, p_3, \ldots\}$ be a set of atomics. We define **CPL**, i.e., the language of classical propositional logic, thereby defining its corresponding logic CPL. First, let $\mathcal{L}_{\mathtt{Prop}}$ be the set of formulas defined recursively as follows (where $p \in \mathtt{Prop}$):

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

Next, let $\mathsf{V} := \mathsf{C}_{\mathbf{CPL}}$ be the set of valuation functions $v\colon \mathtt{Prop} \to \{0, 1\}$. Last, define $\Vdash$ recursively as follows:

$$\begin{aligned} v \Vdash p \quad &\Leftrightarrow \quad v(p) = 1 \\ v \Vdash \neg\phi \quad &\Leftrightarrow \quad v \nVdash \phi \\ v \Vdash \phi \wedge \psi \quad &\Leftrightarrow \quad v \Vdash \phi \text{ and } v \Vdash \psi. \end{aligned}$$

Then $\mathbf{CPL} := \langle \mathcal{L}_{\mathtt{Prop}}, \mathsf{V}, \Vdash \rangle$.

Now define $\mathbf{CPL}^{\uparrow}$, i.e., (the language of) classical propositional logic with the Sheffer stroke analogously. That is, $\mathcal{L}^{\uparrow}_{\mathtt{Prop}}$ is recursively defined as follows:

$$\phi ::= p \mid (\phi \uparrow \phi).$$

The evaluation space is still V. The satisfaction relation $\Vdash^{\uparrow}$ is defined so that:

$$\begin{aligned} v \Vdash^{\uparrow} p \quad &\Leftrightarrow \quad v(p) = 1 \\ v \Vdash^{\uparrow} \phi \uparrow \psi \quad &\Leftrightarrow \quad v \nVdash^{\uparrow} \phi \text{ or } v \nVdash^{\uparrow} \psi. \end{aligned}$$

Then $\mathbf{CPL}^{\uparrow} := \langle \mathcal{L}^{\uparrow}_{\mathtt{Prop}}, \mathsf{V}, \Vdash^{\uparrow} \rangle$. It is easy to show **CPL** and $\mathbf{CPL}^{\uparrow}$ are intertranslatable. First, define $\mathtt{t}\colon \mathcal{L}^{\uparrow}_{\mathtt{Prop}} \to \mathcal{L}_{\mathtt{Prop}}$ so that $\mathtt{t}(p) = p$ and $\mathtt{t}(\phi \uparrow \psi) = \neg(\mathtt{t}(\phi) \wedge \mathtt{t}(\psi))$. Second, define $\mathtt{s}\colon \mathcal{L}_{\mathtt{Prop}} \to \mathcal{L}^{\uparrow}_{\mathtt{Prop}}$ so that $\mathtt{s}(p) = p$, $\mathtt{s}(\neg\phi) = (\mathtt{s}(\phi) \uparrow \mathtt{s}(\phi))$, and $\mathtt{s}(\phi \wedge \psi) = ((\mathtt{s}(\phi) \uparrow \mathtt{s}(\psi)) \uparrow (\mathtt{s}(\phi) \uparrow \mathtt{s}(\psi)))$. Then $\mathtt{t}, \mathtt{s}\colon \mathbf{CPL}^{\uparrow} \leftrightarrows \mathbf{CPL}$ (in fact, $\mathtt{t}, \mathtt{s}\colon \mathbf{CPL}^{\uparrow} \leftrightsquigarrow \mathbf{CPL}$; see Definition 2.2.1).

Example 2.1.7 *(The Standard Translation for Modal Logic)*. We define **K**, i.e., (the language of) minimal normal modal logic, as follows. The syntax $\mathcal{L}_{\mathtt{Prop}}(\Box)$

is defined recursively as follows (where again $p \in \texttt{Prop}$):

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

A ***modal model*** is a triple $\mathcal{M} = \langle W, \twoheadrightarrow, V\rangle$ such that W is a nonempty set, $\twoheadrightarrow \subseteq W \times W$, and $V\colon \texttt{Prop} \to \wp(W)$. The evaluation space $\mathsf{C}_{\mathbf{K}}$ will be the class of pairs $\langle \mathcal{M}, w\rangle$ such that $\mathcal{M} = \langle W, \twoheadrightarrow, V\rangle$ is a modal model and $w \in W$. We define $\Vdash_{\mathbf{K}}$ as follows:

$$\begin{array}{lcl}
\mathcal{M}, w \Vdash_{\mathbf{K}} p & \Leftrightarrow & w \in V(p) \\
\mathcal{M}, w \Vdash_{\mathbf{K}} \neg\phi & \Leftrightarrow & \mathcal{M}, w \nVdash_{\mathbf{K}} \phi \\
\mathcal{M}, w \Vdash_{\mathbf{K}} \phi \wedge \psi & \Leftrightarrow & \mathcal{M}, w \Vdash \phi \text{ and } \mathcal{M}, w \Vdash \psi \\
\mathcal{M}, w \Vdash_{\mathbf{K}} \Box\phi & \Leftrightarrow & \text{for all } v \in W \text{ such that } w \twoheadrightarrow v\text{: } \mathcal{M}, v \Vdash \phi.
\end{array}$$

Next, we define **FOL**, i.e., the language of classical first-order logic (FOL). Let $\texttt{Var} = \{x_1, x_2, x_3, \ldots\}$, let $\texttt{Pred}^n = \{P^n_1, P^n_2, P^n_3, \ldots\}$ for each $n \in \mathbb{N}$, and let $\texttt{Pred} = \bigcup_{n\in\mathbb{N}} \texttt{Pred}^n$. (We single out $\twoheadrightarrow \in \texttt{Pred}^2$ as a special binary predicate.) The syntax $\mathcal{L}_{\texttt{Pred}}$ is defined recursively as follows (where $P^n_i \in \texttt{Pred}^n$ and $x, y, y_1, \ldots, y_n \in \texttt{Var}$):

$$\phi ::= P^n_i(y_1, \ldots, y_n) \mid (x \twoheadrightarrow y) \mid \neg\phi \mid (\phi \wedge \phi) \mid \forall x\, \phi.$$

A ***(first-order) model*** is a pair $\mathcal{M} = \langle D, I\rangle$ where D is a nonempty set and $I\colon \texttt{Pred}^n \to \wp(D^n)$ for each $n \in \mathbb{N}$. A ***variable assignment*** on $\mathcal{M}$ is a map $g\colon \texttt{Var} \to D$. The evaluation space $\mathsf{C}_{\mathbf{FOL}}$ will be the class of pairs of the form $\langle \mathcal{M}, g\rangle$ where $\mathcal{M}$ is a model and g is a variable assignment on $\mathcal{M}$. We define $\Vdash_{\mathbf{FOL}}$ as follows:

$$\begin{array}{lcl}
\mathcal{M}, g \Vdash_{\mathbf{FOL}} P^n(y_1, \ldots, y_n) & \Leftrightarrow & \langle g(y_1), \ldots, g(y_n)\rangle \in I(P^n) \\
\mathcal{M}, g \Vdash_{\mathbf{FOL}} \neg\phi & \Leftrightarrow & \mathcal{M}, g \nVdash_{\mathbf{FOL}} \phi \\
\mathcal{M}, g \Vdash_{\mathbf{FOL}} \phi \wedge \psi & \Leftrightarrow & \mathcal{M}, g \Vdash_{\mathbf{FOL}} \phi \text{ and } \mathcal{M}, g \Vdash_{\mathbf{FOL}} \psi \\
\mathcal{M}, g \Vdash_{\mathbf{FOL}} \forall x\, \phi & \Leftrightarrow & \text{for all } a \in D\text{: } \mathcal{M}, g^x_a \Vdash_{\mathbf{FOL}} \phi,
\end{array}$$

where $g^x_a(x) = a$ and $g^x_a(y) = g(y)$ for $y \neq x$. We can define a series of translations ST_x from **K** to **FOL** for each $x \in \texttt{Var}$ as follows:

$$\begin{array}{lcl}
\mathsf{ST}_x(p_i) & = & P^1_i(x) \\
\mathsf{ST}_x(\neg\phi) & = & \neg\,\mathsf{ST}_x(\phi) \\
\mathsf{ST}_x(\phi \wedge \psi) & = & (\mathsf{ST}_x(\phi) \wedge \mathsf{ST}_x(\psi)) \\
\mathsf{ST}_x(\Box\phi) & = & \forall y\, ((x \twoheadrightarrow y) \to \mathsf{ST}_y(\phi))
\end{array}$$

where $y \neq x$. Such a translation is often called the ***standard translation*** (see, e.g., Blackburn et al. 2001, p. 84).

Example 2.1.8 *(Intuitionistic Logic)*. Let **IPL** be (the language of) intuitionistic propositional logic (IPL) whose syntax is $\mathcal{L}_{\texttt{Prop}}(\vee, \to)$, defined as follows:

$$\phi ::= p \mid \neg\,\phi \mid (\phi \wedge \phi) \mid (\phi \vee \psi) \mid (\phi \to \phi).$$

The famous ***double-negation translation*** N from **CPL**$(\vee, \to)$ (i.e., **CPL** expanded with syntax $\mathcal{L}_{\texttt{Prop}}(\vee, \to)$ in the obvious way) into **IPL** (also known as the ***Gödel-Gentzen translation***) is defined as follows:

$$\begin{aligned}
\mathtt{N}(p) &= \neg\,\neg\,p \\
\mathtt{N}(\neg\,\phi) &= \neg\,\mathtt{N}(\phi) \\
\mathtt{N}(\phi \wedge \psi) &= (\mathtt{N}(\phi) \wedge \mathtt{N}(\psi)) \\
\mathtt{N}(\phi \vee \psi) &= \neg(\neg\,\mathtt{N}(\phi) \wedge \neg\,\mathtt{N}(\psi)) \\
\mathtt{N}(\phi \to \psi) &= (\mathtt{N}(\phi) \to \mathtt{N}(\psi)).
\end{aligned}$$

Let **S4** be the restriction of **K** to the class of pointed models $\langle \mathcal{M}, w \rangle$ where $\twoheadrightarrow$ is reflexive and transitive. The ***Gödel translation*** G from **IPL** into **S4** (also known as the ***Gödel-McKinsey-Tarski translation***) is defined as follows:

$$\begin{aligned}
\mathtt{G}(p) &= \Box p \\
\mathtt{G}(\neg\,\phi) &= \Box\,\neg\,\mathtt{G}(\phi) \\
\mathtt{G}(\phi \wedge \psi) &= (\mathtt{G}(\phi) \wedge \mathtt{G}(\psi)) \\
\mathtt{G}(\phi \vee \psi) &= (\mathtt{G}(\phi) \vee \mathtt{G}(\psi)) \\
\mathtt{G}(\phi \to \psi) &= \Box(\mathtt{G}(\phi) \to \mathtt{G}(\psi)).
\end{aligned}$$

This is sometimes presented just as a validity-preserving map; but by the T-axiom and necessitation, it is also a translation in the sense of **Definition 2.1.1**.

The notion of a translation is fairly minimal. One might worry that without imposing any extra constraints, any logic can be translated into any other logic. As we will see, this worry is completely legitimate (**Theorem 2.6.8**). However, not everything goes: as these next examples show, there are non-trivial cases of failure of translatability.

Example 2.1.9 *(Propositional Logic with Finitely Many Atomics)*. Where $A \subseteq \texttt{At}$, let **CPL**^{A} be the fragment of **CPL** obtained by restricting to the $\mathcal{L}_{\texttt{Prop}}$-formulas whose atomics are all among A. There is no translation from **CPL** into **CPL**$^{\{p_1,\ldots,p_n\}}$. It is well-known that **CPL** is ***locally finite***, meaning there are only finitely many (2^{2^n}) $\mathcal{L}_{\texttt{Prop}}^{\{p_1,\ldots,p_n\}}$-formulas up to equivalence. But there are infinitely many $\mathcal{L}_{\texttt{Prop}}$-formulas up to equivalence. So any mapping from

$\mathcal{L}_{\texttt{Prop}}$ into $\mathcal{L}_{\texttt{Prop}}^{\{p_1,\dots,p_n\}}$ must eventually map some non-equivalent formulas in $\mathcal{L}_{\texttt{Prop}}$ to equivalent formulas in $\mathcal{L}_{\texttt{Prop}}^{\{p_1,\dots,p_n\}}$. Hence, $\mathbf{CPL} \not\rightsquigarrow \mathbf{CPL}^{\{p_1,\dots,p_n\}}$.

Example 2.1.10 *(Infinitary Propositional Logic)*. Let $\mathbf{CPL}(\bigwedge)$ be the extension of **CPL** containing countably infinite conjunctions, so that for any $v \in \mathsf{V}$ and any countable set of $\mathbf{CPL}(\bigwedge)$-formulas Φ:

$$v \Vdash \bigwedge \Phi \quad \Leftrightarrow \quad \text{for all } \phi \in \Phi\text{: } v \Vdash \phi.$$

Then by the same reasoning as that used in Example 2.1.9, $\mathbf{CPL}(\bigwedge) \not\rightsquigarrow \mathbf{CPL}$, since $\mathbf{CPL}(\bigwedge)$ has $2^{2^{\aleph_0}}$-many formulas up to equivalence, whereas **CPL** only has $\aleph_0$-many-formulas in total.

Example 2.1.11 *(**S5**)*. Let **S5** be the restriction of **K** to modal models $\mathcal{M}$ where $\twoheadrightarrow$ is an equivalence relation. Where $A \subseteq \texttt{At}$, define $\mathbf{S5}^A$ analogously to how $\mathbf{CPL}^A$ was defined.

One can verify that there are:

$$2^{2^n \cdot \sum_{k=0}^{2^n-1} \binom{2^n-1}{k}} = 2^{2^{2^n+n-1}}$$

$\mathcal{L}_{\texttt{Prop}}^{\{p_1,\dots,p_n\}}(\Box)$-formulas up to equivalence in $\mathbf{S5}^{\{p_1,\dots,p_n\}}$. This is much larger than the number of formulas up to equivalence in $\mathbf{CPL}^{\{p_1,\dots,p_n\}}$ (viz., 2^{2^n}). Hence, $\mathbf{S5}^{\{p_1,\dots,p_n\}} \not\rightsquigarrow \mathbf{CPL}^{\{p_1,\dots,p_n\}}$.

Similarly, let $\mathbf{S5}(\bigwedge)$ be obtained by adding infinitary conjunction to **S5**. There $\mathbf{S5}(\bigwedge) \not\rightsquigarrow \mathbf{CPL}(\bigwedge)$, since there are $2^{2^{2^{\aleph_0}}}$-many formulas up to equivalence in $\mathbf{S5}(\bigwedge)$, while there are only $2^{2^{\aleph_0}}$-many formulas up to equivalence in $\mathbf{CPL}(\bigwedge)$.

However, it is possible to strike a balance. Given that $m \geqslant 2^n + n - 1$, there will be a translation of $\mathbf{S5}^{\{p_1,\dots,p_n\}}$ into $\mathbf{CPL}^{\{p_1,\dots,p_m\}}$, and in the case where $m = 2^n + n - 1$, the translation will go in the other direction too. Moreover, as we explain below in Theorems 2.6.5 and 2.6.8, there is also a translation from **S5** into **CPL**.

Examples 2.1.9–2.1.11 result from a more general counting observation about translations, viz., the translation of a logic must always contain at least as many formulas up to equivalence as the logic itself (see Corollary 2.6.2). The next examples illustrate alternative ways of proving the non-existence of translations.

Example 2.1.12 *(Logic of Paradox)*. We define **LP**, i.e., the logic of paradox, as follows. The syntax is just $\mathcal{L}_{\texttt{Prop}}$. The evaluation space is the set GV ("glutty" evaluations) of all functions $v\colon \texttt{Prop} \to \{0, \frac{1}{2}, 1\}$. We define the relations $\Vdash$ and $\dashV$ as follows:

$$
\begin{array}{llllll}
v \Vdash p & \Leftrightarrow & v(p) \geqslant \dfrac{1}{2} & v \dashV p & \Leftrightarrow & v(p) \leqslant \dfrac{1}{2} \\
v \Vdash \neg\phi & \Leftrightarrow & v \dashV \phi & v \dashV \neg\phi & \Leftrightarrow & v \Vdash \phi \\
v \Vdash \phi \wedge \psi & \Leftrightarrow & v \Vdash \phi \text{ and } v \Vdash \psi & v \dashV \phi \wedge \psi & \Leftrightarrow & v \dashV \phi \text{ or } v \dashV \psi.
\end{array}
$$

Then $\mathbf{LP} := \langle \mathcal{L}_{\texttt{Prop}}, \mathsf{GV}, \Vdash \rangle$.

Jeřábek [2012, p. 672] showed that there is no translation from **CPL** to **LP**. The proof is as follows. Suppose for *reductio* that $\texttt{t}\colon \mathbf{CPL} \rightsquigarrow \mathbf{LP}$. Let $v_1, \ldots, v_n$ be all the glutty evaluations such that $v(\texttt{t}(\bot)) = 0$ and $v(p) = \frac{1}{2}$ for any atomic p not in $\texttt{t}(\bot)$. Let $\phi_i = p_i$ for $1 \leqslant i \leqslant n$, and let $\phi_0 = \neg \bigwedge_{i=1}^{n} p_i$. Clearly, $\phi_0, \ldots, \phi_n \vDash_{\mathbf{CPL}} \bot$. So $\texttt{t}(\phi_0), \ldots, \texttt{t}(\phi_n) \vDash_{\mathbf{LP}} \texttt{t}(\bot)$. Thus, for each v_i, there must be a $0 \leqslant j_i \leqslant n$ such that $v_i(\phi_{j_i}) = 0$. But then $\{\texttt{t}(\phi_{j_i}) \mid 1 \leqslant i \leqslant n\} \vDash_{\mathbf{CPL}} \texttt{t}(\bot)$, since $v_1, \ldots, v_n$ are all the potential counterexamples (up to equivalence on the atomics in $\texttt{t}(\bot)$) to this inference and each assigns 0 one of the premises. Hence, $\{\phi_{j_i} \mid 1 \leqslant i \leqslant n\} \vDash_{\mathbf{CPL}} \bot$ already; but this cannot be since there must be at least one ϕ_k missing from $\{\phi_{j_i} \mid 1 \leqslant i \leqslant n\}$, and we need all $\phi_0, \ldots, \phi_n$ to derive a contradiction, $\lightning$. So $\mathbf{CPL} \not\rightsquigarrow \mathbf{LP}$. We do, however, have $\mathbf{LP} \rightsquigarrow \mathbf{CPL}$ (**Theorem 2.6.8**).

Example 2.1.13 *(Second-Order Logic)*. We define **SOL**, i.e., classical second-order logic, over the set of formulas generated recursively as follows:

$$\phi ::= P_i^n(y_1, \ldots, y_n) \mid X^n(y_1, \ldots, y_n) \mid \neg\phi \mid (\phi \wedge \phi) \mid \forall x\, \phi \mid \forall X^n\, \phi.$$

The standard semantics for **SOL** can be defined over first-order models but variable assignments on $\mathcal{M} = \langle D, I \rangle$ now also map $g(X^n) \subseteq D^n$. We now just add the following semantic clauses:

$$
\begin{array}{lll}
\mathcal{M}, g \Vdash_{\mathbf{SOL}} X^n(y_1, \ldots, y_n) & \Leftrightarrow & \langle g(y_1), \ldots, g(y_n) \rangle \in g(X^n) \\
\mathcal{M}, g \Vdash_{\mathbf{SOL}} \forall X^n\, \phi & \Leftrightarrow & \text{for all } A^n \subseteq D^n\text{: } \mathcal{M}, g^{X^n}_{A^n} \Vdash_{\mathbf{SOL}} \phi.
\end{array}
$$

Since **SOL** is not compact while **FOL** is compact, $\mathbf{SOL} \not\rightsquigarrow \mathbf{FOL}$ (**Fact 2.1.2**).

Epstein [1990, p. 388] asked whether there was an example of two logics such that neither is translatable into the other. The answer is affirmative. We can generate a whole host of counterexamples instantly using the following method. Let $\langle P_1, \leqslant_1 \rangle$ and $\langle P_2, \leqslant_2 \rangle$ be any partial orders such that neither is order-embeddable into the other. Then we can define a logic whose formulas are members of P_i and where $\Gamma \vDash_i \phi$ just in case there is a $\psi \in \Gamma$ such that $\psi \leqslant_i \phi$ (recall that this is the strategy used to prove **Proposition 1.1.40**). Call the resulting logics $\mathbf{L}_1$ and $\mathbf{L}_2$. It is easy to check that $\mathbf{L}_1 \not\rightsquigarrow \mathbf{L}_2$ and $\mathbf{L}_2 \not\rightsquigarrow \mathbf{L}_1$. There are, however, more natural examples, if we charitably interpret Epstein as seeking a natural pair of logics.

Example 2.1.14 *(Kleene Logic)*. We define **K3**, i.e., the language of (strong) Kleene 3-valued logic, as follows. The syntax is just $\mathcal{L}_{\mathtt{Prop}}$. The evaluation space is the set PV of partial functions $v\colon \mathtt{Prop} \to \{0,1\}$. We define the relations $\Vdash$ and $\dashV$ as follows:

$$
\begin{array}{lll@{\qquad}lll}
v \Vdash p & \Leftrightarrow & v(p) = 1 & v \dashV p & \Leftrightarrow & v(p) = 0 \\
v \Vdash \neg\phi & \Leftrightarrow & v \dashV \phi & v \dashV \neg\phi & \Leftrightarrow & v \Vdash \phi \\
v \Vdash \phi \wedge \psi & \Leftrightarrow & v \Vdash \phi \text{ and } v \Vdash \psi & v \dashV \phi \wedge \psi & \Leftrightarrow & v \dashV \phi \text{ or } v \dashV \psi.
\end{array}
$$

Then $\mathbf{K3} := \langle \mathcal{L}_{\mathtt{Prop}}, \mathsf{PV}, \Vdash \rangle$.

Observe first that there is no ϕ such that $\vDash_{\mathbf{K3}} \phi$, since if $v = \varnothing$, then $v \nVdash \phi$. So $\mathbf{CPL} \not\rightsquigarrow \mathbf{K3}$ and $\mathbf{CPL}^{\{p_1,\dots,p_n\}} \not\rightsquigarrow \mathbf{K3}^{\{p_1,\dots,p_n\}}$. Conversely, observe that if $\phi \equiv_{\mathbf{K3}} \psi$, then $\phi \equiv_{\mathbf{CPL}} \psi$. Hence, there are at least as many formulas up to equivalence in $\mathbf{K3}^{\{p_1,\dots,p_n\}}$ as there are in $\mathbf{CPL}^{\{p_1,\dots,p_n\}}$. In fact, there are strictly more. Consider $p \wedge \neg(q \wedge \neg q)$ where $p \neq q$. Since $p \equiv_{\mathbf{CPL}} p \wedge \neg(q \wedge \neg q)$, $[p]_{\mathbf{CPL}} = [p \wedge \neg(q \wedge \neg q)]_{\mathbf{CPL}}$. But $p \not\equiv_{\mathbf{K3}} p \wedge \neg(q \wedge \neg q)$, so $[p]_{\mathbf{K3}} \neq [p \wedge \neg(q \wedge \neg q)]_{\mathbf{K3}}$. So there are more formulas up to equivalence in $\mathbf{K3}^{\{p_1,\dots,p_n\}}$ than $\mathbf{CPL}^{\{p_1,\dots,p_n\}}$. Hence, $\mathbf{K3}^{\{p_1,\dots,p_n\}} \not\rightsquigarrow \mathbf{CPL}^{\{p_1,\dots,p_n\}}$. However, we do have $\mathbf{K3} \rightsquigarrow \mathbf{CPL}$ (see **Theorems 2.6.5** and **2.6.8**). What is more, Jeřábek [2012, Example 3.3] showed that if we add a verum operator $\top$ such that $v \Vdash \top$ and $v \not\dashV \top$ for all $v \in \mathsf{PV}$, then $\mathbf{CPL} \rightsquigarrow \mathbf{K3}$; likewise, we can show in that case that $\mathbf{CPL}^n \rightsquigarrow \mathbf{K3}^n$ for all n.

Similar results show that neither **LP** nor **K3** are translatable into one another. Since **LP** has validities, $\mathbf{LP} \not\rightsquigarrow \mathbf{K3}$. Moreover, the argument in **Example 2.1.12** showing $\mathbf{CPL} \not\rightsquigarrow \mathbf{LP}$ straightforwardly generalizes to a proof that $\mathbf{K3} \not\rightsquigarrow \mathbf{LP}$. However, there is a sense in which **K3** and **LP** are "duals" of one another. More precisely, one can show that $\phi \vDash_{\mathbf{K3}} \psi$ iff $\neg\psi \vDash_{\mathbf{LP}} \neg\phi$.

§ 2.1.2 Theoretic Formulation

We can define translatability in terms of theory spaces (recall the definition of a theory space, **Definition 1.1.33**).

Notation: Where $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$, we define the map $\mathtt{t}^{\mathrm{Th}}\colon \wp(\mathcal{L}_1) \to \wp(\mathcal{L}_2)$ so that $\mathtt{t}^{\mathrm{Th}}(\Gamma) = \mathsf{Cn}_2(\mathtt{t}[\Gamma])$.

Notation: Let L be a logic. Where $\Sigma \subseteq \wp(\mathcal{L})$, we define $\bigwedge \Sigma := \mathtt{Cn}_{\mathsf{L}}[\bigcup \Sigma]$. Note that this notation is consistent with the special case where $\Sigma \subseteq \mathtt{Th}(\mathsf{L})$.

Lemma 2.1.15 *(Theory Maps Lifted from Translations Preserve Meets)*. Let L_1 and L_2 be logics. If $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, then for all $\Sigma \subseteq \wp(\mathcal{L})$:

$$\mathtt{t}^{\mathtt{Th}}(\bigwedge \Sigma) \quad = \quad \bigwedge \mathtt{t}^{\mathtt{Th}}[\Sigma].$$

Proof: Let $\Sigma \subseteq \wp(\mathcal{L}_1)$. First:

$$\begin{aligned}
\mathtt{t}^{\mathtt{Th}}(\bigwedge \Sigma) &= \mathtt{t}^{\mathtt{Th}}(\mathtt{Cn}_1(\bigcup_{\Gamma\in\Sigma} \Gamma)) \\
&= \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\bigcup_{\Gamma\in\Sigma} \Gamma)]) \\
&= \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\bigcup \Sigma)]).
\end{aligned}$$

Second, observe that $\mathtt{Cn}(\bigcup_{\Gamma\in\Sigma} \mathtt{Cn}(\Gamma)) = \mathtt{Cn}(\bigcup_{\Gamma\in\Sigma} \Gamma)$. Thus:

$$\begin{aligned}
\bigwedge \mathtt{t}^{\mathtt{Th}}[\Sigma] &= \mathtt{Cn}_2(\bigcup_{\Gamma\in\Sigma} \mathtt{t}^{\mathtt{Th}}(\Gamma)) \\
&= \mathtt{Cn}_2(\bigcup_{\Gamma\in\Sigma} \mathtt{Cn}_2(\mathtt{t}[\Gamma])) \\
&= \mathtt{Cn}_2(\bigcup_{\Gamma\in\Sigma} \mathtt{t}[\Gamma]) \\
&= \mathtt{Cn}_2(\mathtt{t}[\bigcup \Sigma]).
\end{aligned}$$

Hence, it suffices to show that:

$$\mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\bigcup \Sigma)]) \quad = \quad \mathtt{Cn}_2(\mathtt{t}[\bigcup \Sigma]).$$

In other words, it suffices to show that $\mathtt{t}[\bigcup \Sigma] \dashv\vdash_2 \mathtt{t}[\mathtt{Cn}_1(\bigcup \Sigma)]$. But this follows from the fact that $\mathtt{t}$ is a translation, since $\bigcup \Sigma \dashv\vdash_1 \mathtt{Cn}_1(\bigcup \Sigma)$. ■

Proposition 2.1.16 *(Theoretic Formulation of Translation)*. Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ iff $\mathtt{t}^{\mathtt{Th}}$ is an order-embedding from $\mathbb{T}_1$ to $\mathbb{T}_2$ such that $\mathtt{t}^{\mathtt{Th}}(\bigwedge \Sigma) = \bigwedge \mathtt{t}^{\mathtt{Th}}[\Sigma]$ for all $\Sigma \subseteq \wp(\mathcal{L}_1)$.

Proof: The left-to-right direction follows from **Lemma 2.1.15** (verifying that $\mathtt{t}^{\mathtt{Th}}$ is an order-embedding on theory spaces is straightforward). For the right-to-left direction, suppose $\mathtt{t}^{\mathtt{Th}}$ is an order-embedding from $\mathbb{T}_1$ to $\mathbb{T}_2$ and that $\mathtt{t}^{\mathtt{Th}}(\bigwedge \Sigma) = \bigwedge \mathtt{t}^{\mathtt{Th}}[\Sigma]$ for all $\Sigma \subseteq \wp(\mathcal{L}_1)$. So:

$$
\begin{aligned}
\mathtt{Cn}_2(\mathtt{t}[\Gamma]) &= \mathtt{t}^{\mathtt{Th}}(\Gamma) \\
&= \mathtt{t}^{\mathtt{Th}}(\Gamma \wedge \mathtt{Cn}_1(\Gamma)) \\
&= \mathtt{t}^{\mathtt{Th}}(\Gamma) \wedge \mathtt{t}^{\mathtt{Th}}(\mathtt{Cn}_1(\Gamma)) \\
&= \mathtt{Cn}_2(\mathtt{Cn}_2(\mathtt{t}[\Gamma]) \cup \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\Gamma)])) \\
&= \mathtt{Cn}_2(\mathtt{t}[\Gamma] \cup \mathtt{t}[\mathtt{Cn}_1(\Gamma)]) \\
&= \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\Gamma)).
\end{aligned}
$$

Thus, $\mathtt{t}[\Gamma] \dashv\vdash_2 \mathtt{t}[\mathtt{Cn}_1(\Gamma)]$. Likewise, $\mathtt{t}(\phi) \dashv\vdash_2 \mathtt{t}[\mathtt{Cn}_1(\phi)]$. Hence:

$$
\begin{aligned}
\Gamma \vdash_1 \phi &\Leftrightarrow \mathtt{Cn}_1(\phi) \subseteq \mathtt{Cn}_1(\Gamma) \\
&\Leftrightarrow \mathtt{t}^{\mathtt{Th}}(\mathtt{Cn}_1(\phi)) \subseteq \mathtt{t}^{\mathtt{Th}}(\mathtt{Cn}_1(\Gamma)) \\
&\Leftrightarrow \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\phi)]) \subseteq \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\Gamma)]) \\
&\Leftrightarrow \mathtt{t}[\mathtt{Cn}_1(\Gamma)] \vdash_2 \mathtt{t}[\mathtt{Cn}_1(\phi)] \\
&\Leftrightarrow \mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi).
\end{aligned}
$$

■

$\mathtt{Cn}_2(\mathtt{t}[\Gamma]) = \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\Gamma)])$ is equivalent to the condition that $\mathtt{t}$ is a quasitranslation. If this holds, and if $\Gamma \vdash_1 \phi$, then $\phi \in \mathtt{Cn}_1(\Gamma)$, and so $\mathtt{t}(\phi) \in \mathtt{t}[\mathtt{Cn}_1(\Gamma)]$. But then $\mathtt{t}(\phi) \in \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\Gamma)]) = \mathtt{Cn}_2(\mathtt{t}[\Gamma])$, i.e., $\mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi)$. Conversely, if $\mathtt{t}$ is a quasitranslation, then since $\Gamma \vdash_1 \mathtt{Cn}_1(\Gamma)$, $\mathtt{t}[\Gamma] \vdash_1 \mathtt{t}[\mathtt{Cn}_1(\Gamma)]$, so $\mathtt{Cn}_2(\mathtt{t}[\Gamma]) = \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\Gamma)])$. A brief examination of **Lemma 2.1.15** reveals that this condition is equivalent to $\mathtt{t}^{\mathtt{Th}}$ preserving meets.[2] However, this property alone does not guarantee that $\mathtt{t}$ is a full translation; to carry out the above reasoning backwards, we would need to infer from the fact that $\mathtt{t}(\phi) \in \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\Gamma)])$ to the claim that $\mathtt{t}(\phi) \in \mathtt{t}[\mathtt{Cn}_1(\Gamma)]$, which is not valid if $\mathtt{t}$ is a merely quasitranslation. So the fact that $\mathtt{t}^{\mathtt{Th}}$ preserves meets is what guarantees that $\mathtt{t}$ is a quasitranslation and the fact that it is an order-embedding on theory spaces is what guarantees that it is a full translation.

[2]Note that by "preserving meets", we do not mean just preserving meets over the theory space. Rather, we mean the more general property of preserving meets over sets of formulas generally.

§ 2.2 Translational Equivalence

Intertranslatable languages are only equivalent in a fairly weak sense. Imagine two speakers a_1 and a_2 of some intertranslatable languages $\mathbf{L}_1$ and $\mathbf{L}_2$ respectively. Let $\mathtt{t}_1$ be a translation from $\mathbf{L}_1$ to $\mathbf{L}_2$ and let $\mathtt{t}_2$ be a translation from $\mathbf{L}_2$ to $\mathbf{L}_1$. Suppose a_1 asserts ϕ and a_2 uses the translation $\mathtt{t}_1$ to infer that whatever a_1 asserted is equivalent to $\mathtt{t}_1(\phi)$. If a_2 then wants to assert what a_1 said in his own language, it seems natural for him to assert $\mathtt{t}_1(\phi)$. If a_1 hears a_2 assert $\mathtt{t}_1(\phi)$, then using $\mathtt{t}_2$, she will infer that what a_2 said is equivalent to $\mathtt{t}_2(\mathtt{t}_1(\phi))$.

But the mere fact that $\mathtt{t}_1$ and $\mathtt{t}_2$ are translations between $\mathbf{L}_1$ and $\mathbf{L}_2$ is not enough to guarantee that a_1 and a_2 can successfully understand each other since $\mathtt{t}_2(\mathtt{t}_1(\phi))$ might *not* be equivalent to what a_1 asserted in the first place, viz., ϕ. From a_1's perspective, if $\mathtt{t}_2(\mathtt{t}_1(\phi))$ is not equivalent to ϕ, then a_1 could interpret a_2 as having misunderstood her, or possibly of having made some logical misstep. Intertranslatability is not, by itself, strong enough to ensure that this does not happen.

What this simple thought experiment illustrates is that the existence of translations that are individually adequate is not sufficient to call two languages or logics "equivalent". Rather, we also need these translations to "agree" with each other. This motivates the following definition:[3]

Definition 2.2.1 *(Translational Equivalence)*. Let L_1 and L_2 be logics. A translation scheme $\mathtt{t}_1, \mathtt{t}_2 \colon \mathsf{L}_1 \overset{\leftsquigarrow}{\rightsquigarrow} \mathsf{L}_2$ is ***reversible*** (written "$\mathtt{t}_1, \mathtt{t}_2 \colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$") if:

(a) for all $\phi \in \mathcal{L}_1$: $\phi \dashv\vdash_1 \mathtt{t}_2(\mathtt{t}_1(\phi))$

(b) for all $\psi \in \mathcal{L}_2$: $\psi \dashv\vdash_2 \mathtt{t}_1(\mathtt{t}_2(\psi))$.

We say L_1 is ***translationally equivalent*** to L_2 (written "$\mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$") if there is a reversible translation scheme $\mathtt{t}_1, \mathtt{t}_2 \colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$. Similarly, we say L_1 and L_2 are ***quasitranslationally equivalent*** if there are quasitranslations $\mathtt{t}$ and $\mathtt{s}$ from one to the other satisfying (a) and (b) above. These notions are lifted from logics to languages in the obvious way.

Lemma 2.2.2 *(Translational Equivalence is an Equivalence Relation)*. $\leftrightsquigarrow$ is reflexive, symmetric, and transitive.

Proof: Reflexivity and symmetry are obvious. For transitivity, suppose that $\mathtt{t}, \mathtt{s} \colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$ and $\mathtt{t}', \mathtt{s}' \colon \mathsf{L}_2 \leftrightsquigarrow \mathsf{L}_3$. Then $\mathtt{t}' \circ \mathtt{t}, \mathtt{s} \circ \mathtt{s}' \colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_3$. ■

[3] This notion is discussed in Pelletier and Urquhart 2003; Caleiro and Gonçalves 2007; Straßburger 2007; French 2010; Turner 2011b,a; Woods 2018. Caleiro and Gonçalves [2007] use the term "equipollent" instead of "translationally equivalent", while Turner [2011b, p. 437] uses the term "recoverable translation" instead of "reversible translation scheme".

While quasitranslatability is relatively uninformative, one might wonder whether quasitranslational equivalence is more informative. The answer is affirmative.

Proposition 2.2.3 *(Quasitranslational Equivalence Implies Translational Equivalence).* If L_1 and L_2 are quasitranslationally equivalent, then L_1 and L_2 are translationally equivalent.

Proof: Let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ be a quasitranslation from L_1 to L_2, and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$ a quasitranslation from L_2 to L_1 such that $\phi \dashv\vdash_1 \mathtt{s}(\mathtt{t}(\phi))$ for $\phi \in \mathcal{L}_1$ and $\psi \dashv\vdash_2 \mathtt{t}(\mathtt{s}(\psi))$ for $\psi \in \mathcal{L}_2$. Suppose $\mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi)$. Since $\mathtt{s}$ is a quasitranslation, $\mathtt{s}[\mathtt{t}[\Gamma]] \vdash_1 \mathtt{s}(\mathtt{t}(\phi))$. Hence, $\Gamma \vdash_1 \phi$. Likewise, if $\mathtt{s}[\Delta] \vdash_1 \mathtt{s}(\psi)$, then $\Delta \vdash_2 \psi$. ■

Thus, to establish the translational equivalence between two logics, it suffices to establish quasitranslational equivalence.

There is yet another apparently weaker condition that suffices to establish translational equivalence: we do not even need to require $\mathtt{s}$ to be a quasitranslation!

Proposition 2.2.4 *(Simpler Definition of Translational Equivalence).* Let L_1 and L_2 be logics such that $\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$. Let $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ and let $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. Suppose $\psi \dashv\vdash_2 \mathtt{t}(\mathtt{s}(\psi))$ for all $\psi \in \mathcal{L}_2$. Then $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$.

Proof: First, observe that $\mathtt{s}$ is a translation. For let $\Delta \subseteq \mathcal{L}_2$ and $\psi \in \mathcal{L}_2$. Then $\Delta \vdash_2 \psi$ iff $\mathtt{t}[\mathtt{s}[\Delta]] \vdash_2 \mathtt{t}(\mathtt{s}(\psi))$ iff $\mathtt{s}[\Delta] \vdash_1 \mathtt{s}(\psi)$.

Next, observe that $\mathtt{t}(\mathtt{s}(\mathtt{t}(\phi))) \dashv\vdash_2 \mathtt{t}(\phi)$ by the given constraint. Hence, since $\mathtt{t}$ is a translation, $\mathtt{s}(\mathtt{t}(\phi)) \dashv\vdash_1 \phi$. ■

Translational equivalence can also be useful as a way of measuring whether one logic really contains another as a fragment. Intuitively, one logic is as expressively powerful as another just in case the latter is equivalent to a fragment of the former. The following result is a way of codifying this intuition.

Proposition 2.2.5 *(Translatability is Translational Equivalence to Image).* Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. The following are equivalent:

(a) $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$.

(b) $\mathsf{L}_1 \overset{\leftsquigarrow}{\rightsquigarrow} \mathtt{t}[\mathsf{L}_1]$.

(c) $\mathsf{L}_1 \leftrightsquigarrow \mathtt{t}[\mathsf{L}_1]$.

Proof: Obviously (c) implies (b), which implies (a) by **Facts 1.1.24** and **2.1.2**. So it suffices to show that (a) implies (c). Now, $\mathtt{t}^{-1}$ may not be a function from $\mathtt{t}[\mathsf{L}_1]$ to L_1, since $\mathtt{t}$ might not be injective. But since $\mathtt{t}^{-1}$ is total on $\mathtt{t}[\mathsf{L}_1]$, we can always find a function $\mathtt{t}^* \subseteq \mathtt{t}^{-1}$ (using the axiom of choice) by selecting a $\psi \in \Delta_\phi = \{\psi' \in \mathcal{L}_1 \mid \mathtt{t}(\psi') = \phi\}$ arbitrarily for each $\phi \in \mathtt{t}[\mathsf{L}_1]$ and setting $\mathtt{t}^*(\phi) = \psi$. (If $\mathtt{t}(\psi) = \mathtt{t}(\psi') = \phi$, then $\psi \dashv\vdash_1 \psi'$.) Clearly $\mathtt{t}^*\colon \mathtt{t}[\mathsf{L}_1] \rightsquigarrow \mathsf{L}_1$. Observe now that $\mathtt{t}^*$ is a right-inverse of $\mathtt{t}$, i.e., for all $\phi \in \mathtt{t}[\mathcal{L}_1]$, $\mathtt{t}(\mathtt{t}^*(\phi)) = \phi$. So clearly, for all $\phi \in \mathtt{t}[\mathcal{L}_1]$, $\phi \dashv\vdash_{\mathtt{t}[1]} \mathtt{t}(\mathtt{t}^*(\phi))$. Conversely, for all $\psi \in \mathcal{L}_1$, $\psi \dashv\vdash_1 \mathtt{t}^*(\mathtt{t}(\psi))$ iff $\mathtt{t}(\psi) \dashv\vdash_{\mathtt{t}[1]} \mathtt{t}(\mathtt{t}^*(\mathtt{t}(\psi))) = \mathtt{t}(\psi)$, which obviously holds. So $\mathtt{t}, \mathtt{t}^*\colon \mathsf{L}_1 \leftrightsquigarrow \mathtt{t}[\mathsf{L}_1]$. ■

Corollary 2.2.6 *(Translatability and Equivalence to Fragments)*. Let L_1 and L_2 be logics. The following are equivalent:

(a) $\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$.

(b) There is a $\mathsf{L}_2' \subseteq \mathsf{L}_2$ such that $\mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2'$.

Corollary 2.2.7 *(Logics are Translationally Equivalent to Images Under Translation)*. Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ iff $\mathsf{L}_1 \leftrightsquigarrow \mathtt{t}[\mathsf{L}_1]$. Hence, if $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$ is such that $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$, then $\mathtt{t}[\mathsf{L}_1] \leftrightsquigarrow \mathtt{s}[\mathsf{L}_2]$.

While translational equivalence implies intertranslatability, the converse does not hold. The following example will be discussed more in § 2.5.

Example 2.2.8 *(Intertranslatability without Translational Equivalence)*. Recall from **Example 2.1.8** that **CPL**$\rightsquigarrow$**IPL**. We will see in **Theorems 2.6.5** and **2.6.8** that **IPL**$\rightsquigarrow$**CPL** too. Thus, **CPL** $\overset{\leftsquigarrow}{\rightsquigarrow}$ **IPL**.

Claim: **CPL** $\not\leftrightsquigarrow$ **IPL**.

Proof: Suppose for *reductio* that $\mathtt{t}, \mathtt{s}\colon \mathbf{CPL} \leftrightsquigarrow \mathbf{IPL}$. First, observe that:

$$\begin{aligned} \mathtt{t}(\bot) \vDash_{\mathbf{IPL}} \bot \quad &\Leftrightarrow \quad \mathtt{s}(\mathtt{t}(\bot)) \vDash_{\mathbf{CPL}} \mathtt{s}(\bot) \\ &\Leftrightarrow \quad \bot \vDash_{\mathbf{CPL}} \mathtt{s}(\bot). \end{aligned}$$

But trivially, $\bot \vDash_{\mathbf{CPL}} \mathtt{s}(\bot)$. Thus, $\mathtt{t}(\bot) \equiv_{\mathbf{IPL}} \bot$. Likewise, $\mathtt{s}(\bot) \equiv_{\mathbf{CPL}} \bot$.

Next, note that:

$$\mathtt{s}(\phi), \mathtt{s}(\neg\,\phi) \vDash_{\mathbf{CPL}} \mathtt{s}(\bot) \equiv_{\mathbf{CPL}} \bot.$$

So $\mathtt{s}(\neg\,\phi) \vDash_{\mathbf{CPL}} \neg\,\mathtt{s}(\phi)$. Conversely, since $\mathtt{s}(\phi), \neg\,\mathtt{s}(\phi) \vDash_{\mathbf{CPL}} \bot$:

$$\mathtt{t}(\mathtt{s}(\phi)), \mathtt{t}(\neg\,\mathtt{s}(\phi)) \vDash_{\mathbf{IPL}} \mathtt{t}(\bot) \vDash_{\mathbf{IPL}} \bot.$$

And since $\mathtt{t}(\mathtt{s}(\phi)) \equiv_{\mathbf{IPL}} \phi$ by translational equivalence, it follows that $\phi, \mathtt{t}(\neg\,\mathtt{s}(\phi)) \vDash_{\mathbf{IPL}} \bot$. So $\mathtt{t}(\neg\,\mathtt{s}(\phi)) \vDash_{\mathbf{IPL}} \neg\,\phi$. Hence:

$$\mathtt{s}(\mathtt{t}(\neg\,\mathtt{s}(\phi))) \equiv_{\mathbf{CPL}} \neg\,\mathtt{s}(\phi) \vDash_{\mathbf{CPL}} \mathtt{s}(\neg\,\phi).$$

Combining these together, we obtain that $\neg\,\mathtt{s}(\phi) \equiv_{\mathbf{CPL}} \mathtt{s}(\neg\,\phi)$. But now we are in trouble. Since $\neg\,\neg\,p \nvDash_{\mathbf{IPL}} p$, it follows that $\mathtt{s}(\neg\,\neg\,p) \nvDash_{\mathbf{CPL}} \mathtt{s}(p)$. But as shown above, $\neg\,\neg\,\mathtt{s}(p) \equiv_{\mathbf{CPL}} \mathtt{s}(\neg\,\neg\,p)$, so $\neg\,\neg\,\mathtt{s}(p) \nvDash_{\mathbf{CPL}} \mathtt{s}(p)$, ↯. Therefore, **CPL** $\not\leftrightsquigarrow$ **IPL**. ■

This result generalizes to intermediate logics, i.e., any logic L closed under modus ponens and uniform substitution such that $\mathsf{IPL} \sqsubseteq \mathsf{L} \sqsubseteq \mathsf{CPL}$.

Claim: Let L_1 and L_2 be distinct intermediate logics. Then $\mathsf{L}_1 \not\leftrightsquigarrow \mathsf{L}_2$.

Proof: Suppose $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$. By **Lemma 3.2.2** (discussed in the next chapter), $\mathtt{t}(\neg\,\phi) \equiv_2 \neg\,\mathtt{t}(\phi)$ and $\mathtt{t}(\phi \star \psi) \equiv_2 \mathtt{t}(\phi) \star \mathtt{t}(\psi)$, where $\star \in \{\wedge, \vee, \rightarrow\}$ (and similarly for $\mathtt{s}$). It follows that if $\Theta(\pi_1, \ldots, \pi_n)$ is a propositional schema, then $\mathtt{t}(\Theta(\phi_1, \ldots, \phi_n)) \dashv\vdash_2 \Theta(\mathtt{t}(\phi_1), \ldots, \mathtt{t}(\phi_n))$ (and similarly for $\mathtt{s}$). Thus, suppose $\Gamma \vdash_1 \phi$. Let $p_1, \ldots, p_n$ enumerate all the atomics in Γ and ϕ, and let $\Gamma^{\mathtt{s}}$ and $\phi^{\mathtt{s}}$ be the result of replacing each instance of p_i with $\mathtt{s}(p_i)$. Since L_1 is closed under uniform substitution, $\Gamma^{\mathtt{s}} \vdash_1 \phi^{\mathtt{s}}$. So $\mathtt{t}[\Gamma^{\mathtt{s}}] \vdash_2 \mathtt{t}(\phi^{\mathtt{s}})$. Pushing $\mathtt{t}$ in, we get $\Gamma^{\mathtt{t}\circ\mathtt{s}} \vdash_2 \phi^{\mathtt{t}\circ\mathtt{s}}$. And since $p_i \equiv_2 \mathtt{t}(\mathtt{s}(p_i))$, we have $\Gamma \vdash_2 \phi$. Likewise for the converse. Hence, $\Gamma \vdash_1 \phi$ iff $\Gamma \vdash_2 \phi$, i.e., $\mathsf{L}_1 = \mathsf{L}_2$. ■

While it is natural to conjecture that two logics are notational variants just in case they are translationally equivalent,[4] we will see in § 2.6 that this proposal has its flaws. Still, it is plausible to propose that translational equivalence be at least a necessary condition for some logics to count as genuine notational variants.

[4] Indeed, Straßburger [2007, p. 139] makes this claim. Mossakowski et al. [2009, p. 7] and French [2010, p. 134] claim translational equivalence is sufficient for notational variance.

Just as for ordinary translations, there is a nice alternative but equivalent definition of translational equivalence in terms of theory spaces.[5]

Lemma 2.2.9 *(Removing an Interior Consequence Operator)*. Let L_1 and L_2 be logics. If $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, then for all $\Gamma \subseteq \mathcal{L}_1$:

$$\mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\Gamma)]) = \mathtt{Cn}_2(\mathtt{t}[\Gamma]).$$

Proof: The $\supseteq$-case follows from (CO2) in **Definition 1.1.8**. For the $\subseteq$-case, suppose $\psi \in \mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\Gamma)])$. Thus, $\mathtt{t}[\mathtt{Cn}_1(\Gamma)] \vDash_2 \psi$. But since $\Gamma \vDash_1 \mathtt{Cn}_1(\Gamma)$, we have $\mathtt{t}[\Gamma] \vDash_2 \mathtt{t}[\mathtt{Cn}_1(\Gamma)]$. Hence, $\mathtt{t}[\Gamma] \vDash_2 \psi$, and so $\psi \in \mathtt{Cn}_2(\mathtt{t}[\Gamma])$. ∎

Proposition 2.2.10 *(Theoretic Formulation of Translational Equivalence)*. Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. Then $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$ iff $\mathtt{t}^{\mathtt{Th}}\colon \mathbb{T}_1 \cong \mathbb{T}_2$ where $\mathtt{t}^{\mathtt{Th}} = \mathtt{s}^{\mathtt{Th}^{-1}}$ and for all $\Sigma_1 \subseteq \wp(\mathcal{L}_1)$ and $\Sigma_2 \subseteq \wp(\mathcal{L}_2)$:

$$\mathtt{t}^{\mathtt{Th}}(\bigwedge \Sigma_1) = \bigwedge \mathtt{t}^{\mathtt{Th}}[\Sigma_1] \qquad \text{and} \qquad \mathtt{s}^{\mathtt{Th}}(\bigwedge \Sigma_2) = \bigwedge \mathtt{s}^{\mathtt{Th}}[\Sigma_2].$$

Proof:

($\Rightarrow$) Suppose $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$. By **Proposition 2.1.16**, $\mathtt{t}^{\mathtt{Th}}$ and $\mathtt{s}^{\mathtt{Th}}$ are order-embeddings on $\mathbb{T}_1$ and $\mathbb{T}_2$ that preserve meets, so it suffices to show that they are inverses, i.e., $\mathtt{s}^{\mathtt{Th}}(\mathtt{t}^{\mathtt{Th}}(\Gamma)) = \Gamma$ for all $\Gamma \in \mathtt{Th}(\mathbf{L}_1)$ and $\mathtt{t}^{\mathtt{Th}}(\mathtt{s}^{\mathtt{Th}}(\Delta)) = \Delta$ for all $\Delta \in \mathtt{Th}(\mathbf{L}_2)$. But for instance, if $\Gamma \in \mathtt{Th}(\mathbf{L}_1)$, then $\mathtt{s}^{\mathtt{Th}}(\mathtt{t}^{\mathtt{Th}}(\Gamma)) = \mathtt{s}^{\mathtt{Th}}(\mathtt{Cn}_2(\mathtt{t}[\Gamma])) = \mathtt{Cn}_1(\mathtt{s}[\mathtt{Cn}_2(\mathtt{t}[\Gamma])]) = \mathtt{Cn}_1(\mathtt{s}[\mathtt{t}[\Gamma]]) = \mathtt{Cn}_1(\Gamma) = \Gamma$.

($\Leftarrow$) Suppose $\mathtt{t}^{\mathtt{Th}}\colon \mathbb{T}_1 \cong \mathbb{T}_2$ where $\mathtt{s}^{\mathtt{Th}} = \mathtt{t}^{\mathtt{Th}^{-1}}$ and that $\mathtt{t}^{\mathtt{Th}}$ and $\mathtt{s}^{\mathtt{Th}}$ preserve meets. Then $\mathtt{t}$ and $\mathtt{s}$ are translations, so it suffices to show that $\mathtt{s}(\mathtt{t}(\phi)) \dashv\vdash_1 \phi$ for $\phi \in \mathcal{L}_1$ and $\mathtt{t}(\mathtt{s}(\psi)) \dashv\vdash_2 \psi$ for $\psi \in \mathcal{L}_2$. But $\mathtt{Cn}_1(\phi) = \mathtt{s}^{\mathtt{Th}}(\mathtt{t}^{\mathtt{Th}}(\mathtt{Cn}_1(\phi))) = \mathtt{Cn}_1(\mathtt{s}[\mathtt{Cn}_2(\mathtt{t}[\mathtt{Cn}_1(\phi)])]) = \mathtt{Cn}_1(\mathtt{s}(\mathtt{t}(\phi)))$. So this completes the proof. ∎

Note that the fact that $\mathtt{t}^{\mathtt{Th}}\colon \mathbb{T}_1 \cong \mathbb{T}_2$ does not guarantee that it preserves meets on *arbitrary* $\Sigma \subseteq \wp(\mathcal{L}_1)$, though it does guarantee preservation of meets on $\Sigma \subseteq \mathtt{Th}(\mathsf{L}_1)$ since order-isomorphisms guarantee the preservation of meets in general.

[5] Caleiro and Gonçalves [2007, pp. 106–107] prove a slightly modified form of **Proposition 2.2.10**.

§ 2.3 Isomorphism

It is worth briefly mentioning an even stronger notion of equivalence between logics that has been occasionally discussed in the literature, viz., that of an isomorphism.[6]

Definition 2.3.1 *(Isomorphism)*. Let L_1 and L_2 be logics. We will say a map $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is an ***isomorphism*** between L_1 and L_2 (written "$\mathtt{t}\colon \mathsf{L}_1 \cong \mathsf{L}_2$") if $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ and $\mathtt{t}$ is bijective. We will say L_1 is ***isomorphic*** to L_2 (written "$\mathsf{L}_1 \cong \mathsf{L}_2$") if there is an isomorphism between them.

The following are easily established via previous results:

Fact 2.3.2 *(Injective Translatability Implies Isomorphism with Image)*. If $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ is injective, then $\mathtt{t}\colon \mathsf{L}_1 \cong \mathtt{t}[\mathsf{L}_1]$ and $\mathtt{t}^{-1}\colon \mathtt{t}[\mathsf{L}_1] \cong \mathsf{L}_1$.

Corollary 2.3.3 *(Isomorphism Implies Translational Equivalence)*. If $\mathtt{t}\colon \mathsf{L}_1 \cong \mathsf{L}_2$, then $\mathtt{t}, \mathtt{t}^{-1}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$.

While isomorphism implies translational equivalence, the converse does not generally hold for very simple cardinality reasons.

Example 2.3.4 *(Translational Equivalence Does Not Imply Isomorphism)*. Where $\mathcal{L}_{\mathtt{Prop}}$ is the standard propositional language, Let $\mathcal{L}_{\mathtt{Prop}}(@_r)_{r\in\mathbb{R}}$ be the result of adding a unary operator $@_r$ to $\mathcal{L}_{\mathtt{Prop}}$ for each $r \in \mathbb{R}$. Let **CPL**(@) be the extension of **CPL** with $\mathcal{L}_{\mathtt{Prop}}(@_r)_{r\in\mathbb{R}}$ where the semantic clause for each $@_r$ is as follows:

$$v \Vdash_@ @_r\phi \quad\Leftrightarrow\quad v \Vdash_@ \phi.$$

Clearly, **CPL**(@) $\leftrightsquigarrow$ **CPL**. But **CPL**(@) $\not\cong$ **CPL**, since there are uncountably many $\mathcal{L}_{\mathtt{Prop}}(@_r)_{r\in\mathbb{R}}$-formulas.

The above example illustrates that isomorphism is too strong of a constraint for notational variance. **CPL**(@), after all, seems to be a (albeit rather silly) notational variant of **CPL**, as each operator $@_r$ is straightforwardly definable in **CPL**. So it would be too demanding to require logics be isomorphic for them to be deemed notational variants. As far as I can tell, no one in the literature has argued that isomorphism is a necessary condition for notational variance. Moreover, in § 2.6, we will see that it is generally not even sufficient—logics that do not appear to be notational variants may nevertheless be isomorphic in the sense of **Definition 2.3.1**.

[6] See, e.g., Wójcicki 1988; Caleiro and Gonçalves 2007; Straßburger 2007.

The theoretic formulation of isomorphism is fairly easy to derive from the corresponding formulation of translatability (**Propositions 2.1.16**, **4.2.11 and 4.2.15**).

Proposition 2.3.5 *(Theoretic Formulation of Isomorphism)*. Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathsf{L}_1 \cong \mathsf{L}_2$ iff $\mathtt{t}^{\mathrm{Th}}\colon \mathbb{T}_1 \cong \mathbb{T}_2$ where $\mathtt{t}^{\mathrm{Th}}(\bigwedge \Sigma) = \bigwedge \mathtt{t}^{\mathrm{Th}}[\Sigma]$ for all $\Sigma \subseteq \wp(\mathcal{L}_1)$.

§ 2.4 Auxiliary Assumptions

Sometimes, the notion of a translation is relativized to some class of auxiliary hypotheses. For instance, recall from **Example 2.1.7** the standard translation ST_n from the modal logic **S4** (= **KT4**) into **FOL**:

$$\begin{array}{lcl} \mathsf{ST}_x(p_i) & = & P^1_i(x) \\ \mathsf{ST}_x(\neg\phi) & = & \neg\,\mathsf{ST}_x(\phi) \\ \mathsf{ST}_x(\phi \wedge \psi) & = & (\mathsf{ST}_x(\phi) \wedge \mathsf{ST}_x(\psi)) \\ \mathsf{ST}_x(\Box\phi) & = & \forall y\,((x \twoheadrightarrow y) \to \mathsf{ST}_y(\phi)) \end{array}$$

This mapping is not quite a translation from **S4** into **FOL**. To illustrate, consider the T-axiom $(\Box p \to p)$. The standard translation of the T-axiom is of the form:

$$\forall y\ ((x \twoheadrightarrow y) \to P(y)) \to P(x).$$

As a first-order formula, however, this sentence is not **FOL**-valid, since $\twoheadrightarrow$ need not be reflexive. Likewise, the standard translation of the 4-axiom $(\Box\Box p \to \Box p)$ is not **FOL**-valid:

$$\forall y\ ((x \twoheadrightarrow y) \to \forall z\ ((y \twoheadrightarrow z) \to P(z))) \to \forall y\ ((x \twoheadrightarrow y) \to P(y))$$

There are two ways to repair the translation. One is to systematically revise the compositional clauses in the definition of ST_n and conjoin the original standard translation with these axioms to form a new translation ST^*_n. For example, the new translation clause for negation would be defined as follows:

$$\begin{array}{rcl} \theta & := & \forall x\ (x \twoheadrightarrow x) \wedge \forall x\,\forall y\,\forall z\ ((x \twoheadrightarrow y) \wedge (y \twoheadrightarrow z) \to (x \twoheadrightarrow z)) \\ \mathsf{ST}^*_n(\neg\phi) & = & \theta \wedge \neg\,\mathsf{ST}^*_n(\phi). \end{array}$$

A more natural approach, however, would be to keep the original definition of ST_n and simply relativize the notion of a translation to the reflexivity and transitivity axioms. For notice that where θ is defined as above, we do have:

$$\Gamma \vDash_{\mathbf{S4}} \phi \quad \Leftrightarrow \quad \theta, \mathsf{ST}_n[\Gamma] \vDash_{\mathbf{FOL}} \mathsf{ST}_n(\phi).$$

This suggests the following generalization of **Definition 2.1.1**.

Definition 2.4.1 *(Translation with Auxiliary Assumptions)*. Let L_1 and L_2 be logics and let $\Lambda \subseteq \mathcal{L}_2$. A map $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is a ***translation given*** Λ, or alternatively a Λ***-translation*** (written "$\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow_\Lambda \mathsf{L}_2$"), if for all $\Gamma \subseteq \mathcal{L}_1$ and all $\phi \in \mathcal{L}_1$:

$$\Gamma \vdash_1 \phi \quad \Leftrightarrow \quad \Lambda, \mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi).$$

Notice that a translation in the sense of **Definition 2.1.1** is just a $\varnothing$-translation. L_1 is ***translatable given*** Λ, or Λ***-translatable***, into L_2 (written "$\mathsf{L}_1 \rightsquigarrow_\Lambda \mathsf{L}_2$") if $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow_\Lambda \mathsf{L}_2$ for some $\mathtt{t}$. Where $\Lambda \subseteq \mathcal{L}_2$ and $\Lambda' \subseteq \mathcal{L}_1$, we write "$\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \overset{\leftsquigarrow \Lambda'}{\rightsquigarrow_\Lambda} \mathsf{L}_2$" to indicate that $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow_\Lambda \mathsf{L}_2$ and $\mathtt{s}\colon \mathsf{L}_2 \rightsquigarrow_{\Lambda'} \mathsf{L}_1$.

L_1 is said to be ***interpretable*** in L_2 (written "$\mathsf{L}_1 \rightsquigarrow_i \mathsf{L}_2$") if $\mathsf{L}_1 \rightsquigarrow_\Lambda \mathsf{L}_2$ for some $\Lambda \subseteq \mathcal{L}_2$. A pair of interpretations $\mathtt{t}_1\colon \mathsf{L}_1 \rightsquigarrow_i \mathsf{L}_2$ and $\mathtt{t}_2\colon \mathsf{L}_2 \rightsquigarrow_i \mathsf{L}_1$ is called an ***interpretation scheme*** (written "$\mathtt{t}_1, \mathtt{t}_2\colon \mathsf{L}_1 \overset{\leftsquigarrow}{\rightsquigarrow}_i \mathsf{L}_2$") between L_1 and L_2. We will say L_1 and L_2 are ***mutually interpretable*** (written "$\mathsf{L}_1 \overset{\leftsquigarrow}{\rightsquigarrow}_i \mathsf{L}_2$") if there is an interpretation scheme between them.

Translations with auxiliary assumptions are not uncommon. A well-known example of such a translation is that from many-sorted logic to one-sorted logic.

Convention: Where $\alpha_1, \ldots, \alpha_n$ is some sequence of objects (variables, terms, or whatever), we will write "$\overline{\alpha}$" for brevity. For example, we may write "$\phi(\overline{y})$" in place of "$\phi(y_1, \ldots, y_n)$". In addition, given that we abbreviate "$\alpha_1, \ldots, \alpha_n$" as "$\overline{\alpha}$", and given that f is a unary function, we may write "$f(\overline{\alpha})$" instead of "$f(\alpha_1), \ldots, f(\alpha_n)$". For example, we may write "$\forall \overline{y}$" for "$\forall y_1 \cdots \forall y_n$". (Generally, it will be clear from context whether f is unary, and thus whether "$f(\overline{\alpha})$" means "$f(\alpha_1, \ldots, \alpha_n)$" or "$f(\alpha_1), \ldots, f(\alpha_n)$".)

We will not generally make explicit the length of a sequence $\overline{\alpha}$. Instead, we let context settle these questions, though unless stated otherwise, it should not be assumed that any sequence mentioned is the same length as any other. For example, in the abbreviation "$\phi(\overline{y}, \overline{z})$", it is *not* assumed that $|\overline{y}| = |\overline{z}|$.

Example 2.4.2 *(Two-Sorted Languages)*. We will define the two-sorted language **TSL** as follows. First, we define the set $\mathtt{StVar} = \{s_1, s_2, s_3, \ldots\}$ and for each $n, m \in \mathbb{N}$, we will let $\mathtt{Pred}^{n;m} = \{P_1^{n;m}, P_2^{n;m}, P_3^{n;m}, \ldots\}$. The syntax $\mathcal{L}_{\mathbf{TSL}}$ is defined as follows:

$$\phi ::= P^{n;m}(y_1, \ldots, y_n; t_1, \ldots, t_m) \mid \neg\phi \mid (\phi \wedge \phi) \mid \forall x\, \phi \mid \forall s\, \phi.$$

A ***two-sorted model*** is a triple $\mathcal{M} = \langle D, W, I \rangle$ where D and W are nonempty disjoint sets and $I(P^{n;m}) \subseteq D^n \times W^m$. A ***variable assignment*** on $\mathcal{M}$ is a map g that sends each member of $\mathtt{Var}$ to a member of D and each member of $\mathtt{StVar}$

to a member of W. We define $\Vdash_{\mathbf{TSL}}$ as follows:

$$\begin{array}{lcl}
\mathcal{M}, g \Vdash_{\mathbf{TSL}} P^{n;m}(\overline{y};\overline{t}) & \Leftrightarrow & \langle g(\overline{y}), g(\overline{t})\rangle \in I(P^{n;m}) \\
\mathcal{M}, g \Vdash_{\mathbf{TSL}} \neg\phi & \Leftrightarrow & \mathcal{M}, g \nVdash_{\mathbf{TSL}} \phi \\
\mathcal{M}, g \Vdash_{\mathbf{TSL}} \phi \wedge \psi & \Leftrightarrow & \mathcal{M}, g \Vdash_{\mathbf{TSL}} \phi \text{ and } \mathcal{M}, g \Vdash_{\mathbf{TSL}} \psi \\
\mathcal{M}, g \Vdash_{\mathbf{TSL}} \forall x\, \phi & \Leftrightarrow & \text{for all } a \in D\text{: } \mathcal{M}, g^x_a \Vdash_{\mathbf{TSL}} \phi \\
\mathcal{M}, g \Vdash_{\mathbf{TSL}} \forall s\, \phi & \Leftrightarrow & \text{for all } w \in W\text{: } \mathcal{M}, g^s_w \Vdash_{\mathbf{TSL}} \phi.
\end{array}$$

Clearly, **FOL**$\rightsquigarrow$**TSL**. On the flip side, let $D := P^1_1$ and $W := P^1_2$. Define $\mathtt{t}\colon \mathcal{L}_{\mathbf{TSL}} \to \mathcal{L}_{\mathbf{FOL}}$ as follows:[a]

$$\begin{array}{rcl}
\mathtt{t}(x_i) & = & x_{2i} \\
\mathtt{t}(s_i) & = & x_{2i+1} \\
\mathtt{t}(P^{n;m}_i(\overline{y};\overline{t}) & = & P^{n+m}_{2^i3^n5^m}(\mathtt{t}(\overline{y}), \mathtt{t}(\overline{t})) \\
\mathtt{t}(\neg\phi) & = & \neg\,\mathtt{t}(\phi) \\
\mathtt{t}(\phi \wedge \psi) & = & \mathtt{t}(\phi) \wedge \mathtt{t}(\psi) \\
\mathtt{t}(\forall x\, \phi) & = & \forall \mathtt{t}(x)\ (D(\mathtt{t}(x)) \to \mathtt{t}(\phi)) \\
\mathtt{t}(\forall s\, \phi) & = & \forall \mathtt{t}(s)\ (W(\mathtt{t}(s)) \to \mathtt{t}(\phi))\,.
\end{array}$$

Also, define Λ as the following set of formulas:

$$\forall x\ (D(x) \leftrightarrow \neg W(x))$$
$$\exists x\, D(x) \wedge \exists x\, W(x)$$
$$\forall\overline{y}\,\forall\overline{z}\left(P^{n+m}_{2^i3^n5^m}(\overline{y},\overline{z}) \to \bigwedge_{i=1}^{n} D(y_i) \wedge \bigwedge_{j=1}^{m} W(j_m)\right).$$

Then $\mathtt{t}$: **TSL**$\rightsquigarrow_\Lambda$ **FOL**. The same strategy could work in principle for logics with n-many sorts for finite n. It will not work for infinitely-sorted logics, since the first axiom in Λ would then need to be infinitary.

While this mapping from **TSL** to **FOL** is often presented as an interpretation in the sense of **Definition 2.4.1**, it is worth noting that it can be converted into a full-blown translation as follows. First, define $\chi^{n;m}_i$ as the **FOL**-formula:

$$\forall\overline{y}\,\forall\overline{z}\left(P^{n+m}_{2^i3^n5^m}(\overline{y},\overline{z}) \to \bigwedge_{i=1}^{n} D(y_i) \wedge \bigwedge_{j=1}^{m} W(j_m)\right).$$

Define also $\delta := \forall x\,(D(x) \leftrightarrow \neg W(x)) \wedge \exists x\, D(x) \wedge \exists x\, W(x)$. Then we can define a translation from **TSL** to **FOL** as follows:

$$\mathtt{t}(x_i) \quad = \quad x_{2i}$$

$$
\begin{aligned}
\mathtt{t}(s_i) &= x_{2i+1} \\
\mathtt{t}(P_i^{n;m}(\overline{y};\overline{t})) &= \delta \wedge \chi_i^{n;m} \to P_{2^i3^n5^m}^{n+m}(\mathtt{t}(\overline{y}), \mathtt{t}(\overline{t})) \\
\mathtt{t}(\neg\phi) &= \delta \wedge \bigwedge_{P_i^{n;m} \in \phi} \chi_i^{n;m} \to \neg\, \mathtt{t}(\phi) \\
\mathtt{t}(\phi \wedge \psi) &= \mathtt{t}(\phi) \wedge \mathtt{t}(\psi) \\
\mathtt{t}(\forall x\, \phi) &= \delta \wedge \bigwedge_{P_i^{n;m} \in \phi} \chi_i^{n;m} \to \forall \mathtt{t}(x)\ (D(\mathtt{t}(x)) \to \mathtt{t}(\phi)) \\
\mathtt{t}(\forall s\, \phi) &= \delta \wedge \bigwedge_{P_i^{n;m} \in \phi} \chi_i^{n;m} \to \forall \mathtt{t}(s)\ (W(\mathtt{t}(s)) \to \mathtt{t}(\phi))\,.
\end{aligned}
$$

Again, this strategy will work for n-sorted logics for any finite n.

[a] Note that we want to map atomic formulas involving $P_i^{n;m}$ to $P_{2^i3^n5^m}^{n+m}$ to avoid accidentally sending distinct predicates to the same predicate.

While the above example does not show that interpretability does not imply translatability, it is easy to manufacture artificial examples where the two notions come apart. In addition, there are examples "in the wild" where they come apart:

Example 2.4.3 *(Interpreting Classical Logic in Kleene Logic)*. We saw in **Example 2.1.14** that **CPL** $\not\rightsquigarrow$ **K3**. But let $\Lambda = \{p \vee \neg p \mid p \in \mathtt{Prop}\}$. Then $\Gamma \vDash_{\mathbf{CPL}} \phi$ iff $\Lambda, \Gamma \vDash_{\mathbf{K3}} \phi$. For if $v \Vdash_{\mathbf{K3}} \Lambda$, then $v(p)\downarrow$ for all $p \in \mathtt{At}$. And it is straightforward by induction to show that if $v(p)\downarrow$ for all $p \in \mathtt{At}$ that occur in ϕ, then $v \Vdash_{\mathbf{K3}} \phi$ iff $v \Vdash_{\mathbf{CPL}} \phi$.

Even so, there is a sense in which interpretability reduces to translatability: interpretability is translatability into a reduction.

Fact 2.4.4 *(Reducing Interpretability to Translatability)*. Let L_1 and L_2 be logics and let $\Lambda \subseteq \mathcal{L}_2$. Then $\mathsf{L}_1 \rightsquigarrow_\Lambda \mathsf{L}_2$ iff $\mathsf{L}_1 \rightsquigarrow (\mathsf{L}_2)_\Lambda$.

From this, the theoretic formulation of interpretability is easily derived:

Proposition 2.4.5 *(Theoretic Formulation of Interpretation)*. Let L_1 and L_2 be logics, let $\Lambda \subseteq \mathcal{L}_2$, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow_\Lambda \mathsf{L}_2$ iff $\mathtt{t}^{\mathrm{Th}}$ is an order-embedding from $\mathbb{T}_1$ to $(\mathbb{T}_2)_\Lambda$ (where $(\mathbb{T}_2)_\Lambda$ is the theory space of $(\mathsf{L}_2)_\Lambda$) such that $\mathtt{t}^{\mathrm{Th}}(\bigwedge \Sigma) = \bigwedge \mathtt{t}^{\mathrm{Th}}[\Sigma]$ for all $\Sigma \subseteq \wp(\mathcal{L}_1)$.

We might try to formulate interpretational equivalence in a way analogous to translational equivalence. The most natural way to do this is as follows:

Definition 2.4.6 *(Interpretational Equivalence)*. Let L_1 and L_2 be some mutually interpretable logics. Where $\mathtt{t}_1\colon \mathsf{L}_1 \rightsquigarrow_\Lambda \mathsf{L}_2$ and $\mathtt{t}_2\colon \mathsf{L}_2 \rightsquigarrow_{\Lambda'} \mathsf{L}_1$, we say the translation scheme $\langle \mathtt{t}_1, \mathtt{t}_2 \rangle$ is ***reversible*** via Λ and Λ' (written "$\mathtt{t}_1, \mathtt{t}_2\colon \mathsf{L}_1 \leftrightsquigarrow_\Lambda^{\Lambda'} \mathsf{L}_2$") if:

(a) for all $\phi \in \mathcal{L}_1$: $\Lambda', \mathtt{t}_2(\mathtt{t}_1(\phi)) \dashv\vdash_1 \Lambda', \phi$

(b) for all $\psi \in \mathcal{L}_2$: $\Lambda, \mathtt{t}_1(\mathtt{t}_2(\psi)) \dashv\vdash_2 \Lambda, \psi$.

We will say L_1 and L_2 are ***interpretationally equivalent*** (written "$\mathsf{L}_1 \leftrightsquigarrow_i \mathsf{L}_2$") if there is a reversible translation scheme between them.

The thought is supposed to be that it is only relative to the background auxiliary assumptions can we assess whether or not $\mathtt{s}(\mathtt{t}(\phi))$ is equivalent to ϕ, since those are the assumptions needed to interpret the other language in the first place. We could have required the even stronger constraint that $\mathtt{s}(\mathtt{t}(\phi)) \dashv\vdash_1 \phi$—but it would not have made a difference, since already we have the following:

Proposition 2.4.7 *(Interpretational Equivalence Implies Translational Equivalence)*. Let L_1 and L_2 be logics. If $\mathsf{L}_1 \leftrightsquigarrow_i \mathsf{L}_2$, then $\mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$.

Proof: Suppose $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow_\Lambda^{\Lambda'} \mathsf{L}_2$. Then observe that by (a) of **Definition 2.4.6**, $\Lambda' \vdash_1 \mathtt{s}[\mathtt{t}[\Lambda']]$. Hence, because $\mathtt{s}$ is a Λ'-interpretation, $\vdash_2 \mathtt{t}[\Lambda']$, from which it follows that $\vdash_1 \Lambda'$. So for $\Sigma \subseteq \mathcal{L}_2$ and $\psi \in \mathcal{L}_2$, $\Sigma \vdash_2 \psi$ iff $\Lambda', \mathtt{s}[\Sigma] \vdash_2 \mathtt{s}(\psi)$ iff $\mathtt{s}[\Sigma] \vdash_2 \mathtt{s}(\psi)$. Thus, $\mathtt{s}\colon \mathsf{L}_2 \rightsquigarrow \mathsf{L}_1$. Likewise, $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$. Moreover, since $\vdash_1 \Lambda'$, we have $\mathtt{s}(\mathtt{t}(\phi)) \dashv\vdash_1 \Lambda', \mathtt{s}(\mathtt{t}(\phi)) \dashv\vdash_1 \Lambda', \phi \dashv\vdash_1 \phi$. Hence, $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$. ■

It seems in general difficult to formulate a notion of interpretational equivalence that does not collapse to translational equivalence. For instance, it does not help to replace (a) and (b) in **Definition 2.4.6** with the weaker constraints:

(a′) for all $\phi \in \mathcal{L}_1$: $\Lambda', \mathtt{s}[\Lambda], \mathtt{s}(\mathtt{t}(\phi)) \dashv\vdash_1 \Lambda', \mathtt{s}[\Lambda], \phi$

(b′) for all $\psi \in \mathcal{L}_2$: $\Lambda, \mathtt{t}[\Lambda'], \mathtt{t}(\mathtt{s}(\psi)) \dashv\vdash_2 \Lambda, \mathtt{t}[\Lambda'], \psi$.

It is easy to check that the same proof in **Proposition 2.4.7** undermines this attempt to define interpretational equivalence as distinct from translational equivalence. It seems that if one can "reverse" an interpretational scheme in any reasonable sense, then the auxiliary assumptions in the interpretations are eliminable.

§ 2.5 Expressive Equivalence vs. Notational Variance

Intuitively, notational variants ought not to differ in their expressive power. After all, if $\mathbf{L}_1$ and $\mathbf{L}_2$ are really *notational* variants, there should be no barrier in principle to expressing everything that one language can express in the other language, just using different notation. Much of the literature seems to implicitly assume that the converse is also true, i.e., if two languages have the same expressive power, then they must somehow be the same system disguised by different notation. But is that right? Must expressively equivalent languages also be notational variants?

Given the remarks at the beginning of § 2.2, I think it is fairly clear that two languages translationally equivalence is a necessary condition for notational variance. Several authors have claimed that translational equivalence is also a sufficient condition for notational variance. We will see in § 2.6 that this cannot be right if we are to seek a precisification of our intuitive notion of expressivity. But the amendments that would need to be imposed on translational equivalence to yield an adequate account of notational variance do not affect the discussion at hand. So to simplify matters, we will set such considerations aside and just stipulate for the moment that two languages are notational variants just in case they are translationally equivalent.

Suppose expressive equivalence coincides with notational variance and hence translational equivalence. Let us use "$\approx$" to stand for expressive equivalence. Thus, we are assuming that $\mathbf{L}_1 \approx \mathbf{L}_2$ iff $\mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$. Now, intuitively, expressive equivalence ought to coincide with expressive bi-inclusion. That is, letting "$\leqslant$" stand for expressive inclusion, we want $\mathbf{L}_1 \approx \mathbf{L}_2$ just in case $\mathbf{L}_1 \leqslant \mathbf{L}_2$ and $\mathbf{L}_2 \leqslant \mathbf{L}_1$. Moreover, expressive inclusion ought to coincide with expressive equivalence to some fragment. Thus, we want $\mathbf{L}_1 \leqslant \mathbf{L}_2$ just in case for some $\mathbf{L}_2' \subseteq \mathbf{L}_2$, we have $\mathbf{L}_1 \approx \mathbf{L}_2'$.

But we cannot have all of these things in general. More specifically, the following three constraints cannot be jointly satisfied for all $\mathbf{L}_1$ and $\mathbf{L}_2$:

(I) $\mathbf{L}_1 \approx \mathbf{L}_2$ iff $\mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$.

(II) If $\mathbf{L}_1 \leqslant \mathbf{L}_2$ and $\mathbf{L}_2 \leqslant \mathbf{L}_1$, then $\mathbf{L}_1 \approx \mathbf{L}_2$.

(III) If $\mathbf{L}_1 \approx \mathbf{L}_2' \subseteq \mathbf{L}_2$, then $\mathbf{L}_1 \leqslant \mathbf{L}_2$.

Suppose for *reductio* that (I)–(III) held. Let $\mathbf{L}_1 \overset{\leftsquigarrow}{\rightsquigarrow} \mathbf{L}_2$. By **Proposition 2.2.5** (from (a) to (c)), that means that there are some $\mathbf{L}_1' \subseteq \mathbf{L}_1$ and $\mathbf{L}_2' \subseteq \mathbf{L}_2$ such that $\mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2'$ and $\mathbf{L}_2 \leftrightsquigarrow \mathbf{L}_1'$. By (I), this means that $\mathbf{L}_1 \approx \mathbf{L}_2'$ and $\mathbf{L}_2 \approx \mathbf{L}_1'$. Since $\mathbf{L}_1 \approx \mathbf{L}_2' \subseteq \mathbf{L}_2$ and $\mathbf{L}_2 \approx \mathbf{L}_1' \subseteq \mathbf{L}_1$, we have $\mathbf{L}_1 \leqslant \mathbf{L}_2$ and $\mathbf{L}_2 \leqslant \mathbf{L}_1$ by (III). So by (II), that means $\mathbf{L}_1 \approx \mathbf{L}_2$, and so $\mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$ by (I) again. Hence, we have shown from (I)–(III) that if $\mathbf{L}_1 \overset{\leftsquigarrow}{\rightsquigarrow} \mathbf{L}_2$, then $\mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$. But as **Example 2.2.8** shows, we can have $\mathbf{L}_1 \overset{\leftsquigarrow}{\rightsquigarrow} \mathbf{L}_2$ without $\mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$, $\lightning$. Therefore, one of (I)–(III) must go.

(III) seems undeniable. We only need three very weak claims to justify maintaining it. First, if $\mathbf{L}_1 \approx \mathbf{L}_2'$, then $\mathbf{L}_1 \leqslant \mathbf{L}_2'$. This just says that expressive equivalence should imply expressive bi-inclusion. Second, if $\mathbf{L}_2' \subseteq \mathbf{L}_2$, then $\mathbf{L}_2' \leqslant \mathbf{L}_2$. Again, this seems uncontroversial: fragments of a language are expressively included in that language. Finally, $\leqslant$ is transitive. If $\leqslant$ were not transitive, it would be unfit to call it a precisification of expressive *inclusion*.

(II) seems less like a constraint that can be denied and more like half of a *definition* of expressive equivalence. Indeed, it seems incoherent for two languages to each be expressively included in the other without them being considered expressively equivalent. For if they were included in each other, then everything you could say with the one you could say in the other.

This leaves just one option: reject (I). Now, we could reject the right-to-left direction of (I). Doing so would mean that we could not equate translational equivalence with notational variance. In particular, translational equivalence could not imply notational variance, since notational variance implies expressive equivalence. If we had strong reasons to think that translational equivalence does not imply notational variance, this would be an attractive option. Indeed, we will see very compelling reasons to think this. But in Chapter 3, we will explore a number of properties that translations could have. And as we said above, we will see that translational systems with some of these properties are not subject to these same worries. So at the end of the day, resolving this trilemma by rejecting the right-to-left direction of (I) does not seem promising.

Instead, I propose we reject the left-to-right direction of (I). That is, while I agree that notational variants ought to agree in their expressive power—that is, expressive power is one of the key properties that notational variants must share to count as notational variants—still, two languages that agree in their expressive power may not necessarily be notational variants of one another.

Consider, again, the case of **CPL** and **IPL** from **Example 2.2.8**, where we observed that **CPL** $\overset{\leftrightsquigarrow}{\rightsquigarrow}$ **IPL** but **CPL** $\not\leftrightsquigarrow$ **IPL**. Thus, **CPL** is translationally equivalent to a fragment of **IPL**, and **IPL** is likewise translationally equivalent to a fragment of **CPL**, but **CPL** and **IPL** are not translationally equivalent full stop. In terms of notational variance, **CPL** and **IPL** are not notational variants of each other, as we would expect. But surprisingly, they are notational variants of a fragment of the other.[7]

Loosely, the situation is like this. Each of **CPL** and **IPL** has a sufficient amount of structure to interpret the other within it—or, as I would prefer to put it, each expressively includes the other in a fragment. But globally, the two languages do not have the same outer structure and thus are not mere notational variants of one

[7]It should be noted that this example depends on the assumption that $\mathbf{L}_1 \leqslant \mathbf{L}_2$ iff $\mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$, which we will later reject. Ultimately, we will argue that $\mathbf{L}_1 \leqslant \mathbf{L}_2$ iff there is a *schematic* $\mathtt{t}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$. And indeed, there is no schematic translation from **IPL** to **CPL**. So **CPL** and **IPL** are not expressively equivalent in this sense. Still, there are other examples of languages $\mathbf{L}_1$ and $\mathbf{L}_2$ such that schematically $\mathbf{L}_1 \overset{\leftrightsquigarrow}{\rightsquigarrow} \mathbf{L}_2$ but $\mathbf{L}_1 \not\leftrightsquigarrow \mathbf{L}_2$. So using **CPL** and **IPL** prior to introducing this constraint still illustrates the point we are making here.

another. This is similar to what we find in, say, order theory. It is well-known that each of $[-1,1]$ and $(-1,1)$ can be order-embedded into the other. But $[-1,1]$ and $(-1,1)$ are not isomorphic orders, since the former but not the latter has endpoints. Each can embed the other, but they have different overall structures. This is particular relevant for the claim that we should separate expressive equivalence and notational variance. We saw in **Proposition 2.1.16** that translatability reduces to the existence of a certain kind of order-embedding on the theory spaces. And in **Proposition 2.2.10**, we saw translational equivalence reduces to the existence of a certain kind of isomorphism on the theory spaces. Isomorphism seems more appropriate as the condition corresponding to notational variance than co-embedding. But co-embedding seems sufficient as a condition for expressive equivalence.

Thus, as a first pass, we can propose the following precisification of the concepts of expressive power and notational variance:[8]

- $\mathbf{L}_1$ is expressively included in $\mathbf{L}_2$ just in case $\mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$.
- $\mathbf{L}_1$ is expressively equivalent to $\mathbf{L}_2$ just in case $\mathbf{L}_1 \mathrel{\substack{\leftsquigarrow \\ \rightsquigarrow}} \mathbf{L}_2$.
- $\mathbf{L}_1$ is a notational variant of $\mathbf{L}_2$ just in case $\mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$.

The philosophical significance of the separation between expressive equivalence and notational variance will be explored in Chapter 5.

§ 2.6 Triviality Results

In § 2.2, we argued that for two logics to be notational variants, they have to be at least translationally equivalent. In other words, translational equivalence is a necessary condition for notational variance. A number of authors in the literature on translations have claimed that translational equivalence between two logics is also a sufficient condition for deeming them notational variants.[9]

Unfortunately, even isomorphism, without further constraints, is relatively uninformative. As we will now show, it turns out that quite a few of the most common logics can be translated into classical propositional logic.[10] Perhaps the most striking example of this is that classical first-order logic (along with a number of nonclassical logics) is isomorphic to classical propositional logic. This may seem

[8]Of course, we may want to replace translatability with interpretability. While this will not affect the account of notational variance (see **Proposition 2.4.7**), it will weaken the notions of expressive inclusion.

[9]For claims like this, see Straßburger 2007, p. 139, Mossakowski et al. 2009, p. 7, and French 2010, p. 134. Straßburger [2007, p. 139] claims translational equivalence is necessary and sufficient. Mossakowski et al. [2009, p. 7] holds that even mere intertranslatability suffices for expressive equivalence. As we argued in § 2.5, this is compatible with the view that intertranslatability is not sufficient for notational variance; but it is unclear whether Mossakowski et al. would agree with the distinction.

[10]*Pace* Carnielli et al. [2009, p. 7], who say "Translations into CPL seem to be hard to obtain."

somewhat paradoxical, since first-order logic is obviously more expressively powerful than (and thus, not a notational variant of) propositional logic. What this result shows, then, is that translational equivalence is simply too unconstrained to serve as an adequate precisification of the concept of notational variance.

There are two ways of going about showing that classical first-order logic **FOL** is translatable into propositional logic **CPL**. A non-constructive algebraic approach is given in § 2.6.1 using Lindenbaum-Tarski algebras. A more constructive approach was given in Jeřábek [2012] and will be reviewed in § 2.6.2. The former has the advantage of showing how **FOL** is isomorphic to **CPL**, whereas the latter has the advantage of constructing a translation that is as computationally efficient as possible.

§ 2.6.1 Algebraic Reduction

We start by giving the algebraic proof that **FOL** $\cong$ **CPL**. To begin, recall the definition of the Lindenbaum-Tarski algebra of a logic (**Definition 1.1.38**).

Proposition 2.6.1 *(Translational Equivalence Implies Lindenbaum-Tarski Isomorphism)*. Let L_1 and L_2 be logics.

(a) If $\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, then $\mathbb{L}_1 \hookrightarrow \mathbb{L}_2$.

(b) If $\mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$, then $\mathbb{L}_1 \cong \mathbb{L}_2$.

Proof:

(a) Let $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$. Define $\mathtt{t}^{\mathbb{L}}\colon \mathcal{L}_1/\dashv\vdash_1 \to \mathcal{L}_2/\dashv\vdash_2$ as follows:

$$\mathtt{t}^{\mathbb{L}}([\phi]_1) = [\mathtt{t}(\phi)]_2 \,.$$

This is a well-defined injective function, since:

$$\begin{aligned} [\phi]_1 = [\psi]_1 \quad &\Leftrightarrow \quad \phi \dashv\vdash_1 \psi \\ &\Leftrightarrow \quad \mathtt{t}(\phi) \dashv\vdash_2 \mathtt{t}(\psi) \\ &\Leftrightarrow \quad [\mathtt{t}(\phi)]_2 = [\mathtt{t}(\psi)]_2 \\ &\Leftrightarrow \quad \mathtt{t}^{\mathbb{L}}([\phi]_1) = \mathtt{t}^{\mathbb{L}}([\psi]_1). \end{aligned}$$

Moreover:

$$\begin{aligned} [\phi]_1 \leqslant_1 [\psi]_1 \quad &\Leftrightarrow \quad \phi \vdash_1 \psi \\ &\Leftrightarrow \quad \mathtt{t}(\phi) \vdash_2 \mathtt{t}(\psi) \\ &\Leftrightarrow \quad [\mathtt{t}(\phi)]_2 \leqslant_2 [\mathtt{t}(\psi)]_2 \\ &\Leftrightarrow \quad \mathtt{t}^{\mathbb{L}}([\phi]_1) \leqslant_2 \mathtt{t}^{\mathbb{L}}([\psi]_1). \end{aligned}$$

Hence, $\mathtt{t}^{\mathbb{L}}\colon \mathbb{L}_1 \hookrightarrow \mathbb{L}_2$.

(b) Let $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$. It is a basic fact from order theory that if $f\colon \mathbb{P}_1 \hookrightarrow \mathbb{P}_2$ and $g\colon \mathbb{P}_2 \hookrightarrow \mathbb{P}_1$ where $g(f(x)) = x$ and $f(g(y)) = y$ for all $x \in P_1$ and all $y \in P_2$, then $f\colon \mathbb{P}_1 \cong \mathbb{P}_2$ where $g = f^{-1}$. So it suffices to show that where $\mathtt{t}^{\mathbb{L}}$ and $\mathtt{s}^{\mathbb{L}}$ are defined as in (a), $\mathtt{s}^{\mathbb{L}}(\mathtt{t}^{\mathbb{L}}([\phi]_1)) = [\phi]_1$ and $\mathtt{t}^{\mathbb{L}}(\mathtt{s}^{\mathbb{L}}([\psi]_2)) = [\psi]_2$ for any $\phi \in \mathcal{L}_1$ and $\psi \in \mathcal{L}_2$. But this is immediate, since $\mathtt{s}^{\mathbb{L}}(\mathtt{t}^{\mathbb{L}}([\phi]_1)) = \mathtt{s}^{\mathbb{L}}([\mathtt{t}(\phi)]_2) = [\mathtt{s}(\mathtt{t}(\phi))]_1$ and $\mathtt{s}(\mathtt{t}(\phi)) \dashv\vdash_1 \phi$. ∎

Corollary 2.6.2 *(Preserving Number of Formulas Up to Equivalence).* Let L_1 and L_2 be logics. If $\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, then $|\mathcal{L}_1/\dashv\vdash_1| \leqslant |\mathcal{L}_2/\dashv\vdash_2|$, i.e., the number of $\mathcal{L}_1$-formulas up to L_1-equivalence is no greater than the number of $\mathcal{L}_2$-formulas up to L_2-equivalence.

Straßburger [2007, p. 139] proposed that we treat two logics as equivalent just in case their Lindenbaum-Tarski algebras were isomorphic. This seems too quick, however. A Lindenbaum-Tarski algebra only really represents the inferences in a logic that have just one premise. In some special cases, inference can be reduced to single-premise inference (e.g., if the logic is compact and conjunctive in the sense of **Definitions 1.2.12** and **1.2.24**). But this is generally not the case.[11] For instance, as is easily verified by **Fact 2.6.6**, **SOL** and **FOL** have isomorphic Lindenbaum-Tarski algebras, even though we do not even have **SOL** $\not\rightsquigarrow$ **FOL** (**Example 2.1.13**). So isomorphism on Lindenbaum-Tarski algebras seems far too weak. A better criteria would be isomorphism on theory spaces, as that is the notion that is captured by translational equivalence (**Proposition 2.2.10**).

With that said, we mentioned in § 1.2.3 that sometimes a logic can be reduced to its Lindenbaum-Tarski algebra. Recall the definitions of (semi)conjunctive and adjunctive logics from § 1.2.3 (**Definitions 1.2.21** and **1.2.24**).

Lemma 2.6.3 *(Translations Preserve Meets).* Let L_1 and L_2 be semiconjunctive logics.

(a) If $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, then $\mathtt{t}(\bigwedge \Gamma) \vdash_2 \bigwedge \mathtt{t}[\Gamma]$ for any finite $\Gamma \subseteq \mathcal{L}_1$.

(b) If $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$, then L_1 is adjunctive iff it is both the case that $\mathtt{t}[\mathsf{L}_1]$ is adjunctive and that $\mathtt{t}(\bigwedge \Gamma) \dashv\vdash_2 \bigwedge \mathtt{t}[\Gamma]$ for any finite $\Gamma \subseteq \mathcal{L}_1$.

(c) If $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$, then $\mathtt{t}(\bigwedge \Gamma) \dashv\vdash_2 \bigwedge \mathtt{t}[\Gamma]$ for any finite $\Gamma \subseteq \mathcal{L}_1$.

If L_1 and L_2 are completely semiconjunctive, then we can drop "finite" from (a)–(c).

[11] Straßburger's claim seems especially surprising, since he explicitly states he wants a definition of notational variance that will apply to non-compact logics as well.

Proof: Throughout, observe that $\bigwedge \Gamma \vdash_1 \Gamma$ and $\bigwedge \mathtt{t}[\Gamma] \vdash_2 \mathtt{t}[\Gamma]$.

(a) Since $[\bigwedge \Gamma]_1 \leqslant_1 \bigwedge_{\gamma \in \Gamma} [\gamma]_1$, it follows that $\mathtt{t}(\bigwedge \Gamma) \leqslant_2 \mathtt{t}[\Gamma]$. Thus, we have that $[\mathtt{t}(\bigwedge \Gamma)]_2 \leqslant_2 [\mathtt{t}(\gamma)]_2$ for all $\gamma \in \Gamma$. By definition of meets, that means that $[\mathtt{t}(\bigwedge \Gamma)]_2 \leqslant_2 \bigwedge_{\gamma \in \Gamma} [\mathtt{t}(\gamma)]_2 = [\bigwedge \mathtt{t}[\Gamma]]_2$. So it follows that $\mathtt{t}(\bigwedge \Gamma) \vdash_2 \bigwedge \mathtt{t}[\Gamma]$.

(b) First, the left-to-right direction. By the above, $\mathtt{t}(\bigwedge \Gamma) \vdash_2 \bigwedge \mathtt{t}[\Gamma]$. Since L_1 is adjunctive, $\Gamma \vdash_1 \bigwedge \Gamma$, and so $\mathtt{t}[\Gamma] \vdash_{\mathtt{t}[1]} \mathtt{t}(\bigwedge \Gamma) \vdash_{\mathtt{t}[1]} \bigwedge \mathtt{t}[\Gamma]$. Thus, $\mathtt{t}[\mathsf{L}_1]$ is adjunctive. Moreover, since $\Gamma \vdash_1 \bigwedge \Gamma$, $\mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\bigwedge \Gamma)$. But $\bigwedge \mathtt{t}[\Gamma] \vdash_2 \mathtt{t}[\Gamma]$. So $\bigwedge \mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\bigwedge \Gamma)$, i.e., $\mathtt{t}(\bigwedge \Gamma) \dashv\vdash_2 \bigwedge \mathtt{t}[\Gamma]$.

For the right-to-left direction, since $\mathtt{t}[\Gamma] \vdash_2 \bigwedge \mathtt{t}[\Gamma] \dashv\vdash_2 \mathtt{t}(\bigwedge \Gamma)$, it immediately follows that $\Gamma \vdash_1 \bigwedge \Gamma$.

(c) Suppose $\mathtt{t}, \mathtt{s} \colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$. As before, $\mathtt{t}(\bigwedge \Gamma) \vdash_2 \bigwedge \mathtt{t}[\Gamma]$ and $\mathtt{s}(\bigwedge \Gamma) \vdash_1 \bigwedge \mathtt{s}[\Gamma]$. This gives us that $\mathtt{s}(\bigwedge \mathtt{t}[\Gamma]) \vdash_1 \bigwedge \mathtt{s}[\mathtt{t}[\Gamma]]$. Thus, we have that $[\mathtt{s}(\bigwedge \mathtt{t}[\Gamma])]_1 \leqslant_1 [\bigwedge \mathtt{s}[\mathtt{t}[\Gamma]]]_1 = \bigwedge_{\gamma \in \Gamma} [\mathtt{s}(\mathtt{t}(\gamma))]_1 = \bigwedge_{\gamma \in \Gamma} [\gamma]_1 = [\bigwedge \Gamma]_1$. That means $\mathtt{s}(\bigwedge \mathtt{t}[\Gamma]) \vdash_1 \bigwedge \Gamma$. So $\mathtt{t}(\mathtt{s}(\bigwedge \mathtt{t}[\Gamma])) \dashv\vdash_2 \bigwedge \mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\bigwedge \Gamma)$, as desired. ∎

Corollary 2.6.4 *(Translatability Implies Meet-Homomorphism for Conjunctive Logics)*. Suppose L_1 and L_2 are logics where L_1 is conjunctive and where $\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$. Then there is a $f \colon \mathbb{L}_1 \hookrightarrow \mathbb{L}_2$ that preserves finite meets.

Not every order-embedding preserves meets, so **Corollary 2.6.4** does not simply follow from **Proposition 2.6.1**(a). The point is just that under the assumption that L_1 is adjunctive, we can take the relevant order-embedding to preserve finite meets, given that the relevant meets are defined in the target logic.

We now return to the original objective, viz., proving that **FOL** $\cong$ **CPL**. To do this, we prove a more general result—essentially the informal remark we made in § 1.2.3.

Theorem 2.6.5 *(Algebraic Reduction of Translatability)*. Suppose L_1 and L_2 are compact conjunctive logics.

(a) If there is a $f \colon \mathbb{L}_1 \hookrightarrow \mathbb{L}_2$ that preserves finite meets, then $\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$.

(b) If $\mathbb{L}_1 \cong \mathbb{L}_2$, then $\mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$.

(c) If $f \colon \mathbb{L}_1 \cong \mathbb{L}_2$ and $|[\phi]_1| = |f([\phi]_1)|$ for each $\phi \in \mathcal{L}_1$, then $\mathsf{L}_1 \cong \mathsf{L}_2$.

Proof:

(a) For each $[\phi]_1 \in \mathcal{L}_1/\dashv\vdash_1$, let $f_{[\phi]_1}\colon [\phi]_1 \to f([\phi]_1)$ be an arbitrary map. Define $\mathtt{t}(\phi) = f_{[\phi]_1}(\phi)$. Since L_1 is compact, $\Gamma \vdash_1 \phi$ iff for some finite $\Gamma' \subseteq \Gamma$, $\Gamma' \vdash_1 \phi$. And if Γ' is finite, then $\Gamma' \vdash_1 \phi$ iff $\bigwedge \Gamma' \vdash_1 \phi$ since L_1 is adjunctive. Likewise, $\mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi)$ iff $\mathtt{t}[\Gamma'] \vdash_2 \mathtt{t}(\phi)$ for some finite $\Gamma' \subseteq \Gamma$, iff $\bigwedge \mathtt{t}[\Gamma'] \vdash_2 \mathtt{t}(\phi)$. Since f preserves meets, $\bigwedge_{\gamma\in\Gamma'} f([\gamma]_1) = f(\bigwedge_{\gamma\in\Gamma'} [\gamma]_1) = f([\bigwedge \Gamma']_1)$. Thus, $\bigwedge_{\gamma\in\Gamma'} \mathtt{t}(\gamma) \dashv\vdash_2 \mathtt{t}(\bigwedge \Gamma')$. So it suffices to show that for any $\phi, \psi \in \mathcal{L}_1$, $\phi \vdash_1 \psi$ iff $\mathtt{t}(\phi) \vdash_2 \mathtt{t}(\psi)$. Indeed:

$$\begin{aligned}
\phi \vdash_1 \psi \quad &\Leftrightarrow \quad [\phi]_1 \leqslant_1 [\psi]_1 \\
&\Leftrightarrow \quad f([\phi]_1) \leqslant_2 f([\psi]_1) \\
&\Leftrightarrow \quad \mathtt{t}(\phi) \vdash_2 \mathtt{t}(\psi).
\end{aligned}$$

So $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$.

(b) Let $f\colon \mathbb{L}_1 \cong \mathbb{L}_2$. For each $[\phi]_1 \in \mathcal{L}_1/\dashv\vdash_1$, let $f_{[\phi]_1}\colon [\phi]_1 \to f([\phi]_1)$ be an arbitrary map. Likewise, for each $[\psi]_2 \in \mathcal{L}_2/\dashv\vdash_2$, let $g_{[\psi]_2}\colon [\psi]_2 \to f^{-1}([\psi]_2)$ be arbitrary. Define $\mathtt{t}(\phi) = f_{[\phi]_1}(\phi)$ and $\mathtt{s}(\psi) = g_{[\psi]_2}(\psi)$. By the above reasoning, $\mathtt{t}$ and $\mathtt{s}$ are translations. Now suppose $\phi \in \mathcal{L}_1$. Then $\phi \dashv\vdash_1 \mathtt{s}(\mathtt{t}(\phi))$ iff $[\phi]_1 = [\mathtt{s}(\mathtt{t}(\phi))]_1$. But:

$$[\mathtt{s}(\mathtt{t}(\phi))]_1 = f^{-1}([\mathtt{t}(\phi)]_2) = f^{-1}(f([\phi]_1)) = [\phi]_1 .$$

So $\phi \dashv\vdash_1 \mathtt{s}(\mathtt{t}(\phi))$ for all $\phi \in \mathcal{L}_1$. Likewise, $\psi \dashv\vdash_1 \mathtt{t}(\mathtt{s}(\psi))$ for all $\psi \in \mathcal{L}_2$.

(c) Under these conditions, we can simply take each $f_{[\phi]_1}$ from (a) to be bijective. ∎

To complete the proof, we need some facts about Boolean algebras and some facts about **FOL** and **CPL**:

Fact 2.6.6 *(Facts about **CPL** and **FOL**).*

(a) Any two countable atomless Boolean algebras are isomorphic.

(b) $\mathbb{L}_{\mathbf{CPL}}$ and $\mathbb{L}_{\mathbf{FOL}}$ are countable atomless Boolean algebras.

(c) For each $\phi \in \mathcal{L}_{\mathtt{Pred}}$ and $\psi \in \mathcal{L}_{\mathtt{Prop}}$, $|[\phi]_{\mathbf{FOL}}| = |[\psi]_{\mathbf{CPL}}| = \aleph_0$.

Corollary 2.6.7 *(Triviality)*. **FOL** $\cong$ **CPL**.

What is more, where **L** is a normal modal logic, $\mathbb{L}_{\mathbf{L}}$ is also a countable atomless Boolean algebra, so $\mathbf{L} \cong \mathbf{CPL}$. Isomorphism to **CPL** is difficult to escape for classical languages.

§ 2.6.2 Jeřábek's Proof

The proof of **Theorem 2.6.5** crucially depends on the adjunctivity and semilattice constraints, as they allowed us to distill a logic down to its Lindenbaum-Tarski algebra. These constraints are fairly benign, since they more-or-less amount to the presence of conjunction introduction and elimination. But we can relax even these constraints and still show that $\mathbf{L} \rightsquigarrow \mathbf{CPL}$ using a technique developed by Jeřábek [2012]. In this section, we review the proof of this result.

Theorem 2.6.8 *(Jeřábek)*. Let **L** be compact where $|\mathcal{L}| \leqslant \aleph_0$. Then there is a $\mathtt{t}\colon \mathbf{L} \rightsquigarrow \mathbf{CPL}$ such that $\mathtt{t}$ is Turing-equivalent to $\vDash_{\mathbf{L}}$.

Proof: Let **L** be compact and countable. If the logic is trivial, meaning that $\vDash \phi$ for all $\phi \in \mathcal{L}$, then we can just map every formula to $\top$. Otherwise, list the elements of $\mathcal{L}$ as $\phi_0, \phi_1, \phi_2, \ldots$. We will define our translation $\mathtt{t}$ in stages. To simplify, we will treat natural numbers as sets in the usual way ($n = \{0, \ldots, n-1\}$), and where $A \subseteq n$, we will write "ϕ_A" for "$\{\phi_i \mid i \in A\}$". I will use "$\vDash$" for "$\vDash_{\mathbf{CPL}}$" throughout.

Suppose $\mathtt{t}(\phi_0), \ldots, \mathtt{t}(\phi_{n-1})$ have been defined in such a way so that the following condition holds for all $A \subseteq n$ and $k < n$:

$$\phi_A \vDash_{\mathbf{L}} \phi_k \quad \Rightarrow \quad \mathtt{t}[\phi_A] \vDash \mathtt{t}[\phi_k].$$

Then define the following $\mathcal{L}_{\mathbf{Prop}}$-formulas (setting $\bigvee \varnothing = \bot$ and $\bigwedge \varnothing = \top$):

$$\alpha_n := \bigwedge_{\substack{A \subseteq n, k < n \\ \phi_A, \phi_n \vDash_{\mathbf{L}} \phi_k}} \left(\bigwedge \mathtt{t}[\phi_A] \to \mathtt{t}(\phi_k) \right)$$

$$\gamma_n := \bigvee_{\substack{A \subseteq n \\ \phi_A \vDash_{\mathbf{L}} \phi_n}} \bigwedge \mathtt{t}[\phi_A].$$

Finally, define $\mathtt{t}(\phi_n) := \alpha_n \wedge (p_n \vee \gamma_n)$.

Claim: $\gamma_n \vDash \alpha_n$.

Subproof: Consider any disjunct from γ_n of the form $\bigwedge \mathtt{t}[\phi_A]$ where $\phi_A \vDash_{\mathbf{L}} \phi_n$, and consider any conjunct form $\bigwedge \mathtt{t}[\phi_B] \to \mathtt{t}(\phi_k)$ where $\phi_B, \phi_n \vDash_{\mathbf{L}} \phi_k$. By transitivity, $\phi_B, \phi_A \vDash_{\mathbf{L}} \phi_k$. But then by assumption, $A, B \subseteq n$ and $k < n$, $\mathtt{t}[\phi_A], \mathtt{t}[\phi_B] \vDash \mathtt{t}(\phi_k)$, i.e., $\bigwedge \mathtt{t}[\phi_A] \vDash \bigwedge \mathtt{t}[\phi_B] \to \mathtt{t}(\phi_k)$. Hence, each disjunct from γ_n implies each conjunct from α_n; so $\gamma_n \vDash \alpha_n$. ∎

By truth-functional reasoning, it follows that $\mathtt{t}(\phi_n) \equiv (\alpha_n \wedge p_n) \vee \gamma_n$.

Claim: For all $A \subseteq n+1$ and $k < n+1$, if $\phi_A \vDash_{\mathbf{L}} \phi_k$, then $\mathtt{t}[\phi_A] \vDash \mathtt{t}[\phi_k]$.

Subproof: We only need to check implications that involve ϕ_n. Thus, it suffices to check that for any $A \subseteq n$, $\phi_A \vDash_{\mathbf{L}} \phi_n$ implies $\mathtt{t}[\phi_A] \vDash \mathtt{t}(\phi_n)$, and for any $A \subseteq n$ and any $k < n$, $\phi_A, \phi_n \vDash_{\mathbf{L}} \phi_k$ implies $\mathtt{t}[\phi_A], \mathtt{t}(\phi_n) \vDash \phi_k$.

Suppose $\phi_A \vDash_{\mathbf{L}} \phi_n$. Then $\bigwedge \mathtt{t}[\phi_A] \vDash \gamma_n$. Hence, $\mathtt{t}[\phi_A] \vDash \mathtt{t}(\phi_n)$. Now instead suppose $\phi_A, \phi_n \vDash_{\mathbf{L}} \phi_k$. Then $\alpha_n \vDash \bigwedge \mathtt{t}[\phi_A] \to \mathtt{t}(\phi_k)$. So $\mathtt{t}(\phi_n) \vDash \bigwedge \mathtt{t}[\phi_A] \to \mathtt{t}(\phi_k)$. Hence, $\mathtt{t}[\phi_A], \mathtt{t}(\phi_n) \vDash \mathtt{t}(\phi_k)$. ∎

We will now show that $\Gamma \vDash_{\mathbf{L}} \phi$ iff $\mathtt{t}[\Gamma] \vDash_{\mathbf{CPL}} \mathtt{t}(\phi)$. First, suppose $\Gamma \vDash_{\mathbf{L}} \phi$. By compactness, there's a finite $\Gamma' \subseteq \Gamma$ such that $\Gamma' \vDash_{\mathbf{L}} \phi$. If $\phi = \phi_n$ occurs in the enumeration after each element in Γ', then γ_n will contain a disjunct of the form $\bigwedge \mathtt{t}[\Gamma']$. Hence, $\mathtt{t}[\Gamma'] \vDash \gamma_n$, and so $\mathtt{t}[\Gamma'] \vDash \mathtt{t}(\phi)$. If instead some element in Γ occurs in the enumeration after ϕ, then where ϕ_n is the last element in Γ to occur, α_n will contain a conjunct of the form $\bigwedge \mathtt{t}[\Gamma - \{\phi_n\}] \to \mathtt{t}(\phi)$. So still, $\mathtt{t}[\Gamma] \vDash \mathtt{t}(\phi)$.

Next, suppose $\Gamma \nvDash_{\mathbf{L}} \phi$. Define $\mathtt{Cn}_{\mathbf{L}}(\Gamma) := \{\psi \in \mathcal{L} \mid \Gamma \vDash_{\mathbf{L}} \psi\}$ and define:

$$V(p_i) = \begin{cases} 1 & \text{if } \phi_i \in \mathtt{Cn}_{\mathbf{L}}(\Gamma) \\ 0 & \text{otherwise.} \end{cases}$$

Claim: $V \Vdash \mathtt{t}(\phi_k)$ iff $V(p_k) = 1$.

Subproof: By induction on n. It's easy to verify for $n = 0$. Suppose for inductive hypothesis that the claim holds for all $k < n$.

Suppose first $V(p_n) = 1$, so that $\phi_n \in \mathtt{Cn}_{\mathbf{L}}(\Gamma)$. Then it suffices to show that $V \Vdash \alpha_n$, i.e., that whenever $A \subseteq n$ and $k < n$ such that $\phi_A, \phi_n \vDash_{\mathbf{L}} \phi_k$, we have that $V \Vdash \bigwedge \mathtt{t}[\phi_A] \to \mathtt{t}(\phi_k)$. So suppose $\phi_A, \phi_n \vDash_{\mathbf{L}} \phi_k$. Either for some $i \in A$ we have $\phi_i \notin \mathtt{Cn}_{\mathbf{L}}(\Gamma)$, or else $\phi_i \in \mathtt{Cn}_{\mathbf{L}}(\Gamma)$ for each $i \in A$ (in which case $\Gamma, \phi_n \vDash \phi_k$) and $\phi_k \neq \phi$. If the former, then by the inductive hypothesis, $V \nVdash \mathtt{t}(\phi_i)$, in which case $V \Vdash \bigwedge \mathtt{t}[\phi_A] \to \mathtt{t}(\phi_k)$. If the latter, then since $\phi_n \in \mathtt{Cn}_{\mathbf{L}}(\Gamma)$, that means $\Gamma \vDash_{\mathbf{L}} \phi_k$, so that $\phi_k \in \mathtt{Cn}_{\mathbf{L}}(\Gamma)$, i.e., $V(p_k) = 1$. So by inductive hypothesis, $V \Vdash \mathtt{t}(\phi_k)$, and hence, $V \Vdash \bigwedge \mathtt{t}[\phi_A] \to \mathtt{t}(\phi_k)$. Either way, $V \Vdash \bigwedge \mathtt{t}[\phi_A] \to \mathtt{t}(\phi_k)$.

Suppose instead that $V(p_n) = 0$, so that $\phi_n \notin \mathtt{Cn}_{\mathbf{L}}(\Gamma)$. It suffices to show that $V \nVdash \gamma_n$, i.e., for each $A \subseteq n$ such that $\phi_A \vDash_{\mathbf{L}} \phi_n$, $V \nVdash \bigwedge \mathtt{t}[\phi_A]$. So suppose $\phi_A \vDash_{\mathbf{L}} \phi_n$. Since $\phi_n \notin \mathtt{Cn}_{\mathbf{L}}(\Gamma)$, there must be some $i \in A$ such that $\phi_i \notin \mathtt{Cn}_{\mathbf{L}}(\Gamma)$. By inductive hypothesis, $V \nVdash \mathtt{t}(\phi_i)$, so $V \nVdash \bigwedge \mathtt{t}[\phi_A]$. ■

Since $\phi \notin \mathtt{Cn}_{\mathbf{L}}(\Gamma)$, we have that $V \Vdash \mathtt{t}[\Gamma]$ but $V \nVdash \mathtt{t}(\phi)$. So $\mathtt{t}[\Gamma] \nvDash \mathtt{t}(\phi)$. Hence, $\mathtt{t}$ is a translation. Since **CPL** is decidable, $\vDash_{\mathbf{L}}$ is Turing-reducible to $\mathtt{t}$. The fact that $\mathtt{t}$ is Turing-reducible to $\vDash_{\mathbf{L}}$ follows by the construction of $\mathtt{t}$ (which appeals to $\vDash_{\mathbf{L}}$ in the definition of α_n and γ_n). ■

This result shows that intertranslatability is simply too weak to really distinguish between many of the most interesting logics. In particular, **Theorem 2.6.8** suffices to show the following:

Corollary 2.6.9 *(Classical and Intuitionistic Logic)*. **IPL**$\rightsquigarrow$**CPL**.

Of course, such a proof does not establish translational equivalence between compact countable languages and **CPL** in general. For one thing, we already showed that **CPL** $\not\leftrightsquigarrow$ **IPL** (**Example 2.2.8**). But moreover, the construction in **Theorem 2.6.8** does not even produce a reversible translation scheme between **CPL** and itself. If, for instance, $\phi_0 = \bot$, then $\mathtt{t}(\phi_0) = \top \wedge (p_0 \vee \bot)$, and no reversible translation scheme from **CPL** to itself maps $\bot$ to a consistent $\mathcal{L}_{\mathtt{Prop}}$-formula.

Now, as the examples explored previously reveal, not every logic can be translated into **CPL**. And the notion of a translation does impose some constraints on the relationship between other nonclassical logics. Still, on account of **Theorem 2.6.8**, it is natural to inquire into stronger notions of equivalence between logics than mere intertranslatability.

§ 2.6.3 A Note on Computability

As we say in § 2.6.1, **FOL** and **CPL** are isomorphic (**Corollary 2.6.7**). But such a translation must be quite complex. In particular:

Fact 2.6.10 *(Undecidability of Translation).* If $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ is decidable, then $\vdash_1$ is Turing-reducible to $\vdash_2$.

Corollary 2.6.11 *(Complexity of Translation from **FOL** to **CPL**).* If $\mathtt{t}\colon$ **FOL**$\rightsquigarrow$**CPL**, then $\mathtt{t}$ is undecidable.

One might think that the way to define notational variance so as to avoid the problems above is to impose a computability constraint on translations. For instance, one might demand that notational variants be translationally equivalent via some *decidable* translations. Or one might instead require that the translations have to be *Turing-reducible* to the target language. These constraints would, of course, block **FOL** and **CPL** from being considered notational variants.

But this proposal is undesirable for a number of reasons. For starters, such an approach would not allow notational variance between logics with uncountably many formulas, since according to the standard definitions, computability implies countability. Thus, **CPL**(@) from **Example 2.3.4** (with uncountably many redundant $@_r$ operators) would not count as equivalent to **CPL**, as counterintuitive as that is. In addition, requiring the decidability (or Turing-reducibility) of translations would not block all potential counterexamples. For instance, monadic first-order logic, being decidable, is isomorphic to propositional logic by **Theorem 2.6.5**. Yet it strikes me as highly undesirable to deem monadic first-order logic a “notational variant” of propositional logic.

Finally, it does not seem to be part of the concept of a notational variant that they have the same complexity: it seems quite intuitive that there could be more computationally efficient versions of the same system.

Example 2.6.12 *(An Undecidable Notational Variant of **CPL**).* Throughout, let $X \subseteq \mathbb{N}$ be some undecidable set. Let $\mathcal{L}_{\mathtt{Prop}}(?)$ be the result of extending $\mathcal{L}_{\mathtt{Prop}}$ with binary connectives $?_0, ?_1, ?_2, \ldots$. Let **CPL**(?) be the extension of **CPL** with $\mathcal{L}_{\mathtt{Prop}}(?)$, where the semantic clause for $?_i$ is given as follows:

$$\begin{aligned} &\text{if } i \in X, \text{ then } v \Vdash_? \phi \mathbin{?_i} \psi \quad \Leftrightarrow \quad v \Vdash_? \phi \text{ and } v \Vdash_? \psi \\ &\text{if } i \notin X, \text{ then } v \Vdash_? \phi \mathbin{?_i} \psi \quad \Leftrightarrow \quad v \Vdash_? \phi \text{ or } v \Vdash_? \psi. \end{aligned}$$

It is easy to verify that **CPL**(?) $\cong$ **CPL** by **Theorem 2.6.5**. Moreover, **CPL**(?) is undecidable. For if it were decidable, we could decide membership in X by deciding whether $p \vDash_? p \mathbin{?_i} q$.

Intuitively, **CPL**(?) is a notational variant of **CPL**—after all, each $?_i$ is simply *definable* in **CPL**. It is just that the way **CPL**(?) presents **CPL** is not computationally tractable. But it is not as though **CPL**(?) can represent more inferential connections that **CPL**: **CPL**(?) is a mere relabelling of **CPL**. So one must take caution when appealing to decidability to avoid the triviality results above.

Chapter 3

Compositionality

It is not as though anything goes when it comes to translations: not every logic is translatable into any other, as **Examples 2.1.9–2.1.14** illustrate. Even so, results such as **Theorems 2.6.5** and **2.6.8** show that translational equivalence is still fairly easy to come by and so cannot be the correct formalization of our intuitive concept of notational variance. We need more stringent requirements to find more adequate formalizations.

To help visualize **Theorem 2.6.5** and where we might successfully constrain the notion of a translation, it is informative to compare very simple versions of **FOL** and **CPL**. Figure 3.1 contains the Lindenbaum-Tarski algebras for (a) the one-variable fragment of **FOL** with just one unary predicate F, which we will label "$\mathbf{FOL}^F$", and (b) propositional logic with two atomics, i.e., $\mathbf{CPL}^{\{p,q\}}$ (**Example 2.1.9**). One can readily verify visually that these two Boolean algebras are isomorphic.[1] The only difference is that the Lindenbaum-Tarski algebra for $\mathbf{FOL}^F$ is equipped with an additional operator $[\forall x\,]$, where:

$$[\forall x\,]([\phi]) = \begin{cases} [\phi] & \text{if } \neg\, q \vDash_{\mathbf{CPL}} \phi \\ [\phi \wedge q] & \text{otherwise.} \end{cases}$$

Notice the definition of $[\forall x\,]$ is given by cases. This feature of the definition is ineliminable: there is no single propositional schema $\Theta(\pi)$[2] such that $[\forall x\,]([\phi]) = [\Theta(\phi)]$. If there were such a Θ, then $[\top] = [\forall x\,]([\top]) = [\Theta(\top)]$. So then $\phi \vDash_{\mathbf{CPL}} \Theta(\phi)$, even though $\Theta(p) \equiv_{\mathbf{CPL}} p \wedge q$ and $p \nvDash_{\mathbf{CPL}} p \wedge q$.

[1]Admittedly, this simple example is not representative. Though we will still have an isomorphism between the Lindenbaum-Tarski algebras of the one-variable fragment **FOL** with m-many predicates and $\mathbf{CPL}^{\{p_1,\dots,p_{4^m}\}}$, the Lindenbaum-Tarski algebra of monadic **FOL** with two or more variables will only be embeddable in the Lindenbaum-Tarski algebra of some $\mathbf{CPL}^{\{p_1,\dots,p_n\}}$. Still, in the infinite case, this difference washes out, and we obtain the isomorphism again, in accordance with **Theorem 2.6.5**.

[2]This notation was explained in **Definition 1.3.9**.

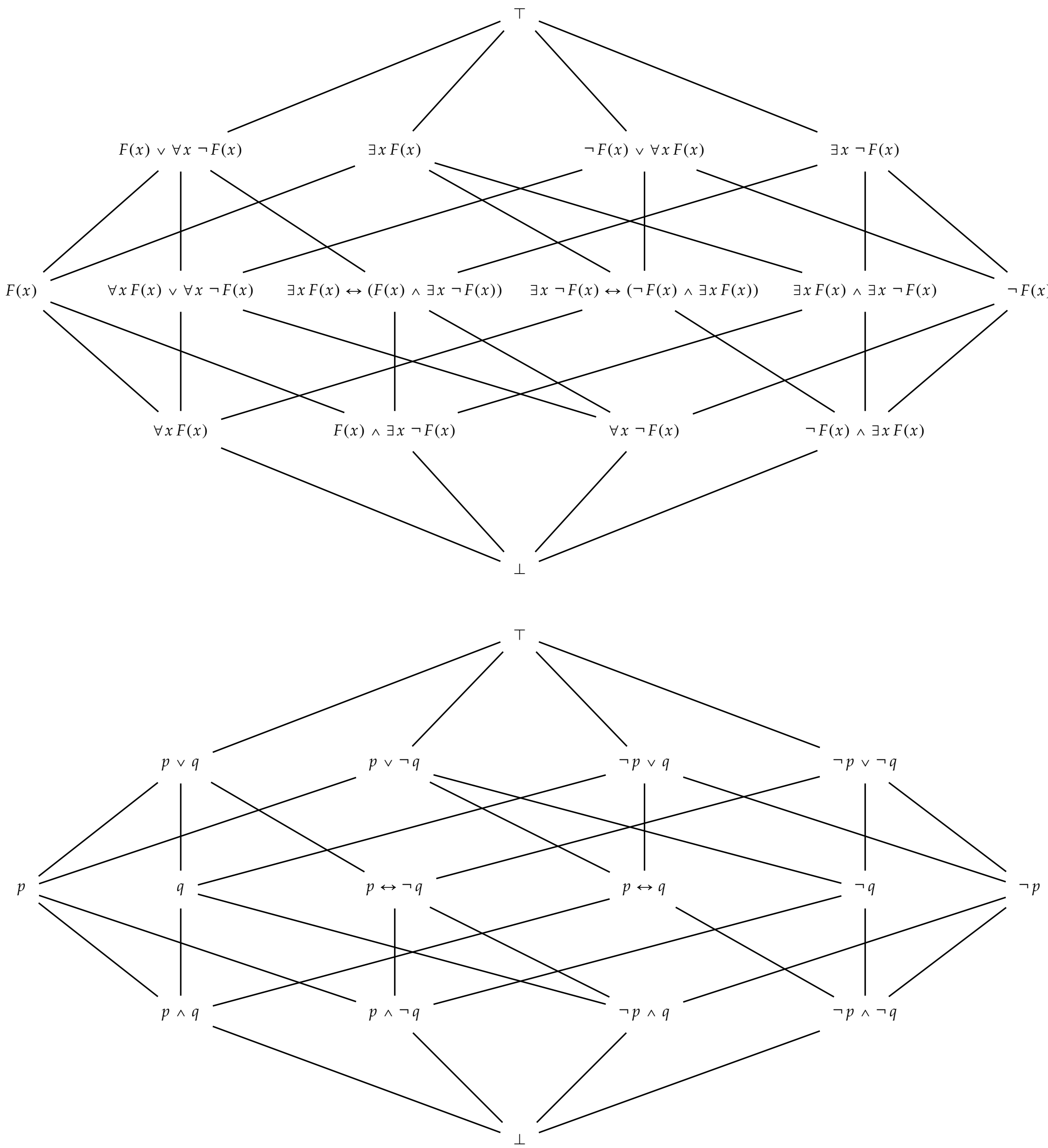


Figure 3.1: Monadic one-variable first-order logic vs. propositional logic.

The fact that $[\forall x\ ]$ cannot be represented as a single schema should give us a clue as to why notions like translational equivalence fall short of notational variance. After all, the notion of a translation as defined in **Definition 2.1.1** is fairly minimal. In theory, a translation could be quite gerrymandered and complex, mapping quantified formulas of the same sort to various different places to achieve a precarious balance. In practice, most translations that have been studied in the literature are fairly uniform—or, in other terms, *compositional*. That is, one defines the translation of a complex formula as some fixed schema of the translation of the parts. This ensures, amongst other things, that translations not only preserve the meanings of whole sentences but also preserve the meanings of the operators used to form complex formulas.

In this chapter, we will explore various ways of fleshing out this idea more precisely. A brief outline is as follows. § 3.1 discusses the simplest definition of compositionality whereby the translation of a complex formula is a function of the translation of the parts. It is shown that this constraint is rather easy to satisfy. In the opposite direction, § 3.2 explores a very strong sense of compositionality that has been considered in the literature. Then in § 3.3, we come to a fairly common definition of compositionality in terms of definition: a translation is compositional if each operator of the source language is definable in the target language. This notion is generalized in § 3.4 to cover a broader class of examples.

For this chapter, it will useful to review some of the more basic definitions of signatures and schemas in § 1.3. From now on, when we use the term "logic" or "language", we will intend these definitions to be read as implicitly carrying a signature. Thus, "logic" means "Σ-logic", where Σ is contextually supplied (and likewise for "language"). We will use the standard labels for the logics that have been discussed previously to stand for the logics with the appropriate signature. For example, we will use "**CPL**" to stand for the language of classical propositional logic with signature $\langle \mathtt{Prop}, \mathtt{Bool} \rangle$, where:

- $\mathtt{Prop} = \{p_1, p_2, p_3, \ldots\}$
- $\mathtt{Bool} = \{\neg, \wedge\}$.

Likewise, we will use "**FOL**" to stand for the language of classical first-order logic with signature $\langle \mathtt{At}, \mathtt{Bool} \cup \mathtt{Quant} \rangle$, where:

- $\mathtt{At} = \{P^n(y_1, \ldots, y_n) \mid P^n \in \mathtt{Pred}^n \text{ and } y_1, \ldots, y_n \in \mathtt{Var}\}$
- $\mathtt{Quant} = \{\forall x \mid x \in \mathtt{Var}\}$.

The signature will be specified explicitly if it is not clear from context.

We generalize the notion of a translation to logics and languages with signatures as follows: a map $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is a translation from a Σ_1-logic L_1 to a Σ_2-logic L_2 if it is a translation from $\langle \mathcal{L}_{\Sigma_1}, \vdash_1 \rangle$ to $\langle \mathcal{L}_{\Sigma_1}, \vdash_2 \rangle$. Similarly for the other definitions stated in Chapter 2. It is easy to check that all of the results from Chapter 2 still hold even when we redefine things in this way: nothing depended on the fact that the logics were unstructured in particular.

§ 3.1 Functionality

Plausibly, an adequate translation from one logic to another ought to at least respect the syntactical structure of the former in the following sense:

Definition 3.1.1 *(Functionality)*. Let L_1 and L_2 be logics (possibly with partial signatures). We say $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is ***functional*** if for each $\triangle \in \mathtt{Op}_1^{\gamma}$, there is a function $\mathtt{t}(\triangle)\colon \mathcal{L}_2^{\gamma} \to \mathcal{L}_2$ such that for all $\rho \in \mathcal{L}_1^{\gamma}$:

(i) $\mathtt{t}(\triangle)(\mathtt{t} \circ \rho)$ is defined iff $\triangle(\rho)$ is defined, and

(ii) when defined, $\mathtt{t}(\triangle(\rho)) = \mathtt{t}(\triangle)(\mathtt{t} \circ \rho)$.

We say that a translation scheme is ***functional*** if each translation in the pair is functional.

Functional translations are ones we can specify by (i) stating how the translation acts on the atomics of our source logic, and (ii) providing the "translation" of each of the operators of the source logic in the target logic as specifying a function of the translations of the components. In practice, this is how most translation translations are specified, as is evident when we write:

$$\begin{aligned}
\mathtt{t}(\chi) &= \cdots \\
\mathtt{t}(\neg\,\phi) &= \neg^{\mathtt{t}}(\mathtt{t}(\phi)) \\
\mathtt{t}(\phi \wedge \psi) &= \wedge^{\mathtt{t}}(\mathtt{t}(\phi), \mathtt{t}(\psi))
\end{aligned}$$

where χ is atomic and both $\neg^{\mathtt{t}}$ and $\wedge^{\mathtt{t}}$ are some functions. So initially, imposing functionality on a translation does not seem to be imposing much at all. Indeed, we can show that translatability from any logic with a simple property implies functional translatability. Recall the definition of intensionality (**Definition 1.3.13**).

Proposition 3.1.2 *(Intensional Translation Implies Functional Translation)*. Let L_1 and L_2 be logics.

(a) If $\mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ and if L_1 is intensional, then there is a functional $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$.

(b) If $\mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$ and if L_1 and L_2 are intensional, then there are some functional $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$.

(c) If $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is injective, then $\mathtt{t}$ is already functional.

Proof:

(a) Let $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$. Using the axiom of choice, let $f\colon (\mathcal{L}_1/\dashv\vdash_1) \to \mathcal{L}_1$ be a choice function such that $f([\phi]_1) \in [\phi]_1$. Define $\mathtt{t}'(\phi) = \mathtt{t}(f([\phi]_1))$ for all $\phi \in \mathcal{L}_1$. Then $\mathtt{t}'$ is automatically a translation, since $\mathtt{t}$ is a translation and $f([\phi]_1) \dashv\vdash_1 \phi$. Notice also that $\mathtt{t}'$ has the property that if $\phi \dashv\vdash_1 \psi$, then $\mathtt{t}'(\phi) = \mathtt{t}'(\psi)$.

We just need to show it is functional. Define $\mathtt{t}'(\triangle)(\mathtt{t}' \circ \rho) = \mathtt{t}'(\triangle(\rho))$ (never mind what it does to the elements outside of $\mathtt{t}'[\mathcal{L}_1]$). Clearly, $\mathtt{t}'(\triangle)(\mathtt{t}' \circ \rho)$ is defined iff $\triangle(\rho)$ is. Suppose $\mathtt{t}' \circ \rho = \mathtt{t}' \circ \rho'$. Then for each $\beta < \gamma$, $\mathtt{t}'(\rho(\beta)) = \mathtt{t}'(\rho'(\beta))$, and thus $\rho(\beta) \dashv\vdash_1 \rho'(\beta)$. By intensionality, that means: $\triangle(\rho) \dashv\vdash_1 \triangle(\rho')$. Hence:

$$\mathtt{t}'(\triangle)(\mathtt{t}' \circ \rho) = \mathtt{t}'(\triangle(\rho)) = \mathtt{t}'(\triangle(\rho')) = \mathtt{t}'(\triangle)(\mathtt{t}' \circ \rho').$$

(b) Define $\mathtt{t}'$ and $\mathtt{s}'$ as in (a) using choice functions $f\colon (\mathcal{L}_1/\dashv\vdash_1) \to \mathcal{L}_1$ and $g\colon (\mathcal{L}_2/\dashv\vdash_2) \to \mathcal{L}_2$. Since $\phi \dashv\vdash_1 f([\phi]_1)$, $\mathtt{t}(\phi) \dashv\vdash_2 \mathtt{t}(f([\phi]_1)) = \mathtt{t}'(\phi)$. Likewise, $\mathtt{s}(\psi) \dashv\vdash_1 \mathtt{s}'(\psi)$. Hence, $\phi \dashv\vdash_1 \mathtt{s}(\mathtt{t}(\phi)) \dashv\vdash_1 \mathtt{s}'(\mathtt{t}'(\phi))$. So $\mathtt{t}', \mathtt{s}'\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$.

(c) Just define $\mathtt{t}(\triangle)(\mathtt{t} \circ \rho) = \mathtt{t}(\triangle(\rho)) = \mathtt{t}(\triangle(\mathtt{t}^{-1} \circ \mathtt{t} \circ \rho))$. ∎

Not every translation can be made to be functional, however. It is possible for one logic to be translatable, but not functionally translatable, into another logic if the underlying logics are hyperintensional.

Example 3.1.3 *(Translatability Without Functional Translatability)*. Define the signature $\Sigma = \langle \{p, q\}, \{\triangle\} \rangle$, where $\triangle$ is unary. Throughout, we define $\triangle^0(\phi) = \phi$ and $\triangle^{n+1}(\phi) = \triangle(\triangle^n(\phi))$. Let $\mathsf{C} = \{x, y\}$ and define $\Vdash$ so that:

- $\mathtt{Diag}_{\mathbf{L}}(x) = \{q\} \cup \{\triangle^n(p) \mid n \in \mathbb{N}\}$
- $\mathtt{Diag}_{\mathbf{L}}(y) = \{p\} \cup \{\triangle^n(q) \mid n \in \mathbb{N}\}$

Set $\mathbf{L} = \langle \Sigma, \mathsf{C}, \Vdash \rangle$. Then $p \equiv_{\mathbf{L}} q$ (since $[\![p]\!]_{\mathbf{L}} = [\![q]\!]_{\mathbf{L}} = \mathsf{C}$) but $\triangle(p) \not\equiv_{\mathbf{L}} \triangle(q)$. Thus, $\mathbf{L}$ is hyperintensional in the sense of Definition 1.3.13.

Now, define $\Sigma^* = \langle \{[p]_{\mathbf{L}}\}, \{\triangle^*\} \rangle$, where $\triangle^*([\phi]_{\mathbf{L}}) = [\triangle(\phi)]_{\mathbf{L}}$, define C as before, and for $z \in \mathsf{C}$, define $z \Vdash^* [\phi]$ iff $z \in [\![\phi]\!]_{\mathbf{L}}$. Call the resulting language $\mathbf{L}^*$. Clearly, there is a $\mathtt{t}\colon \mathbf{L} \rightsquigarrow \mathbf{L}^*$ (for instance, $\mathtt{t}(\phi) = [\phi]_1$). Moreover, it must be that $\mathtt{t}(p) = \mathtt{t}(q) = [p]_{\mathbf{L}}$. But $\mathtt{t}(\triangle(p)) \neq \mathtt{t}(\triangle(q))$, since $\triangle(p) \not\equiv_{\mathbf{L}} \triangle(q)$. Yet if $\mathtt{t}$ were functional, then $\mathtt{t}(\triangle(p)) = \mathtt{t}(\triangle)(\mathtt{t}(p)) = \mathtt{t}(\triangle)(\mathtt{t}(q)) = \mathtt{t}(\triangle(q))$, $\lightning$. So $\mathbf{L}$ is not functionally translatable into $\mathbf{L}^*$.

Still, **Proposition 3.1.2**(c) already shows that having an injective translation suffices for functionality. Many of the translations that appear in practice are injective. In any case, these results clearly show that functionality is too weak to satisfactorily avoid the result that **FOL** and **CPL** are isomorphic (**Corollary 2.6.7**), since **Proposition 3.1.2** implies that isomorphic logics are already functionally isomorphic.

Even so, it is worth remarking how much functionality guarantees. For example, one might wonder whether the image of a functional translation preserves certain nice syntactic features. The answer turns out to be mixed.

Definition 3.1.4 *(Image with Signature)*. Let L_1 be a Σ_1-logic, let L_2 be a Σ_2-logic, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ be functional. Define $\mathtt{Op}_{\mathtt{t}[1]} = \{\mathtt{t}(\triangle) \mid \triangle \in \mathtt{Op}_1\}$, define $\Sigma_{\mathtt{t}[1]} = \langle \mathtt{t}[\mathtt{At}_1], \mathtt{Op}_{\mathtt{t}[1]}\rangle$, and define $\mathcal{L}_{\mathtt{t}[1]} \coloneqq \mathcal{L}_{\Sigma_{\mathtt{t}[1]}}$. The ***image*** of L_1 under $\mathtt{t}$ is the $\Sigma_{\mathtt{t}[1]}$-logic $\mathtt{t}[\mathsf{L}_1] = \langle \Sigma_{\mathtt{t}[1]}, \vdash_{\mathtt{t}[1]}\rangle$, where $\vdash_{\mathtt{t}[1]} = \vdash_2 \restriction_{\mathcal{L}_{\mathtt{t}[1]}}$. Similarly for the image of a language under $\mathtt{t}$.

One might worry about our overloading of the notation for images of maps from **Definitions 1.1.23** and **3.1.4**, since $\mathcal{L}_{\mathtt{t}[1]}$ is not defined to be $\mathtt{t}[\mathcal{L}_1]$. The following lemma alleviates such worries:

Lemma 3.1.5 *(The Functional Image is the Image)*. Let L_1 and L_2 be logics, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ be functional. Then $\mathcal{L}_{\mathtt{t}[1]} = \mathtt{t}[\mathcal{L}_1]$.

Proof: First, we show $\mathcal{L}_{\mathtt{t}[1]} \subseteq \mathtt{t}[\mathcal{L}_1]$. Suppose $\phi \in \mathcal{L}_{\mathtt{t}[1]}$. If $\phi \in \mathtt{t}[\mathtt{At}_1]$, then trivially $\phi \in \mathtt{t}[\mathcal{L}_1]$. Now suppose $\phi = \mathtt{t}(\triangle)(\mathtt{t} \circ \rho)$ for some $\rho \in \mathcal{L}_1^{\gamma}$. Then $\phi = \mathtt{t}(\triangle)(\mathtt{t} \circ \rho) = \mathtt{t}(\triangle(\rho)) \in \mathtt{t}[\mathcal{L}_1]$. So $\mathcal{L}_{\mathtt{t}[1]} \subseteq \mathtt{t}[\mathcal{L}_1]$.

Conversely, suppose $\phi \in \mathtt{t}[\mathcal{L}_1]$. If $\phi \in \mathtt{t}[\mathtt{At}_1]$, then $\phi \in \mathcal{L}_{\mathtt{t}[1]}$. If $\phi = \mathtt{t}(\triangle(\rho))$ where $\mathtt{t} \circ \rho \in \mathcal{L}_{\mathtt{t}[1]}^{\gamma}$, then $\phi = \mathtt{t}(\triangle(\rho)) = \mathtt{t}(\triangle)(\mathtt{t} \circ \rho) \in \mathcal{L}_{\mathtt{t}[1]}$. So $\mathcal{L}_{\mathtt{t}[1]} = \mathtt{t}[\mathcal{L}_1]$. ∎

Thus, there is no harm in overloading the notation "$\mathtt{t}[\mathsf{L}_1]$". Henceforth, we use "$\mathtt{t}[\mathsf{L}_1]$" in the sense of **Definition 3.1.4** unless otherwise stated.

Recall the definition of a total signature (**Definition 1.3.8**).

Fact 3.1.6 *(Functionality Preserves Intensionality and Totality)*. Let $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ be functional.

(a) L_1 is intensional iff $\mathtt{t}[\mathsf{L}_1]$ is intensional.

(b) L_1 is total iff $\mathtt{t}[\mathsf{L}_1]$ is total.

So functionality does guarantee the preservation of some interesting properties. Functionality does not immediately preserve unambiguity under images, however.

Example 3.1.7 *(Ambiguous Image of Unambiguous Language)*. Where $\mathbb{L}_{\mathbf{CPL}}$ is the Lindenbaum-Tarski algebra of **CPL**, define $\mathtt{At}_{\mathbb{L}} := \{[\phi]_{\mathbf{CPL}} \mid \phi \in \mathcal{L}_{\mathtt{Prop}}\}$, and define $\mathtt{Op}_{\mathbb{L}} := \{\sim, \sqcap\}$, where:

$$\begin{aligned} \sim [\phi]_{\mathbf{CPL}} &= [\neg \phi]_{\mathbf{CPL}} \\ [\phi]_{\mathbf{CPL}} \sqcap [\psi]_{\mathbf{CPL}} &= [\phi \wedge \psi]_{\mathbf{CPL}} . \end{aligned}$$

Our evaluation space is just V. Define $\Vdash_{\mathbb{L}}$ so that $v \Vdash_{\mathbb{L}} [\phi]_{\mathbf{CPL}}$ iff $v \Vdash_{\mathbf{CPL}} \phi$. Call the resulting language $\mathbb{L}^*_{\mathbf{CPL}}$. Let $\mathtt{t}\colon \phi \mapsto [\phi]_{\mathbf{CPL}}$ for $\phi \in \mathcal{L}_{\mathtt{Prop}}$. Then $\mathtt{t}\colon \mathbf{CPL} \rightsquigarrow \mathbb{L}^*_{\mathbf{CPL}}$. Moreover, $\mathtt{t}$ is functional:

$$\begin{aligned} \mathtt{t}(\neg)(\mathtt{t}(\phi)) &= [\neg \phi]_{\mathbf{CPL}} \\ \mathtt{t}(\wedge)(\mathtt{t}(\phi), \mathtt{t}(\psi)) &= [\phi \wedge \psi]_{\mathbf{CPL}} . \end{aligned}$$

But $\mathtt{t}[\mathbf{CPL}]$ is ambiguous. For instance, $[\phi]_{\mathbf{CPL}} = [\neg\neg \phi]_{\mathbf{CPL}} = \sim\sim [\phi]_{\mathbf{CPL}}$.

Thus, it is important not assume $\mathtt{t}[\mathsf{L}]$ is automatically unambiguous just because L is. In certain special cases, however, this assumption is justified.

Lemma 3.1.8 *(Unambiguity Preserved Under Injective Image)*. Let L_1 and L_2 be logics, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ be injective. Then L_1 is unambiguous iff $\mathtt{t}[\mathsf{L}_1]$ is unambiguous.

Proof: Automatically, by Proposition 3.1.2(c), $\mathtt{t}$ is functional.

($\Rightarrow$) Let $\chi \in \mathtt{t}[\mathtt{At}_1]$. Then for no $\triangle \in \mathtt{Op}_1$ and $\rho \in \mathcal{L}_1^{\gamma}$ is $\chi = \mathtt{t}(\triangle(\rho))$, and so $\chi \neq \mathtt{t}(\triangle)(\mathtt{t} \circ \rho)$. Now suppose that $\mathtt{t}(\triangle_1)(\mathtt{t} \circ \rho_1) = \mathtt{t}(\triangle_2)(\mathtt{t} \circ \rho_2)$. Then $\mathtt{t}(\triangle_1(\rho_1)) = \mathtt{t}(\triangle_2(\rho_2))$, in which case $\triangle_1(\rho_1) = \triangle_2(\rho_2)$ by injectivity. So by the unambiguity of L_1, $\triangle_1 = \triangle_2$ and $\rho_1 = \rho_2$. Hence, $\mathtt{t}(\triangle_1) = \mathtt{t}(\triangle_2)$ and $\mathtt{t} \circ \rho_1 = \mathtt{t} \circ \rho_2$.

($\Leftarrow$) The atomic case is the same. So suppose $\triangle_1(\rho_1) = \triangle_2(\rho_2)$. Thus, $\mathtt{t}(\triangle_1)(\mathtt{t} \circ \rho_1) = \mathtt{t}(\triangle_2)(\mathtt{t} \circ \rho_2)$. By the unambiguity of $\mathtt{t}[\mathsf{L}_1]$, that means $\mathtt{t} \circ \rho_1 = \mathtt{t} \circ \rho_2$, which means by injectivity that $\rho_1 = \rho_2$. It also means that $\mathtt{t}(\triangle_1) = \mathtt{t}(\triangle_2)$, although this does not yet guarantee that $\triangle_1 = \triangle_2$. Suppose for *reductio* that $\triangle_1 \neq \triangle_2$. Clearly they still must have the same arity, since $\mathtt{t}(\triangle_1) = \mathtt{t}(\triangle_2)$. That means that for some $\rho \in \mathcal{L}_1^{\gamma}$, $\triangle_1(\rho) \neq \triangle_2(\rho)$. By injectivity, $\mathtt{t}(\triangle_1(\rho)) \neq \mathtt{t}(\triangle_2(\rho))$, and so $\mathtt{t}(\triangle_1)(\mathtt{t} \circ \rho) \neq \mathtt{t}(\triangle_2)(\mathtt{t} \circ \rho)$. But that means $\mathtt{t}(\triangle_1) \neq \mathtt{t}(\triangle_2)$, $\lightning$. Hence, $\triangle_1 = \triangle_2$, giving us the unambiguity of L_1. ∎

§ 3.2 Typography

Arguably, functional translational equivalence is at least necessary for notational variance (at least among intensional languages), although it is clearly not sufficient. Conversely, it is worth briefly remarking on a very strong property, proposed by Wójcicki [1988, p. 67], which suffices for notational variance though it is clearly not necessary, viz., typographical variance. In short, typographical variants are those logics where one just "rewrites" the operators.

Definition 3.2.1 *(Typography)*. Let L_1 and L_2 be logics, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. A Σ_2-operator $\triangle_2 \in \mathtt{Op}_2^\gamma$ is the ***typographical variant*** of $\triangle_1 \in \mathtt{Op}_1^\gamma$ via $\mathtt{t}$ if for every $\rho \in \mathcal{L}_1^\gamma$, $\mathtt{t}(\triangle_1(\rho)) = \triangle_2(\mathtt{t} \circ \rho)$. We will say $\mathtt{t}$ is ***typographical*** if every Σ_1-operator has a typographical variant in Σ_2 via $\mathtt{t}$.

If there is a typographical translation from L_1 to L_2, then L_1 is in a very strict sense a notational variant of a fragment of L_2. It is easy to verify that there is no typographical translation from **FOL** to **CPL**, or from **CPL** or **IPL** to the other. Generally speaking, however, if we have a translation that preserves the proof-theoretic operators up to equivalence, we can find a translation that preserves them up to typographical variance. Below, we will show that there is a translational equivalence between **FOL** and **CPL** that is typographical with respect to the boolean operators. But first, a lemma:

Lemma 3.2.2 *(Preservation of Proof-Theoretic Connectives up to Equivalence)*. Let L_1 and L_2 be some logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Suppose $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$. Assuming L_1 and L_2 have the appropriate proof-theoretic operators:

(a) $\mathtt{t}(\top_1) \dashv\vdash_2 \top_2$.

(b) $\mathtt{t}(\phi \wedge_1 \psi) \dashv\vdash_2 \mathtt{t}(\phi) \wedge_2 \mathtt{t}(\psi)$.

Suppose in addition that $\mathtt{t}, \mathtt{s}\colon \mathsf{L}_1 \leftrightsquigarrow \mathsf{L}_2$.

(c) $\mathtt{t}(\bot_1) \dashv\vdash_2 \bot_2$.

(d) $\mathtt{t}(\neg_1 \phi) \dashv\vdash_2 \neg_2 \mathtt{t}(\phi)$.

(e) $\mathtt{t}(\phi \vee_1 \psi) \dashv\vdash_2 \mathtt{t}(\phi) \vee_2 \mathtt{t}(\psi)$.

(f) $\mathtt{t}(\phi \to_1 \psi) \dashv\vdash_2 \mathtt{t}(\phi) \to_2 \mathtt{t}(\psi)$.

(g) $\mathtt{t}(\phi \leftrightarrow_1 \psi) \dashv\vdash_2 \mathtt{t}(\phi) \leftrightarrow_2 \mathtt{t}(\psi)$.

Proof:

(a) By **Fact 2.1.2**(f).

(b) $\{\phi, \psi\} \dashv\vdash_1 \phi \wedge_1 \psi$, so:

$$\mathtt{t}(\phi \wedge_1 \psi) \dashv\vdash_2 \{\mathtt{t}(\phi), \mathtt{t}(\psi)\} \dashv\vdash_2 \mathtt{t}(\phi) \wedge_2 \mathtt{t}(\psi).$$

(c) $\bot_1 \vdash_1 \mathtt{s}(\bot_2)$. So:

$$\mathtt{t}(\bot_1) \vdash_2 \mathtt{t}(\mathtt{s}(\bot_2)) \dashv\vdash_2 \bot_2.$$

Hence, $\mathtt{t}(\bot_1) \dashv\vdash_2 \bot_2$.

(d) Since $\phi, \neg_1 \phi \vdash_1 \mathcal{L}_1 \vdash_1 \mathtt{s}[\mathcal{L}_2]$, we have:

$$\mathtt{t}(\phi), \mathtt{t}(\neg_1 \phi) \vdash_2 \mathtt{t}[\mathtt{s}[\mathcal{L}_2]] \dashv\vdash_2 \mathcal{L}_2.$$

So $\mathtt{t}(\neg_1 \phi) \vdash_2 \neg_2 \mathtt{t}(\phi)$. Likewise, $\mathtt{s}(\neg_2 \psi) \vdash_1 \neg_1 \mathtt{s}(\psi)$. In particular:

$$\mathtt{s}(\neg_2 \mathtt{t}(\phi)) \vdash_1 \neg_1 \mathtt{s}(\mathtt{t}(\phi)) \dashv\vdash_1 \neg_1 \phi.$$

So we have $\neg_2 \mathtt{t}(\phi) \vdash_2 \mathtt{t}(\neg_1 \phi)$ as well.

(e) Since $\phi \vdash_1 \phi \vee_1 \psi$, we have $\mathtt{t}(\phi) \vdash_2 \mathtt{t}(\phi \vee_1 \psi)$. Similarly, $\mathtt{t}(\psi) \vdash_2 \mathtt{t}(\phi \vee_1 \psi)$. So $\mathtt{t}(\phi) \vee_2 \mathtt{t}(\psi) \vdash_2 \mathtt{t}(\phi \vee_1 \psi)$. Likewise, $\mathtt{s}(\phi') \vee_1 \mathtt{s}(\psi') \vdash_1 \mathtt{s}(\phi' \vee_2 \psi')$. In particular:

$$\phi \vee_1 \psi \dashv\vdash_1 \mathtt{s}(\mathtt{t}(\phi)) \vee_1 \mathtt{s}(\mathtt{t}(\psi)) \vdash_1 \mathtt{s}(\mathtt{t}(\phi) \vee_2 \mathtt{t}(\psi)).$$

So we have $\mathtt{t}(\phi \vee_1 \psi) \vdash_2 \mathtt{t}(\phi) \vee_2 \mathtt{t}(\psi)$ as well.

(f) Since $\phi \rightarrow_1 \psi, \phi \vdash_1 \psi$, we have $\mathtt{t}(\phi \rightarrow_1 \psi), \mathtt{t}(\phi) \vdash_2 \mathtt{t}(\psi)$. Hence, we have $\mathtt{t}(\phi \rightarrow_1 \psi) \vdash_2 \mathtt{t}(\phi) \rightarrow_2 \mathtt{t}(\psi)$. Likewise, $\mathtt{s}(\phi \rightarrow_2 \psi) \vdash_1 \mathtt{s}(\phi) \rightarrow_1 \mathtt{s}(\psi)$. In particular:

$$\mathtt{s}(\mathtt{t}(\phi) \rightarrow_2 \mathtt{t}(\psi)) \vdash_1 \mathtt{s}(\mathtt{t}(\phi)) \rightarrow_1 \mathtt{s}(\mathtt{t}(\psi)) \dashv\vdash_1 \phi \rightarrow_1 \psi.$$

So we have $\mathtt{t}(\phi) \rightarrow_2 \mathtt{t}(\psi) \vdash_2 \mathtt{t}(\phi \rightarrow_1 \psi)$ as well.

(g) Since $\phi, \phi \leftrightarrow_1 \psi \vdash_1 \psi$ and $\psi, \phi \leftrightarrow_1 \psi \vdash_1 \phi$, we have $\mathtt{t}(\phi \leftrightarrow_1 \psi) \vdash_1 \mathtt{t}(\phi) \leftrightarrow_2 \mathtt{t}(\psi)$. Likewise, $\mathtt{s}(\phi \leftrightarrow_2 \psi) \vdash_1 \mathtt{s}(\phi) \leftrightarrow_1 \mathtt{s}(\psi)$. In particular:

$$\mathtt{s}(\mathtt{t}(\phi) \leftrightarrow_2 \mathtt{t}(\psi)) \vdash_1 \mathtt{s}(\mathtt{t}(\phi)) \leftrightarrow_1 \mathtt{s}(\mathtt{t}(\psi)) \dashv\vdash_1 \phi \leftrightarrow_1 \psi.$$

So we have $\mathtt{t}(\phi) \leftrightarrow_2 \mathtt{t}(\psi) \vdash_2 \mathtt{t}(\phi \leftrightarrow_1 \psi)$ as well. ∎

Proposition 3.2.3 *(Preserving Booleans from **FOL** to **CPL**)*. There is a translation scheme $\mathtt{t}, \mathtt{s} \colon \mathbf{FOL} \leftrightsquigarrow \mathbf{CPL}$ such that $\phi_1, \phi_2 \in \mathcal{L}_{\mathtt{Prop}}$ and all $\psi_1, \psi_2 \in \mathcal{L}_{\mathtt{Pred}}$:

$$\mathtt{t}(\neg\,\phi_1) = \neg\,\mathtt{t}(\phi_1) \qquad \mathtt{s}(\neg\,\psi_1) = \neg\,\mathtt{s}(\psi_1)$$
$$\mathtt{t}(\phi_1 \wedge \phi_2) = \mathtt{t}(\phi_1) \wedge \mathtt{t}(\phi_2) \qquad \mathtt{s}(\psi_1 \wedge \psi_2) = \mathtt{s}(\psi_1) \wedge \mathtt{s}(\psi_2).$$

Proof: Let $\mathtt{t} \colon \mathbf{FOL} \cong \mathbf{CPL}$ (we know there is such a map by **Corollary 2.6.7**). Define $\mathtt{t}^*$ as follows:

$$\begin{aligned}
\mathtt{t}^*(P^n(y_1, \ldots, y_n)) &= \mathtt{t}(P^n(y_1, \ldots, y_n)) \\
\mathtt{t}^*(\neg\,\phi) &= \neg\,\mathtt{t}^*(\phi) \\
\mathtt{t}^*(\phi \wedge \psi) &= \mathtt{t}^*(\phi) \wedge \mathtt{t}^*(\psi) \\
\mathtt{t}^*(\forall x\,\phi) &= \mathtt{t}(\forall x\,\mathtt{t}^{-1}(\mathtt{t}^*(\phi))).
\end{aligned}$$

Claim: For all $\phi \in \mathbf{FOL}$, $\mathtt{t}^*(\phi) \equiv_{\mathbf{CPL}} \mathtt{t}(\phi)$.

Subproof: By induction. The atomic case follows by definition. And the booleans follow by the inductive hypothesis using **Lemma 3.2.2**. So now we just need to check the quantifier case. Suppose $\mathtt{t}^*(\phi) \equiv_{\mathbf{CPL}} \mathtt{t}(\phi)$. Since $\mathtt{t}^{-1}$ is a translation, $\mathtt{t}^{-1}(\mathtt{t}^*(\phi)) \equiv_{\mathbf{FOL}} \mathtt{t}^{-1}(\mathtt{t}(\phi)) = \phi$. Hence, $\forall x\,\mathtt{t}^{-1}(\mathtt{t}^*(\phi)) \equiv_{\mathbf{FOL}} \forall x\,\phi$. And since $\mathtt{t}$ is a translation, $\mathtt{t}^*(\forall x\,\phi) = \mathtt{t}(\forall x\,\mathtt{t}^{-1}(\mathtt{t}^*(\phi))) \equiv_{\mathbf{CPL}} \mathtt{t}(\forall x\,\phi)$, as desired. ■

Now define $\mathtt{s}^*$ as follows:

$$\begin{aligned}
\mathtt{s}^*(p) &= \mathtt{t}^{-1}(p) \\
\mathtt{s}^*(\neg\,\phi) &= \neg\,\mathtt{s}^*(\phi) \\
\mathtt{s}^*(\phi \wedge \psi) &= \mathtt{s}^*(\phi) \wedge \mathtt{s}^*(\psi).
\end{aligned}$$

Again, by induction, for all $\phi \in \mathcal{L}_{\mathtt{Prop}}$, $\mathtt{s}^*(\phi) \equiv_{\mathbf{FOL}} \mathtt{t}^{-1}(\phi)$.

Now we show that for all $\phi \in \mathbf{FOL}$ and all $\psi \in \mathcal{L}_{\mathtt{Prop}}$:

(i) $\mathtt{s}^*(\mathtt{t}^*(\phi)) \equiv_{\mathbf{FOL}} \phi$

(ii) $\mathtt{t}^*(\mathtt{s}^*(\psi)) \equiv_{\mathbf{CPL}} \psi$.

For (i), $\mathtt{s}^*(\mathtt{t}^*(\phi)) \equiv_{\mathbf{FOL}} \mathtt{t}^{-1}(\mathtt{t}^*(\phi)) \equiv_{\mathbf{FOL}} \mathtt{t}^{-1}(\mathtt{t}(\phi)) = \phi$. Likewise for (ii). ■

In the proof above, we used an isomorphism from **FOL** to **CPL** to construct a reversible translation scheme $\mathtt{t}^*, \mathtt{s}^*$ that was also preserves the boolean operators exactly. But the construction does not guarantee that $\mathtt{s}^* = (\mathtt{t}^*)^{-1}$, and so it does not guarantee that there is a isomorphism from **FOL** to **CPL** that preserves the boolean operators. Can we construct such an isomorphism? The answer is negative:

Proposition 3.2.4 *(No Isomorphism from **FOL** to **CPL** Preserves the Booleans).* There is no $\mathtt{t}\colon \mathbf{FOL} \cong \mathbf{CPL}$ such that $\mathtt{t}(\neg\phi) = \neg\mathtt{t}(\phi)$ and $\mathtt{t}(\phi \wedge \psi) = \mathtt{t}(\phi) \wedge \mathtt{t}(\psi)$.

Proof: Suppose there is such a $\mathtt{t}$. First, observe that:

$$\mathtt{t}^{-1}(\neg\phi) = \mathtt{t}^{-1}(\neg\mathtt{t}(\mathtt{t}^{-1}(\phi))) = \mathtt{t}^{-1}(\mathtt{t}(\neg\mathtt{t}^{-1}(\phi))) = \neg\mathtt{t}^{-1}(\phi)$$
$$\mathtt{t}^{-1}(\phi \wedge \psi) = \mathtt{t}^{-1}(\mathtt{t}(\mathtt{t}^{-1}(\phi)) \wedge \mathtt{t}(\mathtt{t}^{-1}(\psi))) = \mathtt{t}^{-1}(\mathtt{t}(\mathtt{t}^{-1}(\phi) \wedge \mathtt{t}^{-1}(\psi))).$$

So we have $\mathtt{t}^{-1}(\neg\phi) = \neg\mathtt{t}^{-1}(\phi)$ and $\mathtt{t}^{-1}(\phi \wedge \psi) = \mathtt{t}^{-1}(\phi) \wedge \mathtt{t}^{-1}(\psi)$ as well.

Now, observe first that if $\phi \in \mathbf{FOL}$ is atomic, then $\mathtt{t}(\phi) \in \mathtt{Prop}$. For suppose $\mathtt{t}(\phi) = \neg\theta$. Then $\mathtt{t}^{-1}(\mathtt{t}(\phi)) = \phi = \mathtt{t}^{-1}(\neg\theta) = \neg\mathtt{t}^{-1}(\theta)$. But ϕ is atomic, and so not a negation, $\lightning$. Similarly for $\wedge$. Next, note that the same observation applies to $\phi = \forall x\,\psi$. Hence, $\mathtt{t}(\forall x\,\phi) \in \mathtt{Prop}$. Now, $\mathtt{t}(F(x)) \neq \mathtt{t}(\forall x\,F(x))$, since otherwise $F(x) \equiv_{\mathbf{CPL}} \forall x\,F(x)$, $\lightning$. Hence, $\mathtt{t}(F(x)) = p$ and $\mathtt{t}(\forall x\,F(x)) = q$ where $p \neq q$. Thus, $\mathtt{t}(\forall x\,F(x)) \nvDash_{\mathbf{CPL}} \mathtt{t}(F(x))$. But then $\mathtt{t}$ is not a translation, since $\forall x\,F(x) \vDash_{\mathbf{CPL}} F(x)$, $\lightning$. ■

Wójcicki [1988, p. 67] defines two logics to be notational variants just in case there is a bijective typographical translation from one to the other.[3] This seems to be too stringent, partly due to the fact that isomorphism in some cases seems to stringent (**Example 2.3.4**). But even if we were to revise the proposal by replacing isomorphism with translational equivalence, the proposal would still be too strong. For instance, on this proposal, $\mathbf{CPL}^{\vee}$ (i.e., **CPL** with $\vee$ as primitive instead of $\wedge$) would not count as a notational variant of **CPL**. For another illustrative example, compare **CPL**($\rightarrow$) with **CPL**($\leftarrow$), which are like **CPL** except we replace $\wedge$ with $\rightarrow$ and $\leftarrow$ respectively, where the operators are interpreted in their respective languages as follows:

$$v \Vdash_{\rightarrow} \phi \rightarrow \psi \quad\Leftrightarrow\quad v \nVdash_{\rightarrow} \phi \text{ or } v \Vdash_{\rightarrow} \psi$$
$$v \Vdash_{\leftarrow} \phi \leftarrow \psi \quad\Leftrightarrow\quad v \Vdash_{\leftarrow} \phi \text{ or } v \nVdash_{\leftarrow} \psi.$$

[3] To be fair, Wójcicki seems to treat this definition as a mere technical stipulation. He does not seem to intend to interpret this definition as a philosophical thesis about what notational variance amounts to.

Clearly, **CPL**(→) and **CPL**(←) are notational variants: formulas in $\mathcal{L}_{\texttt{Prop}}(\leftarrow)$ seem to just be the result of writing formulas in $\mathcal{L}_{\texttt{Prop}}(\rightarrow)$ backwards. But there is no typographical translation $\texttt{t}\colon \mathbf{CPL}(\rightarrow) \rightsquigarrow \mathbf{CPL}(\leftarrow)$. If $\texttt{t}$ is typographical, then that means $\texttt{t}(\phi \rightarrow \psi) = \texttt{t}(\phi) \leftarrow \texttt{t}(\psi)$, since $\leftarrow$ is the only binary operator in **CPL**(←). Now, since $\vDash_{\rightarrow} \phi \rightarrow (\psi \rightarrow \phi)$, that means that $\vDash_{\leftarrow} \texttt{t}(\phi) \leftarrow (\texttt{t}(\psi) \leftarrow \texttt{t}(\phi))$. But $\texttt{t}(\phi) \leftarrow (\texttt{t}(\psi) \leftarrow \texttt{t}(\phi)) \equiv_{\leftarrow} \texttt{t}(\phi)$. So $\vDash_{\leftarrow} \texttt{t}(\phi)$ for all $\phi \in \mathcal{L}_{\texttt{Prop}}(\rightarrow)$, $\lightning$.

The problem is that notational variants in general can have very different ways of representing the same information. Thus, typographical translational equivalence seems too strict as a precisification of notational variance.

§ 3.3 Schematicity

§ 3.3.1 Definition

Part of the reason typographical equivalence seems too stringent of a criterion for notational variance is that typographical equivalence requires translating each operator as another *operator*. Often, when we translate an operator of one language into another language, we translate it as a *schema*. In fact, most translations that one comes across in practice are ones where the translation of an operator is defined as a schema, while very few are typographical.

This suggests one natural way of strengthening the notion of translational equivalence so as to count as a more adequate formalization of notational variance: require the translation to be *schematic*. In short, a translation is schematic just in case the translation of a complex formula is a schema of the translation of the parts (where each operators is assigned its own schema).[4] Schematic translations respect the underlying syntactic structure of the formulas in the language being translated by ensuring that formulas that share their operator are translated in roughly the same way. They thereby preserve not only the the meanings of sentences of the source language but also the the meanings of operators.

Definition 3.3.1 *(Schematicity)*. Let L_1 and L_2 be logics. A map $\texttt{t}\colon \mathcal{L}_1 \rightarrow \mathcal{L}_2$ is ***schematic*** if for each $\triangle \in \texttt{Op}_1^{\gamma}$, there is a Σ_2-schema $\Theta^{\triangle}(\pi)$ with γ-many parameters π such that for all $\rho \in \mathcal{L}_1^{\gamma}$:

$$\texttt{t}(\triangle(\rho)) = \Theta^{\triangle}(\texttt{t} \circ \rho).$$

We often write "$\Theta^{\triangle}$" as "$\triangle^{\texttt{t}}$". We write "$\texttt{t}\colon \mathsf{L}_1 \multimap \mathsf{L}_2$" to mean $\texttt{t}$ is a schematic

[4] See Epstein 1990, p. 391, Pelletier and Urquhart 2003, p. 269, and Caleiro and Gonçalves 2007, p. 108. The term "compositional" is often used in place of "schematic". Epstein uses the term "grammatical". He argues that two logics should be considered notational variants only if they are grammatically intertranslatable via model-preserving translations (in the sense of **Definition 4.2.1**). Mossakowski et al. [2009, p. 4] argue that we should not require expressively equivalent logics to be schematically intertranslatable, for reasons we will review shortly.

translation from L_1 to L_2. We use "$\overset{\circ\!-}{-\!\circ}$" for schematic intertranslatability, "$\circ\!-\!\circ$" for schematic translational equivalence, and "$\overset{\circ}{=}$" for schematic isomorphism.

Clearly, every typographical translation is schematic, and every schematic translation is functional. The converses of these claims all fail. **Example 2.1.6** is an easy counterexample to the claim that schematic translatability implies typographical translatability. And as we will see below (**Proposition 3.3.11**), **FOL** is not schematically translatable into **CPL**, although **FOL** is functionally translatable into **CPL** by **Corollary 2.6.7** and **Proposition 3.1.2**.

Schematic maps induce a map from schemas in the source logic to schemas in the target logic as follows:

Lemma 3.3.2 *(Induced Schema Map)*. Let L_1 and L_2 be logics, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ be schematic. Define the induced map $\mathtt{t}^*\colon \mathtt{Sch}_{\Sigma_1}(\Pi) \to \mathtt{Sch}_{\Sigma_2}(\Pi)$ as follows:

$$\mathtt{t}^*(\Theta) \quad = \quad \begin{cases} \mathtt{t}(\phi) & \text{if } \Theta = \phi \in \mathcal{L}_1 \\ \xi & \text{if } \Theta = \xi \in \Pi \\ \triangle^{\mathtt{t}}(\mathtt{t}^* \circ \sigma) & \text{if } \Theta = \langle \triangle, \sigma \rangle. \end{cases}$$

Let $\Theta(\pi)$ be a Σ_1-schema with a γ-sequence of distinct parameters π and let $\rho \in \mathcal{L}_1^{\gamma}$. Then $\mathtt{t}(\Theta(\rho)) = \mathtt{t}^*(\Theta)(\mathtt{t} \circ \rho)$.

Proof: By induction. The base cases are trivial. For the inductive case, we proceed in terms of instantiations (**Definition 1.3.9**). Let $\triangle \in \mathtt{Op}_1^{\gamma}$ and let σ be a γ-sequence of Σ_1-schemas with parameters π such that $(\mathtt{t}^* \circ \sigma)[\mathtt{t} \circ \iota] = \mathtt{t} \circ \sigma[\iota]$.[a] We want to show that $\mathtt{t}^*(\langle \triangle, \sigma \rangle)[\mathtt{t} \circ \iota] = \mathtt{t}(\langle \triangle, \sigma \rangle [\iota])$.

First, observe that by induction that if $\Theta(\pi)$ is a schema, and if $\Theta(\sigma)$ is the schema resulting from replacing each $\pi(\beta)$ with $\sigma(\beta)$, then $\Theta(\sigma)[\iota] = \Theta(\sigma[\iota])$ for any instantiation ι. Thus:

$$\begin{aligned} \mathtt{t}^*(\langle \triangle, \sigma \rangle)[\mathtt{t} \circ \iota] \quad &= \quad \triangle^{\mathtt{t}}(\mathtt{t}^* \circ \sigma)[\mathtt{t} \circ \iota] \\ &= \quad \triangle^{\mathtt{t}}((\mathtt{t}^* \circ \sigma)[\mathtt{t} \circ \iota]) \\ &= \quad \triangle^{\mathtt{t}}(\mathtt{t} \circ \sigma[\iota]) \\ &= \quad \mathtt{t}(\triangle(\sigma[\iota])) \\ &= \quad \mathtt{t}(\langle \triangle, \sigma \rangle [\iota]). \end{aligned}$$

This completes the proof. ■

[a] That is: $\mathtt{t}^*(\sigma(\beta))[\mathtt{t} \circ \iota] = \mathtt{t}(\sigma(\beta)[\iota])$ for all $\beta < \gamma$.

Convention: In what follows, we write "$\Theta^{\mathtt{t}}$" in place of "$\mathtt{t}^*(\Theta)$".

Corollary 3.3.3 *(Composition of Schematic Maps are Schematic)*. Let L_1 and L_2 be logics. If $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_3$ are schematic, then so is $\mathtt{s} \circ \mathtt{t}$.

Proof: Simply define $\triangle^{\mathtt{s}\circ\mathtt{t}} = (\triangle^{\mathtt{t}})^{\mathtt{s}}$ and use **Lemma 3.3.2**. ■

Corollary 3.3.4 *(Schematic Translatability is Transitive)*. If $\mathsf{L}_1 \multimap \mathsf{L}_2$ and $\mathsf{L}_2 \multimap \mathsf{L}_3$, then $\mathsf{L}_1 \multimap \mathsf{L}_3$. Similarly, if $\mathsf{L}_1 \mathrel{\circ\!\!-\!\!\circ} \mathsf{L}_2$ and $\mathsf{L}_2 \mathrel{\circ\!\!-\!\!\circ} \mathsf{L}_3$, then $\mathsf{L}_1 \mathrel{\circ\!\!-\!\!\circ} \mathsf{L}_3$.

Before we state the next result, two remarks are in order. First, recall the definition of an image logic under a functional map (**Definition 3.1.4**). If $\mathtt{t}\colon \mathsf{L}_1 \multimap \mathsf{L}_2$, note that where $\rho \in \mathcal{L}_2^{\gamma}$, $\mathtt{t}(\triangle)(\rho) = \triangle^{\mathtt{t}}(\rho)$. However, one should still distinguish between $\mathtt{t}(\triangle)$ and $\triangle^{\mathtt{t}}$ conceptually. $\mathtt{t}(\triangle)$ is a function from sequences of formulas to the instantiation of $\triangle^{\mathtt{t}}$ with those formulas, while $\triangle^{\mathtt{t}}$ is a schema, i.e., a sequence of parameters, formulas, and operators. Thus, $\mathtt{t}(\triangle)$ lives on a higher-level than $\triangle^{\mathtt{t}}$, metaphorically speaking, even though *applying* arguments to $\mathtt{t}(\triangle)$ yields the same result as *instantiating* $\triangle^{\mathtt{t}}$ with those same arguments—that is, $\mathtt{t}(\triangle)(\mathtt{t} \circ \rho) = \triangle^{\mathtt{t}}(\mathtt{t} \circ \rho)$. One should be careful not to blur these concepts together, since $\mathtt{t}(\triangle)$ might be an operator in $\mathtt{t}[\mathsf{L}_1]$ even if there is no corresponding schema $\triangle^{\mathtt{t}}$. This will be relevant for the interpretation of **Propositions 3.3.8** and **3.3.9** below.

In any case, we want a special term to describe when we *can* safely conflate $\mathtt{t}(\triangle)$ with $\triangle^{\mathtt{t}}$.

Definition 3.3.5 *(Schematic Image Operator)*. Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ be functional. An operator $\mathtt{t}(\triangle) \in \mathtt{Op}^{\gamma}_{\mathtt{t}[1]}$ is ***t-schematic*** in L_2 if there is a Σ_2-schema $\triangle^{\mathtt{t}}(\pi)$ such that for all $\rho \in \mathcal{L}_1^{\gamma}$, $\mathtt{t}(\triangle)(\mathtt{t} \circ \rho) = \triangle^{\mathtt{t}}(\mathtt{t} \circ \rho)$. $\mathtt{t}[\mathsf{L}_1]$ as a whole is ***schematic*** in L_2 if each $\mathtt{t}(\triangle) \in \mathtt{Op}_{\mathtt{t}[1]}$ is schematic in L_2.

Fact 3.3.6 *(Schematic Map and Schematic Image)*. Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ be functional. Then $\mathtt{t}$ is schematic iff $\mathtt{t}[\mathsf{L}_1]$ is schematic.

Second, we still interpret extensions and fragments in terms of **Definition 1.1.15**. Thus, for $\mathsf{L}_1 \subseteq \mathsf{L}_2$, we only require that $\mathcal{L}_1 \subseteq \mathcal{L}_2$ and $\Gamma \vdash_1 \phi$ iff $\Gamma \vdash_2 \phi$. In particular, we do *not* require that every Σ_1-operator be a Σ_2-operator. Otherwise, the fragment of first-order logic picked out by the standard translation of modal logic (for instance) would not be considered a *fragment* of first-order logic, since it is not closed under (unguarded) universal quantification. So while it would be interesting to study the fragments of logics which do only contain operators from their extensions, we do not build this into the definition of a fragment as such. Thus, where $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$, we still have $\mathtt{t}[\mathsf{L}_1] \subseteq \mathsf{L}_2$, even though $\mathtt{Op}_{\mathtt{t}[1]} \nsubseteq \mathtt{Op}_2$.

Still, it would be unnatural for a fragment of a compositional language to have operators that were not, in any sense, definable in terms of its extension. After all, the operators of, say, the fragment of first-order logic picked out by the standard translation are still broadly first-order schemas. So there are special reasons for focusing on the *schematic* fragments of a language in particular. For that reason, we adopt the following convention:

Convention: Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ be functional. We write "$\mathtt{t}[\mathsf{L}_1] \overset{\circ}{\subseteq} \mathsf{L}_2$" to mean that $\mathtt{t}[\mathsf{L}_1]$ is a schematic fragment of L_2.

Fact 3.3.7 *(Schematic Images are Schematically Translatable)*. Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ be functional. If $\mathtt{t}[\mathsf{L}_1] \overset{\circ}{\subseteq} \mathsf{L}_2$, then $\mathsf{id}\colon \mathtt{t}[\mathsf{L}_1] \multimap \mathsf{L}_2$.

With these qualifications in place, we can now prove the schematic counterpart of **Proposition 2.2.5**.

Proposition 3.3.8 *(Alternative Formulation of Schematic Translatability)*. Let L_1 and L_2 be logics. Suppose that L_1 is intensional and that $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is functional. Then the following are equivalent:

(a) $\mathtt{t}\colon \mathsf{L}_1 \multimap \mathsf{L}_2$.

(b) $\mathsf{L}_1 \overset{\circ\!-}{\multimap} \mathtt{t}[\mathsf{L}_1] \overset{\circ}{\subseteq} \mathsf{L}_2$.

(c) $\mathsf{L}_1 \circ\!\!\multimap \mathtt{t}[\mathsf{L}_1] \overset{\circ}{\subseteq} \mathsf{L}_2$.

Proof: It suffices to show that (a) implies (c), since the other implications are obvious. Clearly, $\mathtt{t}[\mathsf{L}_1]$ is schematic. Define $\mathtt{s}$ inductively as follows. First, where $\mathtt{t}(\chi) \in \mathtt{t}[\mathtt{At}_1]$, let $\mathtt{s}(\mathtt{t}(\chi))$ be an arbitrary member of $[\chi]_1$. Next, where $\mathtt{t}(\triangle) \in \mathtt{Op}_{\mathtt{t}[1]}$, use the axiom of choice to set $\mathtt{s}(\mathtt{t}(\triangle))$ to be an arbitrary member of $\{\triangle' \in \mathtt{Op}_1 \mid \mathtt{t}(\triangle') = \mathtt{t}(\triangle)\}$. Finally, define:

$$\mathtt{s}(\mathtt{t}(\triangle)(\mathtt{t} \circ \rho)) = \mathtt{s}(\mathtt{t}(\triangle))(\mathtt{s} \circ \mathtt{t} \circ \rho).$$

First, observe that $\mathtt{s}$ is a translation by **Proposition 2.2.4**. Next, we show that $\mathtt{s}(\mathtt{t}(\phi)) \equiv_1 \phi$ for all $\phi \in \mathcal{L}_1$. If $\phi \in \mathtt{At}_1$, then this follows by definition.

Now suppose $\rho \in \mathcal{L}_1^\gamma$ is such that $\mathtt{s} \circ \mathtt{t} \circ \rho \equiv_1 \rho$. Then:

$$\begin{aligned} \mathtt{s}(\mathtt{t}(\triangle(\rho))) &= \mathtt{s}(\mathtt{t}(\triangle)(\mathtt{t} \circ \rho)) \\ &= \mathtt{s}(\mathtt{t}(\triangle))(\mathtt{s} \circ \mathtt{t} \circ \rho) \\ &\equiv_1 \triangle(\mathtt{s} \circ \mathtt{t} \circ \rho) \\ &\equiv_1 \triangle(\rho). \end{aligned}$$

The third step follows since if $\mathtt{t}(\mathtt{s}(\mathtt{t}(\triangle))) = \mathtt{t}(\triangle)$, then $\mathtt{s}(\mathtt{t}(\triangle))(\rho') \equiv_1 \triangle(\rho')$ for all $\rho' \in \mathcal{L}_1^\gamma$. The last step follows by intensionality. Finally, observe that automatically $\mathtt{t}(\mathtt{s}(\psi)) \equiv_{\mathtt{t}[1]} \psi$ for $\psi \in \mathtt{t}[\mathcal{L}_1]$, since we have that $\mathtt{t}(\mathtt{s}(\mathtt{t}(\phi))) \equiv_{\mathtt{t}[1]} \mathtt{t}(\phi)$. ∎

We can relax the requirement that L_1 be intensional if $\mathtt{t}$ is injective.

Proposition 3.3.9 *(Alternative Formulation of Injective Schematic Translatability)*. Let L_1 and L_2 be logics. Suppose that $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is injective. Then the following are equivalent:

(a) $\mathtt{t}\colon \mathsf{L}_1 \multimap \mathsf{L}_2$.

(b) $\mathsf{L}_1 \overset{\circ\!-}{\underset{-\!\circ}{\rightleftarrows}} \mathtt{t}[\mathsf{L}_1] \overset{\circ}{\subseteq} \mathsf{L}_2$.

(c) $\mathsf{L}_1 \circ\!\!-\!\!\circ\ \mathtt{t}[\mathsf{L}_1] \overset{\circ}{\subseteq} \mathsf{L}_2$.

(d) $\mathsf{L}_1 \overset{\circ}{=} \mathtt{t}[\mathsf{L}_1] \overset{\circ}{\subseteq} \mathsf{L}_2$.

Proof: We just need to show (a) implies (d). First, notice that **Fact 2.3.2** establishes that $\mathtt{t}, \mathtt{t}^{-1}\colon \mathsf{L}_1 \cong \mathtt{t}[\mathsf{L}_1]$. So we just need to show that $\mathtt{t}^{-1}$ is schematic. For each $\mathtt{t}(\triangle) \in \mathtt{Op}_{\mathtt{t}[1]}$, define $\mathtt{t}(\triangle)^{-1}(\pi) = \triangle(\pi)$. Then:

$$\begin{aligned} \mathtt{t}^{-1}(\mathtt{t}(\triangle)(\mathtt{t} \circ \rho)) &= \mathtt{t}^{-1}(\mathtt{t}(\triangle(\rho))) \\ &= \triangle(\rho) \\ &= \triangle(\mathtt{t}^{-1} \circ \mathtt{t} \circ \rho) \\ &= \mathtt{t}(\triangle)^{-1}(\mathtt{t}^{-1} \circ \mathtt{t} \circ \rho). \end{aligned}$$

So $\mathtt{t}^{-1}$ is schematic. ∎

We cannot in general drop both the injectivity and intensionality assumptions, even in moving from (a) to (b). For instance, consider the logics from **Example 3.1.7**. Because L' is ambiguous, we cannot find a schematic translation from $\mathsf{L}' = \mathtt{t}[\mathsf{CPL}]$

into CPL. If there were a $\mathtt{s}\colon \mathtt{t}[\mathsf{CPL}] \multimap \mathsf{CPL}$, then where $\mathtt{s}([\phi]_{\mathsf{CPL}}) \sqcap^{\mathtt{s}} \mathtt{s}([\psi]_{\mathsf{CPL}}) = \mathtt{s}([\phi]_{\mathsf{CPL}} \sqcap [\psi]_{\mathsf{CPL}})$, we would have $\mathtt{s}([p]_{\mathsf{CPL}}) = \mathtt{s}([p]_{\mathsf{CPL}} \sqcap [p]_{\mathsf{CPL}}) = \mathtt{s}([p]_{\mathsf{CPL}}) \sqcap^{\mathtt{s}} \mathtt{s}([p]_{\mathsf{CPL}})$. Since $\mathtt{s}([p]_{\mathsf{CPL}})$ cannot be a subformula of itself, that must mean that $\pi_1 \sqcap^{\mathtt{s}} \pi_2 = \pi_1$ or $\pi_1 \sqcap^{\mathtt{s}} \pi_2 = \pi_2$—either way, $\mathtt{s}$ would not be a translation.

Moreover, we cannot drop the condition that $\mathtt{t}[\mathsf{L}_1]$ be schematic. In particular, $\mathsf{L}_1 \multimapboth \mathtt{t}[\mathsf{L}_1]$, without the assumption that $\mathtt{t}[\mathsf{L}_1] \overset{\circ}{\subseteq} \mathsf{L}_2$, does *not* imply that $\mathtt{t}\colon \mathsf{L}_1 \multimap \mathsf{L}_2$. Even if $\mathtt{t}$ is schematic for $\mathtt{t}[\mathsf{L}_1]$, whose operators are functions of the form $\mathtt{t}(\triangle)$ where $\triangle \in \mathtt{Op}_1$, it does not follow that $\mathtt{t}$ will be schematic in L_2, whose operators are not necessarily $\mathtt{t}(\triangle)$ and where $\mathtt{t}(\triangle)$ may not even be a schema. One way to bring this out is with the following observation:

Proposition 3.3.10 *(From Injective Translatability to Schematic Equivalence)*. Let L_1 and L_2 be logics, and let $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ be injective. Then $\mathtt{t}, \mathtt{t}^{-1}\colon \mathsf{L}_1 \multimapboth \mathtt{t}[\mathsf{L}_1]$.

Proof: First, since $\mathtt{t}(\triangle) \in \mathtt{Op}_{\mathtt{t}[1]}$, $\mathtt{t}$ is automatically schematic. Moreover:

$$\begin{aligned} \mathtt{t}^{-1}(\mathtt{t}(\triangle)(\mathtt{t} \circ \rho)) &= \mathtt{t}^{-1}(\mathtt{t}(\triangle(\rho))) \\ &= \triangle(\rho) \\ &= \triangle(\mathtt{t}^{-1} \circ \mathtt{t} \circ \rho). \end{aligned}$$

So $\mathtt{t}^{-1}$ is also schematic. And by **Fact 2.3.2**, $\mathtt{t}, \mathtt{t}^{-1}\colon \mathsf{L}_1 \leftrightsquigarrow \mathtt{t}[\mathsf{L}_1]$. ■

So injective translatability already guarantees schematic equivalence with the image logic. Thus, where $\mathtt{t}\colon$ **FOL** $\cong$ **CPL**, **FOL** $\overset{\circ}{=}$ $\mathtt{t}$[**FOL**] $\cong$ **CPL**. But we already observed that, in the case of **FOL**, it does not seem as though **FOL** is compositionally translatable into **CPL**; and indeed, we show in a moment that it is not. While this might seem odd, it does make sense in light of the fact the operators of $\mathtt{t}$[**FOL**] are not really just the standard boolean operators, but also include more gerrymandered operators such as $\mathtt{t}(\forall x\ )$.

§ 3.3.2 Applications

Schematicity seems to strike a nice balance between functionality and typographicality. We illustrate this with two examples. First, we show that there is no schematic translation from **FOL** into **CPL**. Thus, schematic translations can differentiate between **FOL** and **CPL**. Second, we show that there is no schematic translation from **IPL** into **CPL**. This suggests that schematic equivalence is closer to notational variance than mere translational equivalence. The match is not perfect, however, as will be illustrated by applying the notions to modal logic.

Proposition 3.3.11 *(No Schematic Translation from First-Order Logic to Propositional Logic)*. There is no $\mathtt{t}\colon \mathbf{FOL} \multimap \mathbf{CPL}$.

Proof: Suppose there were such a $\mathtt{t}$. Let $\Pi_x(\xi)$ be the propositional schema such that $\mathtt{t}(\forall x\, \phi) = \Pi_x(\mathtt{t}(\phi))$. Then:

$$\top \equiv_{\mathbf{CPL}} \mathtt{t}(\top) \equiv_{\mathbf{CPL}} \mathtt{t}(\forall x\, \top) = \Pi_x(\mathtt{t}(\top)) \equiv_{\mathbf{CPL}} \Pi_x(\top).$$

Thus:

$$\begin{aligned} \mathtt{t}(\phi) &\vDash_{\mathbf{CPL}} \mathtt{t}(\phi) \leftrightarrow \top \\ &\vDash_{\mathbf{CPL}} \Pi_x(\mathtt{t}(\phi)) \leftrightarrow \Pi_x(\top) \\ &\vDash_{\mathbf{CPL}} \Pi_x(\mathtt{t}(\phi)) = \mathtt{t}(\forall x\, \phi). \end{aligned}$$

So $\mathtt{t}(\phi) \vDash_{\mathbf{CPL}} \mathtt{t}(\forall x\, \phi)$, contrary to the fact that generally $\phi \nvDash_{\mathbf{FOL}} \forall x\, \phi$, ↯. ∎

Example 2.2.8 mentions that $\mathbf{CPL} \overset{\leftrightsquigarrow}{\rightsquigarrow} \mathbf{IPL}$, which follows from **Theorem 2.6.5**. Interestingly, schematicity introduces an asymmetry. Since the double-negation translation (**Example 2.1.8**) is schematic, we still have $\mathbf{CPL} \multimap \mathbf{IPL}$. And yet:

Proposition 3.3.12 *(No Schematic Translation from a Non-Classical Intermediate Logic to Classical Logic)*. Let L be an intermediate logic such that $\mathsf{L} \multimap \mathsf{CPL}$. Then $\mathsf{L} = \mathsf{CPL}$.

Proof: Suppose there were a $\mathtt{t}\colon \mathsf{L} \multimap \mathsf{CPL}$. Let $\Theta(\xi)$ be the propositional schema such that $\mathtt{t}(\neg\, \phi) = \Theta(\mathtt{t}(\phi))$. Then:

$$\begin{aligned} \neg\, \mathtt{t}(\phi) &\vdash_{\mathsf{CPL}} \mathtt{t}(\phi) \leftrightarrow \bot \\ &\vdash_{\mathsf{CPL}} \Theta(\Theta(\mathtt{t}(\phi))) \leftrightarrow \Theta(\Theta(\bot)) \\ &\vdash_{\mathsf{CPL}} \Theta(\Theta(\mathtt{t}(\phi))) \leftrightarrow \Theta(\Theta(\mathtt{t}(\bot))) \\ &\vdash_{\mathsf{CPL}} \mathtt{t}(\neg\neg\, \phi) \leftrightarrow \mathtt{t}(\neg\neg\, \bot) \\ &\vdash_{\mathsf{CPL}} \mathtt{t}(\neg\neg\, \phi) \leftrightarrow \mathtt{t}(\bot). \end{aligned}$$

The step from $\Theta(\Theta(\mathtt{t}(\phi))) \leftrightarrow \Theta(\Theta(\bot))$ to $\Theta(\Theta(\mathtt{t}(\phi))) \leftrightarrow \Theta(\Theta(\mathtt{t}(\bot)))$ follows from the fact that $\neg\, \mathtt{t}(\phi) \vdash_{\mathsf{CPL}} \neg\, \mathtt{t}(\bot) \dashv\vdash_{\mathsf{CPL}} \mathtt{t}(\bot) \leftrightarrow \bot$. Hence, $\mathtt{t}(\neg\neg\, \phi), \neg\, \mathtt{t}(\phi) \vdash_{\mathsf{CPL}} \mathtt{t}(\bot)$. So:

$$\begin{aligned} \mathtt{t}(\neg\neg\, \phi) &\vdash_{\mathsf{CPL}} \mathtt{t}(\phi) \vee \mathtt{t}(\bot) \\ &\vdash_{\mathsf{CPL}} \mathtt{t}(\phi \vee \bot) \end{aligned}$$

$$\vdash_{\mathsf{CPL}} \mathtt{t}(\phi).$$

But if $\neg\neg\phi \vdash_{\mathsf{L}} \phi$ for all ϕ, then $\mathsf{L} = \mathsf{CPL}$. ■

Corollary 3.3.13 *(No Schematic Translation from Intuitionistic Logic to Classical Logic)*. There is no $\mathtt{t}\colon \mathbf{IPL} \multimap \mathbf{CPL}$.

Theorem 2.6.5 also shows that for any normal modal logic L, $\mathsf{L} \cong \mathsf{CPL}$. But again, schematicity is strong enough to distinguish CPL from most normal modal logics:[5]

Proposition 3.3.14 *(Schematic Translations from Modal Logic to Propositional Logic)*. Let L be a normal modal logic. If $\mathsf{L} \multimap \mathsf{CPL}$, then $\phi \vdash_{\mathsf{L}} \Box\phi$.

Proof: Let $\mathtt{t}\colon \mathsf{L} \multimap \mathsf{CPL}$ and let $\Theta(\xi)$ be a propositional schema such that $\mathtt{t}(\Box\phi) = \Theta(\mathtt{t}(\phi))$. Observe that there are $\mathcal{L}_{\mathtt{Prop}}$-formulas λ and μ such that for all $\phi \in \mathcal{L}$:

$$\Theta(\mathtt{t}(\phi)) \dashv\vdash_{\mathsf{CPL}} (\mathtt{t}(\phi) \wedge \lambda) \vee (\neg\,\mathtt{t}(\phi) \wedge \mu)\,.$$

Now, since $\vdash_{\mathsf{L}} \Box\top$, we have that:

$$\vdash_{\mathsf{CPL}} \mathtt{t}(\Box\top) = (\mathtt{t}(\top) \wedge \lambda) \vee (\neg\,\mathtt{t}(\top) \wedge \mu) \dashv\vdash_{\mathsf{CPL}} \lambda.$$

Hence, $\mathtt{t}(\Box\phi) \dashv\vdash_{\mathsf{CPL}} \mathtt{t}(\phi) \vee \mu$. This implies that $\mathtt{t}(\phi) \vdash_{\mathsf{CPL}} \mathtt{t}(\Box\phi)$, and so $\phi \vdash_{\mathsf{L}} \Box\phi$. ■

If L is a normal modal logic such that $\phi \vdash_{\mathsf{L}} \Box\phi$, then $\Box\phi \dashv\vdash_{\mathsf{L}} \Box\bot \vee \phi$.[6] Hence, if L also extends KD, then $\phi \dashv\vdash_{\mathsf{L}} \Box\phi$, making L the trivial modal logic Id. Thus, if $\mathtt{t}\colon \mathsf{L} \rightsquigarrow \mathsf{CPL}$ is schematic, then either $\mathsf{L} = \mathsf{Id}$ or L does not satisfy the D-axiom ($\Diamond\top$). This means that if L is schematically translatable into CPL, it is already a fairly uninteresting modal logic. In particular:

Corollary 3.3.15 *(No Schematic Translation from **S5** to **CPL**)*. There is no $\mathtt{t}\colon \mathbf{S5} \multimap \mathbf{CPL}$.

[5] Here, I am understanding "modal logic" in terms of local model validity, i.e., $\Gamma \vdash_{\mathsf{L}} \phi$ if for every pointed model $\langle \mathcal{M}, w\rangle$, if $\mathcal{M}, w \Vdash \Gamma$, then $\mathcal{M}, w \Vdash \phi$. The results below apply equally to local frame validity, but not to global (model or frame) validity.

[6] Proof: Clearly $\Box\bot \vee \phi \vdash_{\mathsf{L}} \Box\phi$. Conversely, $\Box\phi, \neg\phi \vdash_{\mathsf{L}} \Box\neg\phi$ and $\Box\phi, \Box\neg\phi \vdash_{\mathsf{L}} \Box(\phi \wedge \neg\phi) \vdash_{\mathsf{L}} \Box\bot$. So $\Box\phi, \neg\phi \vdash_{\mathsf{L}} \Box\bot$.

While schematic translatability seems to conform to our intuitive judgments in the cases above, it unfortunately does not make all of the right predictions. In particular, notice that the standard translation $\mathtt{ST}_x$ of **K** into **FOL** (**Example 2.1.7**) is not technically schematic in the sense of **Definition 3.3.1**. The problem is that the $\Box$-clause is not schematically defined:

$$\mathtt{ST}_x(\Box\phi) \quad = \quad \forall y\,((x \twoheadrightarrow y) \to \mathtt{ST}_y(\phi)).$$

Since $\mathtt{ST}_x(\phi)$ is not a subformula of $\mathtt{ST}_x(\Box\phi)$, $\mathtt{ST}_x(\Box\phi)$ cannot be construed as a schema Θ such that $\mathtt{ST}_x(\Box\phi) = \Theta(\mathtt{ST}_x(\phi))$. Moreover, this defect is not repairable in general:

Proposition 3.3.16 *(Schematic Translations from Modal Logic to First-Order Logic).* Let L be a normal modal logic. If $\mathtt{t}\colon \mathsf{L} \multimap \mathsf{FOL}$, then $\Box\phi \dashv\vdash_{\mathsf{L}} \Box\Box\phi$.

Proof: Let $\Theta(\xi)$ be the first-order schema such that $\mathtt{t}(\Box\phi) = \Theta(\mathtt{t}(\phi))$. Without loss of generality, we may assume $\Theta(\xi)$ is in prenex normal form, i.e., that:

$$\Theta(\mathtt{t}(\phi)) = \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \; ((\mathtt{t}(\phi) \wedge \lambda) \vee (\neg\, \mathtt{t}(\phi) \wedge \mu))$$

where λ and μ are boolean combinations of atomic $\mathcal{L}_{\mathtt{Pred}}$-formulas. Observe that since $\vdash_{\log K} \Box\top$:

$$\begin{aligned} \vdash_{\mathsf{FOL}} \quad & \mathtt{t}(\Box\top) = \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \; ((\mathtt{t}(\top) \wedge \lambda) \vee (\neg\, \mathtt{t}(\top) \wedge \mu)) \\ \dashv\vdash_{\mathsf{FOL}} \quad & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \, \lambda. \end{aligned}$$

So $\vdash_{\mathsf{FOL}} \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \, \lambda$.

First, we show $\Box\phi \vdash_{\mathsf{L}} \Box\Box\phi$. Using the fact that $\vdash_{\mathsf{FOL}} \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \, \lambda$:

$$\begin{aligned} \mathtt{t}(\Box\phi) \quad & \vdash_{\mathsf{FOL}} \quad \mathtt{t}(\Box\phi) \wedge \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \, \lambda \\ & \dashv\vdash_{\mathsf{FOL}} \quad \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \; (\mathtt{t}(\Box\phi) \wedge \lambda)\,, \end{aligned}$$

since $y_1, \ldots, y_n$ are already bound in $\mathtt{t}(\Box\phi)$. So:

$$\begin{aligned} \mathtt{t}(\Box\phi) & \vdash_{\mathsf{FOL}} \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \; (\mathtt{t}(\Box\phi) \wedge \lambda) \\ & \vdash_{\mathsf{FOL}} \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \; ((\mathtt{t}(\Box\phi) \wedge \lambda) \vee (\neg\, \mathtt{t}(\Box\phi) \wedge \mu)) = \mathtt{t}(\Box\Box\phi). \end{aligned}$$

Hence, $\mathtt{t}(\Box\phi) \vdash_{\mathsf{FOL}} \mathtt{t}(\Box\Box\phi)$, and thus, $\Box\phi \vdash_{\mathsf{L}} \Box\Box\phi$.

Next, we show $\Box\Box\phi \vdash_{\mathsf{L}} \Box\phi$. Observe that:

$$\Theta(\mathtt{t}(\phi)) \dashv\vdash_{\mathsf{FOL}} \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \; ((\mathtt{t}(\phi) \vee \mu) \wedge (\neg\, \mathtt{t}(\phi) \vee \lambda))\,.$$

From this, we have the following:

$$
\begin{array}{rl}
& \mathtt{t}(\Box\Box\phi) \\
\dashv\vdash_{\mathsf{FOL}} & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\ ((\mathtt{t}(\Box\phi) \vee \mu) \wedge (\neg\,\mathtt{t}(\Box\phi) \vee \lambda)) \\
\vdash_{\mathsf{FOL}} & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\ (\mathtt{t}(\Box\phi) \vee \mu) \\
\dashv\vdash_{\mathsf{FOL}} & \mathtt{t}(\Box\phi) \vee \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, \mu \\
\dashv\vdash_{\mathsf{FOL}} & \mathtt{t}(\Box\phi) \vee (\neg\,\mathtt{t}(\Box\phi) \wedge \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, \mu) \\
\dashv\vdash_{\mathsf{FOL}} & (\mathtt{t}(\Box\phi) \wedge \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, \lambda) \vee (\neg\,\mathtt{t}(\Box\phi) \wedge \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, \mu) \\
\dashv\vdash_{\mathsf{FOL}} & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\ (\mathtt{t}(\Box\phi) \wedge \lambda) \vee \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\ (\neg\,\mathtt{t}(\Box\phi) \wedge \mu) \\
\dashv\vdash_{\mathsf{FOL}} & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\ ((\mathtt{t}(\Box\phi) \wedge \lambda) \vee (\neg\,\mathtt{t}(\Box\phi) \wedge \mu)) = \mathtt{t}(\Box\Box\phi).
\end{array}
$$

In particular, $\mathtt{t}(\Box\Box\phi) \dashv\vdash_{\mathsf{FOL}} \mathtt{t}(\Box\phi) \vee \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, \mu$. Moreover, unpacking $\mathtt{t}(\Box(\phi \wedge \psi))$ using **Lemma 3.2.2**:

$$
\begin{array}{rl}
& \mathtt{t}(\Box(\phi \wedge \psi)) \\
\dashv\vdash_{\mathsf{FOL}} & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\ ((\mathtt{t}(\phi \wedge \psi) \wedge \lambda) \vee (\neg\,\mathtt{t}(\phi \wedge \psi) \wedge \mu)) \\
\dashv\vdash_{\mathsf{FOL}} & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\ ((\mathtt{t}(\phi \wedge \psi) \wedge \lambda) \vee (\neg\,(\mathtt{t}(\phi) \wedge \mathtt{t}(\psi)) \wedge \mu)) \\
\dashv\vdash_{\mathsf{FOL}} & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\ ((\mathtt{t}(\phi \wedge \psi) \wedge \lambda) \vee ((\neg\,\mathtt{t}(\phi) \vee \neg\,\mathtt{t}(\psi)) \wedge \mu)) \\
\dashv\vdash_{\mathsf{FOL}} & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\ ((\mathtt{t}(\phi \wedge \psi) \wedge \lambda) \vee (\neg\,\mathtt{t}(\phi) \wedge \mu) \vee (\neg\,\mathtt{t}(\psi) \wedge \mu))\,.
\end{array}
$$

Now, since $\Box(\phi \wedge \psi) \vdash_{\mathsf{L}} \Box\phi$, and since $\mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, (\neg\,\mathtt{t}(\psi) \wedge \mu) \vdash_{\mathsf{FOL}} \mathtt{t}(\Box(\phi \wedge \psi)$, that means $\mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, (\neg\,\mathtt{t}(\psi) \wedge \mu) \vdash_{\mathsf{FOL}} \mathtt{t}(\Box\phi)$ for any ϕ and ψ. In particular:

$$\mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, (\neg\,\mathtt{t}(\Box\phi) \wedge \mu) \dashv\vdash_{\mathsf{FOL}} \neg\,\mathtt{t}(\Box\phi) \wedge \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, \mu \vdash_{\mathsf{FOL}} \mathtt{t}(\Box\phi).$$

Hence, $\mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n\, \mu \vdash_{\mathsf{FOL}} \mathtt{t}(\Box\phi)$. Thus, $\mathtt{t}(\Box\Box\phi) \dashv\vdash_{\mathsf{FOL}} \mathtt{t}(\Box\phi)$. ■

Corollary 3.3.17 *(No Schematic Translation from the Minimal Normal Modal Logic to First-Order Logic)*. There is no $\mathtt{t}\colon \mathbf{K} \multimap \mathbf{FOL}$. Similarly, for no $\Gamma \subseteq \mathcal{L}_{\mathtt{Pred}}$ do we have that $\mathbf{K} \multimap (\mathbf{FOL})_\Gamma$.

The proof does not carry over to every normal modal logic. It is straightforward to show that the following schematic map is a translation from **S5** to **FOL**:

$$
\begin{array}{rcl}
\mathtt{t}(p_i) & = & P_i(x) \\
\mathtt{t}(\neg\,\phi) & = & \neg\,\mathtt{t}(\phi) \\
\mathtt{t}(\phi \wedge \psi) & = & \mathtt{t}(\phi) \wedge \mathtt{t}(\psi) \\
\mathtt{t}(\Box\phi) & = & \forall x\, \mathtt{t}(\phi).
\end{array}
$$

However, I have yet to find a proof that settles which logics validating $\Box\phi \equiv_{\mathsf{L}} \Box\Box\phi$ can be schematically translated into **FOL**. Still, the fact that there is no schematic $\mathtt{t}\colon \mathbf{K} \multimap \mathbf{FOL}$ already seems unfortunate enough, since it is common to treat modal logics as notational variants of fragments of first-order logic already. If we want to do justice to this practice, it would be ill-advised to require notational variants to be schematically intertranslatable.

§ 3.3.3 Definitional Extensions

Schematic equivalence is sometimes characterized differently as what is called "definitional equivalence". Informally, two logics are definitionally equivalent if there is an extension of each where the operators of each logic are definable as schemas from the other.

Definition 3.3.18 *(Definitional Extension)*. Let L_1 and L_2 be logics. We say that L_2 is a ***definitional extension*** of L_1 (written "$\mathsf{L}_1 :\subseteq \mathsf{L}_2$") if:

- $\mathsf{L}_1 \subseteq \mathsf{L}_2$
- $\mathtt{At}_1 \subseteq \mathtt{At}_2$ and $\mathtt{Op}_1 \subseteq \mathtt{Op}_2$
- for each $\chi \in \mathtt{At}_2 - \mathtt{At}_1$, there is an $\mathcal{L}_1$-formula θ^χ such that $\chi \dashv\vdash_2 \theta^\chi$
- for each γ-ary $\triangle \in \mathtt{Op}_2 - \mathtt{Op}_1$, there is a Σ_1-schema $\Theta^\triangle(\pi)$ such that for all $\rho \in \mathcal{L}_2^\gamma$, $\triangle(\rho) \dashv\vdash_2 \Theta^\triangle(\rho)$.

A definitional extension of a language is likewise defined.

The following is essentially due to Wójcicki [1988, Lemma 1.7.8, p. 62], though the result below is significantly generalized.[7]

Proposition 3.3.19 *(Wójcicki's Lemma)*. Let L_1 and L_2 be intensional logics where $\mathtt{At}_1 \subseteq \mathtt{At}_2$ and $\mathtt{Op}_1 \subseteq \mathtt{Op}_2$. The following are equivalent:

(a) $\mathsf{L}_1 :\subseteq \mathsf{L}_2$.

(b) There is a $\mathtt{t}\colon \mathsf{L}_2 \multimap \mathsf{L}_1$ such that for all $\phi \in \mathcal{L}_1$, $\mathtt{t}(\phi) = \phi$.

(c) There is a map $\alpha\colon (\mathtt{At}_2 - \mathtt{At}_1) \to \mathcal{L}_1$ and a map σ mapping each $\triangle \in \mathtt{Op}_2^\gamma - \mathtt{Op}_1^\gamma$ to a Σ_1-schema with γ-many parameters such that $\vdash_2$ is the smallest consequence relation $\vdash^*$ over $\mathcal{L}_2$ where $\vdash_1 \subseteq \vdash^*$ and for any

[7] **Proposition 3.3.19** encodes two generalizations of Wójcicki's lemma. First, we do not require that $\mathtt{At}_1 = \mathtt{At}_2$. Second, we allow infinitary operators. This second generalization requires a substantial modification to part (c). In particular, the last step of the proof of what corresponds in Wójciciki's version to the implication from (b) to (c) is verified "by a straightforward inductive argument". This step in our version is much more complicated because of this generalization, however.

Σ_2-schema $\Theta(\pi)$ and any $\rho_1, \rho_2 \in \mathcal{L}_2^{|\pi|}$, if for each $\beta < |\pi|$, one of the following holds:

(i) $\rho_2(\beta) = \rho_1(\beta)$

(ii) $\rho_2(\beta) = \alpha(\rho_1(\beta))$ and $\rho_1 \in \mathtt{At}_2 - \mathtt{At}_1$

(iii) $\rho_2(\beta) = \sigma(\triangle)(\rho_3)$ where $\rho(\beta) = \triangle(\rho_3)$

then $\Theta(\rho) \dashv\vdash^* \Theta(\rho')$.

Part (a) just says L_2 extends L_1 with some definable atomics and operators. Part (b) says one can translate the new expressions in $\mathcal{L}_2$ into $\mathcal{L}_1$ without touching the old expressions. Part (c) says there is a way of assigning $\mathcal{L}_1$-definitions to new $\mathcal{L}_2$-expressions so that L_2 is the smallest logic that validates replacement of an expression with its $\mathcal{L}_1$-definition.

Proof:

(a) ⇒ (b). Define $\mathtt{t}$ as follows:

$$\mathtt{t}(\phi) = \begin{cases} \phi & \text{if } \phi \in \mathtt{At}_1 \\ \theta^\phi & \text{if } \phi \in \mathtt{At}_2 - \mathtt{At}_1 \\ \triangle(\mathtt{t} \circ \rho) & \text{if } \phi = \triangle(\rho) \text{ where } \triangle \in \mathtt{Op}_1 \\ \Theta^\triangle(\mathtt{t} \circ \rho) & \text{if } \phi = \triangle(\rho) \text{ where } \triangle \in \mathtt{Op}_2 - \mathtt{Op}_1. \end{cases}$$

We first show that $\phi \dashv\vdash_2 \mathtt{t}(\phi)$ for all $\phi \in \mathcal{L}_2$. The proof is by induction. If $\phi \in \mathtt{At}_2$, then clearly $\phi \dashv\vdash_2 \mathtt{t}(\phi)$. Next, let $\triangle \in \mathtt{Op}_2$ and let $\rho \in \mathcal{L}_2^\gamma$ such that $\rho \dashv\vdash_2 \mathtt{t} \circ \rho$. Since L_2 is intensional, $\triangle(\rho) \dashv\vdash_2 \triangle(\mathtt{t} \circ \rho)$. If $\triangle \in \mathtt{Op}_1$, then $\triangle(\rho) \dashv\vdash_2 \triangle(\mathtt{t} \circ \rho) = \mathtt{t}(\triangle(\rho))$. Otherwise, $\triangle(\mathtt{t} \circ \rho) \dashv\vdash_2 \Theta^\triangle(\mathtt{t} \circ \rho) = \mathtt{t}(\triangle(\rho))$. This completes the induction.

Now, since $\mathsf{L}_1 \subseteq \mathsf{L}_2$, $\mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi)$ iff $\mathtt{t}[\Gamma] \vdash_1 \mathtt{t}(\phi)$. Hence, $\Gamma \vdash_2 \phi$ iff $\mathtt{t}[\Gamma] \vdash_2 \mathtt{t}(\phi)$ iff $\mathtt{t}[\Gamma] \vdash_1 \mathtt{t}(\phi)$. So $\mathtt{t}\colon \mathsf{L}_2 \multimap \mathsf{L}_1$.

(b) ⇒ (c). Assume without loss of generality that $\mathtt{t}$ is defined as above (since any such translation must be of this form). We let $\alpha(\chi) = \mathtt{t}(\chi)$ and $\sigma(\triangle) = \Theta^\triangle$. Let $\vdash^*$ be the smallest consequence relation over $\mathcal{L}_2$ satisfying the properties in (c).[a] Since $\mathtt{t}(\phi) = \phi$ for $\phi \in \mathcal{L}_1$ and since $\mathtt{t}\colon \mathsf{L}_2 \rightsquigarrow \mathsf{L}_1$, for all $\Gamma \subseteq \mathcal{L}_1$ and $\phi \in \mathcal{L}_1$:

$$\begin{aligned} \Gamma \vdash_1 \phi &\Leftrightarrow \mathtt{t}[\Gamma] \vdash_1 \mathtt{t}(\phi) \\ &\Leftrightarrow \Gamma \vdash_2 \phi. \end{aligned}$$

Thus, $\vdash_1 \subseteq \vdash_2$. Moreover, since L_2 is intensional, by induction (using **Proposition 1.3.15**), $\vdash_2$ satisfies the replacement property in part (c). Hence, $\vdash^* \subseteq \vdash_2$. So we just need to show that $\vdash_2 \subseteq \vdash^*$.

Suppose $\Gamma \vdash_2 \phi$. Then $\mathtt{t}[\Gamma] \vdash_1 \mathtt{t}(\phi)$, and so $\mathtt{t}[\Gamma] \vdash^* \mathtt{t}(\phi)$ by (i).

> **Claim**: Let $\Theta(\pi)$ be any Σ_2-schema and let $\rho_1, \rho_2 \in \mathcal{L}_2^{|\pi|}$. Suppose that for each $\beta < |\pi|$, either $\rho_2(\beta) = \rho_1(\beta)$ or $\rho_2(\beta) = \mathtt{t}(\rho_1(\beta))$. Then $\Theta(\rho) \dashv\vdash^* \Theta(\rho')$.

Assuming the claim can be shown, it follows that $\mathtt{t}(\phi) \dashv\vdash^* \phi$ for all $\phi \in \mathcal{L}_2$. Hence, $\Gamma \vdash^* \phi$. Therefore, $\vdash_2 \subseteq \vdash^*$. The rest of the proof of this part is spent on proving the claim.

The claim is proved by induction on $\mathcal{L}_2$. First, the atomic case. Let $\Theta(\pi)$ be a Σ_2-schema and let $\rho_1, \rho_2 \in \mathcal{L}_2^{|\pi|}$ be such that for each $\beta < |\pi|$, either $\rho_2(\beta) = \rho_1(\beta)$ or $\rho_2(\beta) = \alpha(\rho_1(\beta))$ where $\rho_1(\beta) \in \mathtt{At}_2 - \mathtt{At}_1$. Then by definition of $\vdash^*$ (and $\mathtt{t}$), $\Theta(\rho_1) \dashv\vdash^* \Theta(\rho_2)$.

Next, the inductive case. Suppose $\Phi \subseteq \mathcal{L}_2$ is such that the following holds: if $\Theta(\pi)$ is a Σ_2-schema and if $\rho_1, \rho_2 \in \mathcal{L}_2^{|\pi|}$ are such that for all $\beta < |\pi|$, either $\rho_2(\beta) = \rho_1(\beta)$ or $\rho_2(\beta) = \mathtt{t}(\rho_1(\beta))$ where $\rho_1(\beta) \in \Phi$, then $\Theta(\rho_1) \dashv\vdash^* \Theta(\rho_2)$. (The atomic case showed that $\mathtt{At}_2$ was such a Φ.) We now show that this also holds for Φ^+, where:

$$\Phi^+ := \Phi \cup \{\triangle(\rho) \mid \exists\gamma:\ \rho \in \Phi^\gamma \ \&\ \triangle \in \mathtt{Op}_2^\gamma\}.$$

Let $\Theta(\pi)$ be a Σ_2-schema and let $\rho_1, \rho_2 \in \mathcal{L}_2^{|\pi|}$ be such that for all $\beta < |\pi|$, either $\rho_2(\beta) = \rho_1(\beta)$ or $\rho_2(\beta) = \mathtt{t}(\rho_1(\beta))$ where $\rho_1(\beta) \in \Phi^+$. Define $\rho_3 \in \mathcal{L}_2^{|\pi|}$ so that for all $\beta < |\pi|$:

$$\rho_3(\beta) = \begin{cases} \rho_1(\beta) & \text{if } \rho_2(\beta) = \rho_1(\beta) \\ \rho_1(\beta) & \text{if } \rho_1(\beta) \in \Phi \\ \triangle(\rho) & \text{if } \rho_1(\beta) \notin \Phi \ \&\ \rho_1(\beta) = \triangle(\rho) \ \&\ \triangle \in \mathtt{Op}_1 \\ \sigma(\triangle)(\rho) & \text{if } \rho_1(\beta) \notin \Phi \ \&\ \rho_1(\beta) = \triangle(\rho) \ \&\ \triangle \notin \mathtt{Op}_1. \end{cases}$$

Then for any $\beta < |\pi|$, either $\rho_3(\beta) = \rho_1(\beta)$ (first three cases) or $\rho_3(\beta) = \sigma(\triangle)(\rho)$ where $\rho_1 = \triangle(\rho)$. So $\Theta(\rho_1) \dashv\vdash^* \Theta(\rho_3)$ by definition of $\vdash^*$. Thus, it suffices to show that $\Theta(\rho_3) \dashv\vdash^* \Theta(\rho_2)$. This will be done by reconceiving $\Theta(\rho_3)$ as the instantiation of another schema with another sequence and then using the inductive hypothesis on this new schema.

Now, we define a new Σ_2-schema $\Theta'(\pi')$ as follows. First, let $\beta < |\pi|$. We define the sequence π' as follows. If for some $\triangle \in \mathtt{Op}_1$, $\rho_1(\beta) = \triangle(\rho)$ and $\rho_1(\beta) \notin \Phi$ (so $\rho_3(\beta) = \triangle(\rho)$), then define $\pi'(\beta) = \triangle(\pi^\beta)$, where π^β are some distinct parameters not used yet. If for some $\triangle \in \mathtt{Op}_2 - \mathtt{Op}_1$, $\rho_1(\beta) = \triangle(\rho)$ and $\rho_1(\beta) \notin \Phi$ (so $\rho_3(\beta) = \sigma(\triangle)(\rho)$), then define $\pi'(\beta) = \sigma(\triangle)(\pi^\beta)$, where π^β are some distinct parameters not used yet. Otherwise, define $\pi'(\beta) = \pi(\beta)$. This defines π'. Now define $\Theta'(\pi')$ to be the result of replacing each $\pi(\beta)$ in Θ with $\pi'(\beta)$.

Define the instantiation ι_1 so that:

$$\begin{array}{rcl} \iota_1(\pi(\beta)) & = & \rho_1(\beta) \\ \iota_1(\pi^\beta(\gamma)) & = & \rho(\gamma) \text{ where } \rho_1 = \triangle(\rho) \text{ and } \gamma < \beta. \end{array}$$

Then it is easy to check that $\Theta(\rho_3) = \Theta'[\iota_1]$ (ι_1 just fills in the argument places for the $\triangle$s and $\sigma(\triangle)$s built into Θ' that yield $\Theta(\rho_3)$). Define also the instantiation ι_2 so that:

$$\begin{array}{rcl} \iota_2(\pi(\beta)) & = & \begin{cases} \rho_1(\beta) & \text{if } \rho_2(\beta) = \rho_1(\beta) \\ \mathtt{t}(\rho_1(\beta)) & \text{if } \rho_2(\beta) \neq \rho_1(\beta) \end{cases} \\ \iota_2(\pi^\beta(\gamma)) & = & \mathtt{t}(\rho(\gamma)) \text{ where } \rho_1 = \triangle(\rho) \text{ and } \gamma < \beta. \end{array}$$

Then $\Theta'[\iota_2] = \Theta(\rho_2)$ by definition of $\mathtt{t}$. But $\Theta'[\iota_1] \dashv\vdash^* \Theta'[\iota_2]$ by the inductive hypothesis (ι_2 is just the result of applying $\mathtt{t}$ to some $\iota_1(\xi)$s that are in Φ). So $\Theta(\rho_3) \dashv\vdash^* \Theta(\rho_2)$.

(c) ⇒ (a). It suffices to show that $\mathsf{L}_1 \sqsubseteq \mathsf{L}_2$ (i.e., that $\vdash_1 = \vdash_2 \restriction_{\mathcal{L}_1}$), since the properties in (c) guarantee that $\mathsf{L}_1 :\sqsubseteq \mathsf{L}_2$ if $\mathsf{L}_1 \sqsubseteq \mathsf{L}_2$. Define $\mathtt{t}$ as follows:

$$\mathtt{t}(\phi) = \begin{cases} \phi & \text{if } \phi \in \mathtt{At}_1 \\ \alpha(\phi) & \text{if } \phi \in \mathtt{At}_2 - \mathtt{At}_1 \\ \triangle(\mathtt{t} \circ \rho) & \text{if } \phi = \triangle(\rho) \text{ where } \triangle \in \mathtt{Op}_1 \\ \sigma(\triangle)(\mathtt{t} \circ \rho) & \text{if } \phi = \triangle(\rho) \text{ where } \triangle \in \mathtt{Op}_2 - \mathtt{Op}_1. \end{cases}$$

Observe that $\mathtt{t}(\phi) = \phi$ if $\phi \in \mathcal{L}_1$ by induction. Hence, if $\Gamma \nvdash_1 \phi$, then $\mathtt{t}[\Gamma] \nvdash_1 \mathtt{t}(\phi)$. Define:

$$\vdash^* \quad := \quad \vdash_2 - \{\langle \Gamma, \phi \rangle \mid \mathtt{t}[\Gamma] \nvdash_1 \mathtt{t}(\phi)\}.$$

By definition, if $\mathtt{t}[\Gamma] \nvdash_1 \mathtt{t}(\phi)$, then $\Gamma \nvdash^* \phi$. So we just need to show that $\vdash^* = \vdash_2$ (in particular, that $\vdash_2 \subseteq \vdash^*$).

To show that $\vdash^* = \vdash_2$, it suffices to show the following: To show that $\vdash^* = \vdash_2$, it suffices to show the following:

(i) $\vdash^*$ is a consequence relation

(ii) $\vdash_1 \subseteq \vdash^*$

(iii) $\vdash^*$ is intensional

(iv) for all $\chi \in \mathtt{At}_2 - \mathtt{At}_1$: $\mathtt{t}(\chi) = \mathtt{t}(\alpha(\chi))$

(v) for all $\triangle \in \mathtt{Op}_2 - \mathtt{Op}_1$ and all $\rho \in \mathcal{L}_2^{\gamma}$: $\mathtt{t}(\triangle(\rho)) = \mathtt{t}(\sigma(\triangle)(\rho))$.

If we show (iii)–(v), then $\vdash^*$ satisfies the replacement property from part (c). Together with (i)–(ii), this implies that $\vdash^* \subseteq \vdash_2$.

(i) By **Proposition 1.1.9**, we just need to show that $\vdash^*$ is monotonically reflexive and transitive. First, monotonic reflexivity. If $\phi \in \Gamma$, then $\mathtt{t}(\phi) \in \mathtt{t}[\Gamma]$. So $\Gamma \vdash_2 \phi$ and $\mathtt{t}[\Gamma] \vdash_1 \mathtt{t}(\phi)$. Thus, $\Gamma \vdash^* \phi$.

Next, transitivity. If $\Gamma \vdash^* \Delta \vdash^* \phi$, that means $\Gamma \vdash_2 \Delta \vdash_2 \phi$ and $\mathtt{t}[\Gamma] \vdash_1 \mathtt{t}[\Delta] \vdash_1 \mathtt{t}(\phi)$. So $\Gamma \vdash_2 \phi$ and $\mathtt{t}[\Gamma] \vdash_1 \mathtt{t}(\phi)$. Thus, $\Gamma \vdash^* \phi$.

(ii) If $\Gamma \vdash_1 \phi$, then $\Gamma \vdash_2 \phi$ and $\mathtt{t}[\Gamma] \vdash_1 \mathtt{t}(\phi)$ (since $\mathtt{t}$ maps every $\mathcal{L}_1$-formula to itself). Thus, $\Gamma \vdash^* \phi$. Therefore, $\vdash_1 \subseteq \vdash^*$.

(iii) Let $\triangle \in \mathtt{Op}_2$. Then if $\rho_1 \dashv\vdash^* \rho_2$, then by definition of $\vdash^*$, $\rho_1 \dashv\vdash_2 \rho_2$ and $\mathtt{t} \circ \rho_1 \dashv\vdash_1 \mathtt{t} \circ \rho_2$. Since L_2 is intensional, $\triangle(\rho_1) \dashv\vdash_2 \triangle(\rho_2)$. If $\triangle \in \mathtt{Op}_1$, then since L_1 is intensional:

$$\mathtt{t}(\triangle(\rho_1)) = \triangle(\mathtt{t} \circ \rho_1) \dashv\vdash_1 \triangle(\mathtt{t} \circ \rho_2) = \mathtt{t}(\triangle(\rho_2)).$$

If $\triangle \notin \mathtt{Op}_1$, then again since L_1 is intensional:

$$\mathtt{t}(\triangle(\rho_1)) = \sigma(\triangle)(\mathtt{t} \circ \rho_1) \dashv\vdash_1 \sigma(\triangle)(\mathtt{t} \circ \rho_2) = \mathtt{t}(\triangle(\rho_2)).$$

Either way, $\triangle(\rho_1) \dashv\vdash^* \triangle(\rho_2)$.

(iv) By definition.

(v) By induction, since $\sigma(\triangle)$ is a Σ_1-schema.

This completes the proof. ■

[a] We know there is a consequence relation with these properties since the trivial consequence relation has them. The smallest such consequence relation can be taken to just be the intersection of all such consequence relations.

Proposition 3.3.20 *(Alternative Characterization of Schematic Equivalence).* Let L_1 and L_2 be intensional logics. The following are equivalent:

(a) $\mathsf{L}_1 \multimapboth \mathsf{L}_2$.

(b) There are some intensional L_1', L_2', and L_3 such that $\mathsf{L}_1 \overset{\circ}{=} \mathsf{L}_1' :\subseteq \mathsf{L}_3$ and $\mathsf{L}_2 \overset{\circ}{=} \mathsf{L}_2' :\subseteq \mathsf{L}_3$. In a diagram:

$$\begin{tikzcd}
 & \mathsf{L}_3 & \\
\mathsf{L}_1' \arrow[ur, ":\subseteq"] & & \mathsf{L}_2' \arrow[ul, "\supseteq:"'] \\
\mathsf{L}_1 \arrow[u, "\overset{\circ}{=}"] & \leftrightsquigarrow & \mathsf{L}_2 \arrow[u, "\overset{\circ}{="}']
\end{tikzcd}$$

Proof:

(a) ⇒ (b). We start by defining L_1' and L_2' from L_1 and L_2. Define $\Sigma_1' = \langle \mathtt{At}_1', \mathtt{Op}_1' \rangle$ so that:

- $\mathtt{At}_1' = \{\langle \psi, 1\rangle \mid \psi \in \mathtt{At}_1\}$, and
- for each $\triangle \in \mathtt{Op}_1^{\gamma}$, we define $\triangle^*$ to be the function takes a γ-sequence ρ and returns the pair $\langle\langle \triangle, 1\rangle, \rho\rangle$
- $\mathtt{Op}_1' = \{\triangle^* \mid \triangle \in \mathtt{Op}_1\}$.

Define the map $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_1'$ so that $\mathtt{t}(\chi) = \langle \chi, 1\rangle$ if $\chi \in \mathtt{At}_1$, and $\mathtt{t}(\triangle(\rho)) = \langle\langle \triangle, 1\rangle, \mathtt{t} \circ \rho\rangle$. Clearly, $\mathtt{t}$ is schematic. Moreover, since Σ_1 is unambiguous, $\mathtt{t}$ is bijective. Now, define $\vdash_1$ so that $\mathtt{t}[\Gamma] \vdash_1' \mathtt{t}(\phi)$ iff $\Gamma \vdash_1 \phi$. Let $\mathsf{L}_1' = \langle \mathcal{L}_1', \vdash_1' \rangle$. Then $\mathtt{t}\colon \mathsf{L}_1 \overset{\circ}{=} \mathsf{L}_1'$ and L_1' is intensional. We can also define L_2' and a $\mathtt{s}$ in a likewise manner so that $\mathtt{s}\colon \mathsf{L}_2 \overset{\circ}{=} \mathsf{L}_2'$ (using '2', say, instead of '1' to ensure that $\mathcal{L}_2' \cap \mathcal{L}_1' = \varnothing$). Hence, $\mathsf{L}_1' \multimapboth \mathsf{L}_2'$. So it suffices to show that L_1' and L_2' have a common definitional extension.

Let $\mathtt{t}', \mathtt{s}'\colon \mathsf{L}_1' \multimapboth \mathsf{L}_2'$. Define $\Sigma_3 = \langle \mathtt{At}_1' \cup \mathtt{At}_2', \mathtt{Op}_1' \cup \mathtt{Op}_2' \rangle$. Define $\vdash_3$ to be the smallest consequence $\vdash^*$ over $\mathcal{L}_3$ such that:[a]

(i) $\vdash^* \supseteq \vdash_1' \cup \vdash_2'$

(ii) $\langle \mathcal{L}_3, \vdash^* \rangle$ is intensional

(iii) for all $\chi \in \mathtt{At}_1'$, $\chi \dashv\vdash^* \mathtt{t}'(\chi)$
(iv) for all $\zeta \in \mathtt{At}_2'$, $\zeta \dashv\vdash^* \mathtt{s}'(\zeta)$
(v) for all $\triangle \in (\mathtt{Op}_1')^\gamma$ and all $\rho \in \mathcal{L}_3^\gamma$, $\triangle(\rho) \dashv\vdash^* \triangle^{\mathtt{t}'}(\rho)$
(vi) for all $\triangledown \in (\mathtt{Op}_2')^\gamma$ and all $\rho \in \mathcal{L}_3^\gamma$, $\triangledown(\rho) \dashv\vdash^* \triangledown^{\mathtt{s}'}(\rho)$.

Let $\mathsf{L}_3 = \langle \mathcal{L}_3, \vdash_3 \rangle$. Define $\mathtt{t}^* \colon \mathcal{L}_3 \to \mathcal{L}_1$ as follows:

$$\mathtt{t}^*(\phi) = \begin{cases} \phi & \text{if } \phi \in \mathtt{At}_1' \\ \mathtt{s}'(\phi) & \text{if } \phi \in \mathtt{At}_2' \\ \triangle(\mathtt{t}^* \circ \rho) & \text{if } \triangle \in \mathtt{Op}_1' \\ \triangle^{\mathtt{s}'}(\mathtt{t}^* \circ \rho) & \text{if } \triangle \in \mathtt{Op}_2'. \end{cases}$$

It is easy to check by induction that $\mathtt{t}^*(\phi) = \phi$ for all $\phi \in \mathcal{L}_1$. By **Proposition 3.3.19**, to show that $\mathsf{L}_1 :\subseteq \mathsf{L}_3$, it suffices to show that $\mathtt{t}^* \colon \mathsf{L}_3 \multimap \mathsf{L}_1$.

First, note that each of the following is easily proved by induction:

- $\mathtt{t}^*(\psi) \dashv\vdash_1 \mathtt{s}'(\psi)$ for all $\psi \in \mathcal{L}_2$
- $\mathtt{t}^*(\phi) \dashv\vdash_3 \phi$ for all $\phi \in \mathcal{L}_3$.

Thus, if $\mathtt{t}^*[\Gamma] \vdash_1 \mathtt{t}^*(\phi)$, then $\Gamma \vdash_3 \phi$. To show the converse, define $\vdash^*$ as follows:

$$\vdash^* \quad := \quad \vdash_3 - \{\langle \Gamma, \phi \rangle \mid \mathtt{t}^*[\Gamma] \nvdash_1 \mathtt{t}^*(\phi)\}\,.$$

If we show that $\vdash^*$ satisfies (i)–(vi), then it follows that $\vdash^* = \vdash_3$, in which case $\Gamma \vdash_3 \phi$ implies $\mathtt{t}^*[\Gamma] \vdash_3 \mathtt{t}^*(\phi)$ (and so $\mathtt{t}^* \colon \mathsf{L}_3 \multimap \mathsf{L}_1$).

First, note that $\vdash^*$ is a consequence relation, since it is monotonically reflexive and transitive. Now for (i)–(vi):

(i) Since $\mathtt{t}^*(\phi) = \phi$ for all $\phi \in \mathcal{L}_1$:

$$\Gamma \vdash_1 \phi \quad \Rightarrow \quad \Gamma \vdash_3 \phi \text{ and } \mathtt{t}^*[\Gamma] \vdash_1 \mathtt{t}^*(\phi) \quad \Rightarrow \quad \Gamma \vdash^* \phi.$$

Similarly, since $\mathtt{s}'$ is a translation and since $\mathtt{t}^*(\psi) \dashv\vdash_1 \mathtt{s}'(\psi)$ for all $\psi \in \mathcal{L}_2$:

$$\begin{aligned} \Gamma \vdash_2 \phi \quad &\Rightarrow \quad \Gamma \vdash_3 \phi \text{ and } \mathtt{s}'[\Gamma] \vdash_1 \mathtt{s}'(\phi) \\ &\Rightarrow \quad \Gamma \vdash_3 \phi \text{ and } \mathtt{t}^*[\Gamma] \vdash_1 \mathtt{t}^*(\phi) \\ &\Rightarrow \quad \Gamma \vdash^* \phi. \end{aligned}$$

So $\vdash_1 \cup \vdash_2 \subseteq \vdash^*$.

(ii) Suppose $\rho_1 \dashv\vdash^* \rho_2$. Then $\rho_1 \dashv\vdash_3 \rho_2$ and $\mathtt{t}^* \circ \rho_1 \dashv\vdash_1 \mathtt{t}^* \circ \rho_2$ by definition of $\vdash^*$. If $\triangle \in \mathtt{Op}_3$, then $\triangle(\rho_1) \dashv\vdash_3 \triangle(\rho_2)$ since L_3 is intensional. If $\triangle \in \mathtt{Op}'_1$, then since L_1 is intensional:

$$\mathtt{t}^*(\triangle(\rho_1)) = \triangle(\mathtt{t}^* \circ \rho_1) \dashv\vdash_1 \triangle(\mathtt{t}^* \circ \rho_2) = \mathtt{t}^*(\triangle(\rho_2)).$$

If $\triangle \in \mathtt{Op}'_2$, then since L_1 is intensional:

$$\mathtt{t}^*(\triangle(\rho_1)) = \triangle^{\mathtt{s}'}(\mathtt{t}^* \circ \rho_1) \dashv\vdash_1 \triangle^{\mathtt{s}'}(\mathtt{t}^* \circ \rho_2) = \mathtt{t}^*(\triangle(\rho_2)).$$

Hence, $\triangle(\rho_1) \dashv\vdash^* \triangle(\rho_2)$.

(iii) Let $\chi \in \mathtt{At}'_1$. Since $\mathtt{t}^*(\psi) \dashv\vdash_1 \mathtt{s}'(\psi)$ for $\psi \in \mathcal{L}_2$:

$$\mathtt{t}^*(\chi) = \chi \dashv\vdash_1 \mathtt{s}'(\mathtt{t}'(\chi)) \dashv\vdash_1 \mathtt{t}^*(\mathtt{t}'(\chi)).$$

And since $\chi \dashv\vdash_3 \mathtt{t}'(\chi)$, $\chi \dashv\vdash^* \mathtt{t}'(\chi)$.

(iv) Let $\zeta \in \mathtt{At}'_2$. Then $\mathtt{t}^*(\zeta) = \mathtt{s}'(\zeta) = \mathtt{t}^*(\mathtt{s}(\zeta))$ (since $\mathtt{t}^*(\phi) = \phi$ for $\phi \in \mathcal{L}_1$). And since $\zeta \dashv\vdash_3 \mathtt{s}'(\zeta)$, $\zeta \dashv\vdash^* \mathtt{s}'(\zeta)$.

(v) Let $\triangle \in (\mathtt{Op}'_1)^\gamma$ and let $\rho \in \mathcal{L}_3^\gamma$. Then $\triangle(\rho) \dashv\vdash_3 \triangle^{\mathtt{t}'}(\rho)$. Moreover:

$$\mathtt{t}^*(\triangle(\rho)) = \triangle(\mathtt{t}^* \circ \rho) \dashv\vdash_1 \triangle^{\mathtt{s}'\circ\mathtt{t}'}(\mathtt{t}^* \circ \rho) = \mathtt{t}^*(\triangle^{\mathtt{t}'}(\rho)).$$

So $\triangle(\rho) \dashv\vdash^* \triangle^{\mathtt{t}'}(\rho)$.

(vi) Let $\triangledown \in (\mathtt{Op}'_2)^\gamma$ and let $\rho \in \mathcal{L}_3^\gamma$. Then $\triangledown(\rho) \dashv\vdash_3 \triangledown^{\mathtt{s}'}(\rho)$. Moreover:

$$\mathtt{t}^*(\triangledown(\rho)) = \triangledown^{\mathtt{s}'}(\mathtt{t}^* \circ \rho) = \mathtt{t}^*(\triangledown^{\mathtt{s}'}(\rho)).$$

So $\triangledown(\rho) \dashv\vdash^* \triangle^{\mathtt{s}'}(\rho)$.

This completes the proof.

(b) ⇒ (a). It suffices to show that if $\mathsf{L}_1 :\subseteq \mathsf{L}_3$ and $\mathsf{L}_2 :\subseteq \mathsf{L}_3$, then $\mathsf{L}_1 \multimapboth \mathsf{L}_2$. We define $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$ as follows. First, let $\chi \in \mathtt{At}_1$. Set $\mathtt{t}(\chi) = \chi$ if $\chi \in \mathtt{At}_2$ and otherwise pick an arbitrary $\theta^\chi \in \mathcal{L}_2$ such that $\chi \dashv\vdash_3 \theta^\chi$ and set $\mathtt{t}(\chi) = \theta^\chi$. Next, let $\phi = \triangle(\rho)$, where $\mathtt{t} \circ \rho$ is already defined. Set $\mathtt{t}(\triangle(\rho)) = \triangle(\mathtt{t} \circ \rho)$ if $\triangle \in \mathtt{Op}_2$ and otherwise set $\mathtt{t}(\triangle(\rho)) = \Theta^\triangle(\mathtt{t} \circ \rho)$, where $\Theta^\triangle(\rho') \dashv\vdash_3 \triangle(\rho')$ for all $\rho' \in \mathcal{L}_3^\gamma$. Define $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$ similarly.

Claim: For all $\phi \in \mathcal{L}_1$, $\phi \dashv\vdash_3 \mathtt{t}(\phi)$.

Subproof: By induction. The atomic case is trivial. Let $\triangle \in \mathtt{Op}_1^{\gamma}$, and let $\rho \in \mathcal{L}_1^{\gamma}$ be such that $\rho(\beta) \dashv\vdash_3 \mathtt{t}(\rho(\beta))$ for all $\beta < \gamma$. If $\triangle \in \mathtt{Op}_2^{\gamma}$, then $\mathtt{t}(\triangle(\rho)) = \triangle(\mathtt{t} \circ \rho) \dashv\vdash_3 \triangle(\rho)$ since L_3 is intensional. If $\triangle \notin \mathtt{Op}_2^{\gamma}$, then $\mathtt{t}(\triangle(\rho)) = \Theta^{\triangle}(\mathtt{t} \circ \rho) \dashv\vdash_3 \triangle(\mathtt{t} \circ \rho) \dashv\vdash_3 \triangle(\rho)$. Either way, $\triangle(\rho) \dashv\vdash_3 \mathtt{t}(\triangle(\rho))$. ■

Thus, $\Gamma \vdash_1 \phi$ iff $\Gamma \vdash_3 \phi$ iff $\mathtt{t}[\Gamma] \vdash_3 \mathtt{t}(\phi)$. Likewise for $\mathtt{s}$. So both $\mathtt{t}$ and $\mathtt{s}$ are schematic translations. Moreover, let $\psi \in \mathcal{L}_2$. Since $\psi \dashv\vdash_3 \mathtt{s}(\psi) \dashv\vdash_3 \mathtt{t}(\mathtt{s}(\psi))$ and since $\mathsf{L}_2 :\subseteq \mathsf{L}_3$, $\psi \dashv\vdash_2 \mathtt{t}(\mathtt{s}(\psi))$. Hence, by **Proposition 2.2.4**, $\mathtt{t}, \mathtt{s} \colon \mathsf{L}_1 \multimapboth \mathsf{L}_2$. ■

[a] Again, there is a consequence relation with these properties, since the trivial consequence relation has these properties, and the smallest can be taken to be the intersection of all such consequence relations.

One way to think about the significance of **Proposition 3.3.20** is as follows: if two logics are schematically equivalent, one can combine them into a single logic in a way that faithfully preserves the relationship between the two logics. Normally, we want to say that when we have a translation from L_1 into L_2 that an $\mathcal{L}_1$-formula and its translation are "equivalent". But there is no easy way to state this in general, since the notion of equivalence is always relative to a logic. The best we can say, in general, is that an $\mathcal{L}_1$-formula and its translation are inferentially equivalent, in the sense that they play the same inferential roles in their respective logics. But when you have schematic equivalence between L_1 and L_2, you can find a combined logic where you *can* directly state that an $\mathcal{L}_1$-formula and its translation are equivalent.

§ 3.4 Schematic Interdependence

Proposition 3.3.16 shows that schematic intertranslatability is, in general, too strong of a constraint to measure the expressive power of a language if we want propositional modal logic to be expressively weaker than first-order logic. But that case does suggest a natural modification of the notion of schematicity.

The problem with the definition of schematicity (**Definition 3.3.1**) is that sometimes a translation can only be defined *simultaneously* with other translations. This is what the standard translation of modal logic into first-order logic illustrates. But intuitively, that should not matter. What is important is not that the translation of a complex formula is strictly a schema of the translation of the parts, but rather that the translation of a complex formula is uniform and fixed solely by its syntactic structure. This motivates a more general notion of schematicity along the following lines:[8]

Definition 3.4.1 *(Interdependent Schematicity)*. Let L_1 and L_2 be logics and let $\mathtt{T}$ be a class of maps $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Where $\rho \in \mathcal{L}_1^{\gamma}$ and $\tau \in \mathtt{T}^{\gamma}$, define the map $(\tau \cdot \rho)\colon \gamma \to \mathcal{L}_2$ so that:

$$(\tau \cdot \rho)(\beta) = \tau(\beta)(\rho(\beta)).$$

We say $\mathtt{T}$ is ***schematically interdependent*** if for each $\mathtt{t} \in \mathtt{T}$ and each $\triangle \in \mathtt{Op}_1^{\gamma}$, there is a Σ_2-schema $\Theta^{\triangle}(\pi)$ with γ-many parameters π and a $\tau \in \mathtt{T}^{\gamma}$ such that $\mathtt{t}(\triangle(\rho)) = \Theta^{\triangle}(\tau \cdot \rho)$ for all $\rho \in \mathcal{L}_1^{\gamma}$. As before, we write "$\Theta^{\triangle}$" as "$\triangle^{\mathtt{t}}$". For convenience, we write "$\triangle^{\mathtt{t}}_{\tau}(\rho)$" instead of "$\triangle^{\mathtt{t}}(\tau \cdot \rho)$".

A translation $\mathtt{t}\colon \mathsf{L}_1 \rightsquigarrow \mathsf{L}_2$ is ***recursive*** if it is a member of a schematically interdependent class of translations from $\mathcal{L}_1$ to $\mathcal{L}_2$. We write "$\mathtt{t}\colon \mathsf{L}_1 \multimap \mathsf{L}_2$" to mean $\mathtt{t}$ is a recursive translation. Likewise, we use "$\rightleftarrows$" for recursive intertranslatability, "$\bullet\!\!-\!\!\bullet$" for recursive equivalence, and "$\dot{=}$" for recursive isomorphism (where these notions are defined in the obvious way).

Note that as we are defining it, for a translation to count as recursive, it must be a member of a schematically interdependent class of *translations*. It cannot be merely that there are maps in terms of which it is schematically interdependently defined. If we only required recursive translations to be members of schematically interdependent classes of maps, then it would be too easy to get recursive translations in general.

If $\mathtt{t}$ is schematic, then it is a member of a schematically interdependent set, but not *vice versa*, as the standard translation from **K** into **FOL** shows. So the fact that no schematic translation from **FOL** to **CPL** exists does not immediately imply that there is no recursive translation from **FOL** to **CPL**. Similarly for **IPL** and **CPL**. Fortunately, with a little more work, we can achieve this result as well.

Proposition 3.4.2 *(No Recursive Translation from First-Order Logic into Propositional Logic)*. There is no $\mathtt{t}\colon$ **FOL** $\multimap$ **CPL**.

[8]The definition is inspired by the definition of recursive translations from French 2010, p. 16, who attributes the definition to Steven Kuhn. We could also require recursive translations to translate atomic formulas schematically. Such a constraint seems well-motivated, but it was not included in this definition for purposes of generality, as it was not necessary in the results to follow.

Proof: Suppose there were such a $\mathtt{t}$. Let $\Theta(\pi)$ be a $\Sigma_{\mathbf{CPL}}$-schema with a single parameter π and let $\mathtt{t}'$: $\mathbf{FOL} \rightsquigarrow \mathbf{CPL}$ be such that $\mathtt{t}(\exists x\, \phi) = \Theta(\mathtt{t}'(\phi))$. Since $\top \equiv_{\mathbf{CPL}} \mathtt{t}(\top)$, we have:

$$\vDash_{\mathbf{CPL}} \mathtt{t}(\top) \equiv_{\mathbf{CPL}} \mathtt{t}(\exists x\, \top) = \Theta(\mathtt{t}'(\top)) \equiv_{\mathbf{CPL}} \Theta(\top).$$

Hence:

$$\begin{aligned} \mathtt{t}'(\phi) &\vDash_{\mathbf{CPL}} \mathtt{t}'(\phi) \leftrightarrow \top \\ &\vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\phi)) \leftrightarrow \Theta(\top) \\ &\vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\phi)) = \mathtt{t}(\exists x\, \phi). \end{aligned}$$

So $\mathtt{t}'(\phi) \vDash_{\mathbf{CPL}} \mathtt{t}(\exists x\, \phi)$ for all $\phi \in \mathbf{FOL}$. Now, $\Theta(\mathtt{t}'(\phi)), \neg\, \mathtt{t}'(\phi) \vDash_{\mathbf{CPL}} \Theta(\bot)$. Moreover:

$$\begin{aligned} \neg\, \mathtt{t}'(\bot) &\vDash_{\mathbf{CPL}} \mathtt{t}'(\bot) \leftrightarrow \bot \\ &\vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\bot)) \leftrightarrow \Theta(\bot). \end{aligned}$$

So we have:

$$\mathtt{t}(\exists x\, \phi), \neg\, \mathtt{t}'(\phi), \neg\, \mathtt{t}'(\bot) \vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\bot)).$$

But $\mathtt{t}'(\bot) \vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\bot))$, too. So using proof by cases on $\mathtt{t}'(\bot)$:

$$\mathtt{t}(\exists x\, \phi), \neg\, \mathtt{t}'(\phi) \vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\bot)) = \mathtt{t}(\exists x\, \bot) \equiv_{\mathbf{CPL}} \mathtt{t}(\bot).$$

Hence, $\mathtt{t}(\exists x\, \phi) \vDash_{\mathbf{CPL}} \mathtt{t}'(\phi) \vee \mathtt{t}(\bot)$. Moreover, the converse holds too, since $\mathtt{t}'(\phi) \vDash_{\mathbf{CPL}} \mathtt{t}(\exists x\, \phi)$ (from above) and $\mathtt{t}(\bot) \vDash_{\mathbf{CPL}} \mathtt{t}(\exists x\, \phi)$. So for all $\phi \in \mathbf{FOL}$:

$$\mathtt{t}(\exists x\, \phi) \equiv_{\mathbf{CPL}} \mathtt{t}'(\phi) \vee \mathtt{t}(\bot).$$

But it is easy to check that $\mathtt{t}(\phi \wedge \psi) \equiv_{\mathbf{CPL}} \mathtt{t}(\phi) \wedge \mathtt{t}(\psi)$ for any $\phi, \psi \in \mathbf{FOL}$. Likewise for $\mathtt{t}'$. Thus, we have:

$$\begin{aligned} \mathtt{t}(\exists x\, \phi \wedge \exists x\, \neg\, \phi) &\equiv_{\mathbf{CPL}} \mathtt{t}(\exists x\, \phi) \wedge \mathtt{t}(\exists x\, \neg\, \phi) \\ &\vDash_{\mathbf{CPL}} (\mathtt{t}'(\phi) \wedge \mathtt{t}'(\neg\, \phi)) \vee \mathtt{t}(\bot) \\ &\equiv_{\mathbf{CPL}} (\mathtt{t}'(\phi \wedge \neg\, \phi)) \vee \mathtt{t}(\bot) \\ &\equiv_{\mathbf{CPL}} \mathtt{t}'(\bot) \vee \mathtt{t}(\bot) \\ &\equiv_{\mathbf{CPL}} \mathtt{t}(\exists x\, \bot) \\ &\equiv_{\mathbf{CPL}} \mathtt{t}(\bot). \end{aligned}$$

But, $\exists x\, Fx \wedge \exists x\, \neg\, Fx \nvDash_{\mathbf{FOL}} \bot$, ↯. ■

Proposition 3.4.3 *(No Recursive Translation from Intuitionistic Logic into Classical Logic)*. There is no $\mathtt{t}\colon \mathbf{IPL} \multimap \mathbf{CPL}$.

Proof: Suppose there were a such a $\mathtt{t}$. Let $\Theta(\pi)$ be a $\Sigma_{\mathbf{CPL}}$-schema and let $\mathtt{t}'\colon \mathbf{IPL} \rightsquigarrow \mathbf{CPL}$ be such that $\mathtt{t}(\neg\phi) = \Theta(\mathtt{t}'(\phi))$. Then:

$$\top \equiv_{\mathbf{CPL}} \mathtt{t}(\top) = \mathtt{t}(\neg\bot) = \Theta(\mathtt{t}'(\bot)).$$

So $\vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\bot))$. Thus:

$$\begin{aligned} \mathtt{t}'(\neg\phi) &\vDash_{\mathbf{CPL}} \mathtt{t}'(\phi \leftrightarrow \bot) \\ &\vDash_{\mathbf{CPL}} \mathtt{t}'(\phi) \leftrightarrow \mathtt{t}'(\bot) \\ &\vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\phi)) \leftrightarrow \Theta(\mathtt{t}'(\bot)) \\ &\vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\phi)). \end{aligned}$$

So $\mathtt{t}'(\neg\phi) \vDash_{\mathbf{CPL}} \mathtt{t}(\neg\phi)$.

Since $\vDash_{\mathbf{CPL}} \mathtt{t}'(\bot) \vee \neg\mathtt{t}'(\bot)$ and $\vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\bot))$, we have:

$$\vDash_{\mathbf{CPL}} \Theta(\top) \vee \Theta(\bot).$$

Now, $\mathtt{t}(\bot) \vDash_{\mathbf{CPL}} \mathtt{t}(\neg\phi)$ trivially. So:

$$\begin{aligned} \neg\mathtt{t}(\neg\phi) &\vDash_{\mathbf{CPL}} \neg\mathtt{t}(\bot) \\ &\equiv_{\mathbf{CPL}} \neg\mathtt{t}(\neg\top) = \neg\Theta(\mathtt{t}'(\top)) \\ &\equiv_{\mathbf{CPL}} \neg\Theta(\top) \\ &\vDash_{\mathbf{CPL}} \Theta(\bot) \\ &\vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\neg\phi)). \end{aligned}$$

(The last stop holds since $\neg\mathtt{t}(\neg\phi) \vDash_{\mathbf{CPL}} \neg\mathtt{t}'(\neg\phi) \vDash_{\mathbf{CPL}} \mathtt{t}'(\neg\phi) \leftrightarrow \bot$.) Thus, $\neg\mathtt{t}(\neg\phi) \vDash_{\mathbf{CPL}} \Theta(\mathtt{t}'(\neg\phi)) = \mathtt{t}(\neg\neg\phi)$. Hence:

$$\begin{aligned} &\vDash_{\mathbf{CPL}} \mathtt{t}(\neg\phi) \vee \neg\mathtt{t}(\neg\phi) \\ &\vDash_{\mathbf{CPL}} \mathtt{t}(\neg\phi) \vee \mathtt{t}(\neg\neg\phi) \\ &\vDash_{\mathbf{CPL}} \mathtt{t}(\neg\phi \vee \neg\neg\phi). \end{aligned}$$

But $\nvDash_{\mathbf{IPL}} \neg p \vee \neg\neg p$, ↯. ■

So far, I have not been able to find a proof that generalizes this result to intermediate logics below **CPL** that satisfy weak excluded middle ($\vdash \neg\phi \vee \neg\neg\phi$).

§ 3.5 Limitations

Propositions 3.4.2 and **3.4.3** suggest that recursive translatability is a viable candidate for the precisification of the intuitive notion of expressive power. And given that it is able to capture our initial judgments about the relationship between modal logic and first-order logic, it is arguably more suited for that role than schematicity. Even so, there are limitations to the thesis that expressive power is to be measured by recursive translatability, some of which we will now discuss.

First, we motivated each of the different compositionality constraints explored in this chapter with the idea that adequate translations ought to preserve the syntactic structure of the language being translated and not just the (in)validity of its entailment relation. But throughout this chapter, we have been working with a relatively coarse-grained view of syntactic structure, whereby formulas are built from atomic formulas using some operators. More fine-grained syntaxes ought to be considered as well, however. For instance, one might impose syntactic structure on the atomic formulas by requiring that they be freely constructed out of terms and predicates. One might even isolate some atomic terms and atomic predicates as well as term-forming and predicate-forming operators out of which more complex atomic formulas are constructed (at which point, one might deem the term "atomic formula" inaccurate). Just as one might require adequate translations to preserve the meanings of individual operators, so too one might require adequate translations to preserve the meanings of terms, predicates, and other syntactic categories.

Whether such requirements are generally justified is another matter. Even the requirement that every schema be translated as a schema might seem too stringent for some purposes. As an example, consider an alternative formulation of first-order, which we call "$\mathbf{FOL}^{\lambda}$", where quantifiers of the form $\forall x$ are constructed out of a single quantifier operator $\forall$, which maps predicates to formulas, and a variable binding operator λx, which maps formulas to predicates.[9] It is natural to view these different versions of first-order logic as mere notational variants. Yet, since schemas are sequences built from the atomic formulas and the operators of a language, no schema of $\mathbf{FOL}^{\lambda}$ adequately translates the **FOL**-schema $\forall x\, \xi$. Similarly, **FOL** in its formulation here does not have any kind of predicate-to-formula or formula-to-predicate operators that could correlate with $\forall$ or λx. At minimum, before one can even address whether these languages are equiexpressive, one would need to develop a more generalized notion of a "schema" that would allow for constructions other than those given by **Definition 1.3.9**. And even then, we are left with a difficult question regarding what syntactic structures need to be preserved by adequate translations. For instance, does the fact that $\mathbf{FOL}^{\lambda}$ have predicate-to-formula operators (and *vice versa*) matter for the purposes of expressivity?

Another class of examples comes from languages with partial signatures (**Definition 1.3.8**). A natural example comes from the study of propositional logic. It is

[9] I have Seth Yalcin to thank for this example.

well-known that every **CPL**-formula is equivalent to a formula in disjunctive normal form, i.e., a disjunction of conjunctions of literals (atomic formulas and their negations).[10] Let $\mathbf{CPL}^{DNF}$ be the restriction of **CPL** to formulas in disjunctive normal form and where the conjunction operator $\wedge$ is partial in that $(\phi \wedge \psi)$ is defined in $\mathbf{CPL}^{DNF}$ iff ϕ and ψ are not disjunctions ($\vee$ is still total). On the one hand, there is a temptation to say that **CPL** and $\mathbf{CPL}^{DNF}$ are expressively equivalent. On the other hand, there does not seem to be any way to map the operators of **CPL** to schemas of $\mathbf{CPL}^{DNF}$, since it does not seem possible to define the translation of $\wedge$ and $\vee$ schematically so that the translation of $(\phi \wedge (\psi \vee \theta))$ is always in disjunctive normal form. For instance, a natural suggestion is to use the following as a translation:

$$\begin{array}{rcl} \mathtt{t}(p) & = & p \\ \mathtt{t}(\phi \wedge \psi) & = & (\mathtt{t}(\phi) \wedge \mathtt{t}(\psi)) \vee (\mathtt{t}(\phi) \wedge \mathtt{t}(\psi)) \\ \mathtt{t}(\phi \vee \psi) & = & \mathtt{t}(\phi) \vee \mathtt{t}(\psi). \end{array}$$

But then the translation of $p \wedge (q \vee r)$ is $(p \wedge (q \vee r)) \vee (p \wedge (q \vee r))$, which is not in disjunctive normal form. Intuitively, we want the translation of $p \wedge (q \vee r)$ to be something like $(p \wedge q) \vee (p \wedge r)$. While this is in disjunctive normal form, it does not contain $q \vee r$ as a subformula and it is not clear what subformula of this could possibly act as the translation of $q \vee r$. Nor is it clear how appealing to a collection of schematically interdependent mappings might help. A compositionality constraint that will render **CPL** and $\mathbf{CPL}^{DNF}$ equivalent while failing to render **CPL** and **FOL** equivalent is still wanted.

These examples suggest that recursivity is not perfect as a measure of expressive power. Perhaps there is a better measure out there waiting to be discovered. Or perhaps there simply is no unique measure of expressive power that does justice to all of our intuitions. Still, it is worth emphasizing that recursivity has thus far done better than other inference-preserving notions considered so far. Such observations about its limitations ought to be viewed as a recognition that recursivity is an idealization and as a call to investigate other measures of expressive power further.

[10] For simplicity, I assume this allows for "disjunctions" and "conjunctions" of length one, so that p counts as being in disjunctive normal form.

Chapter 4

Logical Space

In the previous chapters, we explored in depth the notion of a translation, i.e., a consequence-preserving map between the formulas of the language. The notion of translatability is especially promising as a precisification of the notion of expressive power (or, if you like, interpretability power) for logics. But of course, the notion of expressive power is often thought of as a property not of *logics* but of *languages*. And as we mentioned near the beginning of Chapter 2, translatability does not seem sufficient as an explication of the notion of expressive power for languages. So we return to the question with which we started: how do we precisify the expressivity of a *language*?

Intuitively, the expressive power of a language is connected to a language's ability to make distinctions between different possibilities. It is natural to think that one language is more expressive than another just in case the former recognizes more possibilities than the latter, i.e., just in case (a) any two possibilities that the latter language can distinguish can be distinguished by the former language, and (b) there are some possibilities that the former language can distinguish that are indistinguishable according to the latter language. Indeed, this is the most common approach found in the literature for defining a precise measure of a language's expressive power. This chapter categorizes different ways of making precise this notion of "distinguishing" among the possibilities and develops in more mathematical detail the relationship between these different notions.

The guiding metaphor throughout this chapter is of carving "logical space" (or "modal space"). Very roughly, we can think of logical space as the space of all genuine possibilities—that is, the space of all ways the world could have been.[1] Initially, this space is unstructured; there is no underlying metric, order, or anything of the sort placed on this space. That is where the role of the language comes to the foreground: the role of a language is to articulate structure on this space.

[1] The sense of "could" here should not be interpreted as metaphysical or epistemic. We should think of it less reductively in terms of "semantic" possibilities. See Schwarz 2018.

The simplest kind of structure a language can articulate is that of division. This is enacted by pairing a syntax (which for now can be thought of as a class of well-formed formulas) with a semantics (a relation between points in logical space and formulas). When a formula is interpreted by a semantics, it can be thought of as determining a cut in logical space between the possibilities that satisfy that formula and the possibilities that do not. A language might impose other kinds of structure as well, though our main focus in this chapter will be on division.

Using this guiding metaphor, then, we can think of the expressive power of a language as being determined by the structures it can articulate. Thus, two languages have equal expressive power if they each can articulate the same structures on logical space as the other. So for instance, if the main kind of structure we care about is division, then two languages have equal expressive power if every division one of the languages can make can be made by the other language. Put differently, two languages will have equal expressive power if every formula from one language is satisfied by the same possibilities that some other formula from the other language is satisfied by. This idea is explored in § 4.1.

However, this notion of expressive power, though quite useful for many practical purposes, is not sufficiently general for other purposes. Often times, we want to compare the expressive power of languages that are defined over different conceptions of logical space (or are at least not assumed to defined over the same conception of logical space). For instance, we might want to compare the expressive powers of classical and nonclassical languages, or of non-modal and modal languages, where the model space is not generally the same. Arguably, this should be achievable: expressive power is a measure of a language's ability to articulate a certain kind of *structure* on a logical space. It does not crucially hinge on the *contents* of that space. So in § 4.2, we explore ways that we can generalize the notions of expressive power from § 4.1 so as to be able to compare languages with competing conceptions of logical space. The metaphor used in this section is that of *transforming* logical space: if two languages can articulate the same structures, one should be able to transform the logical space of one language into the other while preserving that structure.

§ 4.1 Carving Logical Space

The guiding idea behind the notions of expressive power developed here is that of carving logical space: two languages should be deemed expressively equivalent if they carve logical space in the same way.[2] Each of the notions discussed below are motivated by this intuitive idea and each attempts to make it more precise.

[2]This notion crucially assumes that the existence of the semantic value of a formula does not depend on the point of evaluation. This might be called into question if one maintained an externalist view of semantic values. See Stalnaker 2012. For now I will set this issue aside. Still, I wish to thank Seth Yalcin for bringing this to my attention.

§ 4.1.1 Expressibility

The first notion of expressive power we discuss is perhaps the simplest: two languages are expressively equivalent if there is a translation from one to the other that preserves meaning. Here, "meaning" should be interpreted in terms of semantic value (**Definition 1.1.2**). In terms of logical space, two languages are expressively equivalent if every cut in logical space one language can make can equally be made in the other. We now make this idea precise.

Definition 4.1.1 *(C-language)*. Where C is a class of points, a ***C-language*** is a language $\mathbf{L}$ where $C_{\mathbf{L}} = C$.

Definition 4.1.2 *(Expressibility)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be C-languages. We say that $\mathbf{L}_1$ is ***expressible in*** (or ***included in***) $\mathbf{L}_2$ (written "$\mathbf{L}_1 \leqslant \mathbf{L}_2$") if for every $\mathcal{L}_1$-formula ϕ, there is an $\mathcal{L}_2$-formula ψ such that $[\![\phi]\!]_1 = [\![\psi]\!]_2$. We say that $\mathbf{L}_1$ is ***strictly expressible in*** (***strictly included in***) $\mathbf{L}_2$ (written "$\mathbf{L}_1 < \mathbf{L}_2$") if $\mathbf{L}_1 \leqslant \mathbf{L}_2$ but $\mathbf{L}_2 \nleqslant \mathbf{L}_1$. Finally, we say that $\mathbf{L}_1$ is ***(expressively) equivalent to*** $\mathbf{L}_2$ (written "$\mathbf{L}_1 \approx \mathbf{L}_2$") if $\mathbf{L}_1 \leqslant \mathbf{L}_2$ and $\mathbf{L}_2 \leqslant \mathbf{L}_1$.

With the axiom of choice (for classes), we can restate this definition in terms of a translation as follows:

Lemma 4.1.3 *(Expressibility and Translations)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be C-languages.

(a) $\mathbf{L}_1 \leqslant \mathbf{L}_2$ iff there is a function $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ such that for all $\phi \in \mathcal{L}_1$:

$$[\![\phi]\!]_1 = [\![\mathtt{t}(\phi)]\!]_2 .$$

(b) $\mathbf{L}_1 \approx \mathbf{L}_2$ iff there are functions $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$ such that for all $\phi \in \mathcal{L}_1$ and $\psi \in \mathcal{L}_2$:

$$[\![\phi]\!]_1 = [\![\mathtt{t}(\phi)]\!]_2 = [\![\mathtt{s}(\mathtt{t}(\phi))]\!]_1$$
$$[\![\psi]\!]_2 = [\![\mathtt{s}(\psi)]\!]_1 = [\![\mathtt{t}(\mathtt{s}(\psi))]\!]_2 .$$

Proof: (b) immediately follows from (a), so we just prove (a). The right-to-left direction is trivial. For the left-to-right direction, suppose $\mathbf{L}_1 \leqslant \mathbf{L}_2$. Define the set of $\mathcal{L}_2$-formulas $|\phi| := \{\psi \in \mathcal{L}_2 \mid [\![\phi]\!]_1 = [\![\psi]\!]_2\}$. By the axiom of choice, there is a function $\mathtt{t}$ that maps every $\phi \in \mathcal{L}_1$ to a member of $|\phi|$. ∎

Notation: We write "$\mathtt{t}\colon \mathbf{L}_1 \leqslant \mathbf{L}_2$" to indicate that $\mathbf{L}_1$ is expressively included in $\mathbf{L}_2$ via $\mathtt{t}$, i.e., for all $\phi \in \mathcal{L}_1$, $[\![\phi]\!]_1 = [\![\mathtt{t}(\phi)]\!]_2$.

Example 4.1.4 *(Hybrid Logic)*. Where $\mathtt{Prop} = \{p_1, p_2, p_3, \ldots\}$ and where $\mathtt{Var} = \{x_1, x_2, x_3, \ldots\}$ (as before), let $\mathcal{L}_\mathbf{H}$ be the following syntax:

$$\phi ::= p \mid x \mid \neg\phi \mid (\phi \wedge \phi) \mid \Box\phi \mid @_x\phi \mid {\downarrow}x.\phi.$$

Define $\mathsf{C}_\mathbf{H} := \mathsf{C}_\mathbf{FOL}$ and define $\Vdash_\mathbf{H}$ as follows (where $\twoheadrightarrow$ is some designated binary predicate and "$a \twoheadrightarrow b$" means "$\langle a, b\rangle \in I^\mathcal{M}(\twoheadrightarrow)$"):

$$\begin{array}{lcl}
\mathcal{M}, g \Vdash_\mathbf{H} p_i & \Leftrightarrow & g(x_1) \in I^\mathcal{M}(P_i^1) \\
\mathcal{M}, g \Vdash_\mathbf{H} x & \Leftrightarrow & g(x_1) = g(x) \\
\mathcal{M}, g \Vdash_\mathbf{H} \neg\phi & \Leftrightarrow & \mathcal{M}, g \nVdash_\mathbf{H} \phi \\
\mathcal{M}, g \Vdash_\mathbf{H} \phi \wedge \psi & \Leftrightarrow & \mathcal{M}, g \Vdash_\mathbf{H} \phi \text{ and } \mathcal{M}, g \Vdash_\mathbf{H} \psi \\
\mathcal{M}, g \Vdash_\mathbf{H} \Box\phi & \Leftrightarrow & \forall a \in D^\mathcal{M}\colon\ g(x_1) \twoheadrightarrow a \Rightarrow \mathcal{M}, g^{x_1}_a \Vdash_\mathbf{H} \phi \\
\mathcal{M}, g \Vdash_\mathbf{H} @_x\phi & \Leftrightarrow & \mathcal{M}, g^{x_1}_{g(x)} \Vdash_\mathbf{H} \phi \\
\mathcal{M}, g \Vdash_\mathbf{H} {\downarrow}x.\phi & \Leftrightarrow & \mathcal{M}, g^{x}_{g(x_1)} \Vdash_\mathbf{H} \phi.
\end{array}$$

Define $\mathbf{H} := \langle \mathcal{L}_\mathbf{H}, \mathsf{C}_\mathbf{H}, \Vdash_\mathbf{H}\rangle$. That is, $\mathbf{H}$ is a formulation of propositional hybrid logic. Now, let $\mathbf{FOL}^{\twoheadrightarrow}$ be the fragment of $\mathbf{FOL}(=)$ where all quantifiers are bounded by $\twoheadrightarrow$ (i.e., of the form $\forall y\ (x \twoheadrightarrow y \to \cdots)$ where $x \neq y$) and where the atomic formulas are restricted to those containing the unary predicates $P_1^1, P_2^1, P_3^1, \ldots$, the identity predicate $=$, or the single binary predicate $\twoheadrightarrow$. Then $\mathbf{H} \approx \mathbf{FOL}^{\twoheadrightarrow}$.

To show that $\mathbf{H} \leqslant \mathbf{FOL}^{\twoheadrightarrow}$, define $\mathtt{ST}_n\colon \mathcal{L}_\mathbf{H} \to \mathcal{L}_\mathtt{Pred}$ for each $x \in \mathtt{Var}$ as follows:

$$\begin{array}{rcl}
\mathtt{ST}_y(p_i) & = & P_i^1(y) \\
\mathtt{ST}_y(x) & = & (y = x) \\
\mathtt{ST}_y(\neg\phi) & = & \neg\,\mathtt{ST}_y(\phi) \\
\mathtt{ST}_y(\phi \wedge \psi) & = & (\mathtt{ST}_y(\phi) \wedge \mathtt{ST}_y(\psi)) \\
\mathtt{ST}_y(\Box\phi) & = & \forall z\ ((y \twoheadrightarrow z) \to \mathtt{ST}_z(\phi))\ \text{ where } z \text{ is not in } \mathtt{ST}_y(\phi) \\
\mathtt{ST}_y(@_x\phi) & = & \mathtt{ST}_x(\phi) \\
\mathtt{ST}_y({\downarrow}x.\phi) & = & \mathtt{ST}_y(\phi)[x/y],
\end{array}$$

where $\alpha[x/y]$ is the result of replacing each free instance of x in α with y. Then for all $\phi \in \mathcal{L}_\mathbf{H}$, $[\![\phi]\!]_\mathbf{H} = [\![\mathtt{ST}_{x_1}(\phi)]\!]_{\mathbf{FOL}^{\twoheadrightarrow}}$.

Conversely, to show that $\mathbf{FOL}^{\rightarrowtriangle} \leqslant \mathbf{H}$, define $\mathtt{HT}\colon \mathcal{L}_{\mathtt{Pred}} \to \mathcal{L}_{\mathbf{H}}$ as follows:

$$
\begin{array}{rcl}
\mathtt{HT}(P_i^1(x)) & = & @_x p_i \\
\mathtt{HT}(x = y) & = & @_x y \\
\mathtt{HT}(x \rightarrowtriangle y) & = & @_x \Diamond y \\
\mathtt{HT}(\neg \phi) & = & \neg \mathtt{HT}(\phi) \\
\mathtt{HT}(\phi \wedge \psi) & = & (\mathtt{HT}(\phi) \wedge \mathtt{HT}(\psi)) \\
\mathtt{HT}(\forall y \ ((x \rightarrowtriangle y) \to \phi)) & = & @_x \Box {\downarrow} y.\mathtt{HT}(\phi).
\end{array}
$$

Then for all $\alpha \in \mathcal{L}^{\rightarrowtriangle}_{\mathtt{Pred}}$, $[\![\alpha]\!]_{\mathbf{FOL}^{\rightarrowtriangle}} = [\![\mathtt{HT}(\alpha)]\!]_{\mathbf{H}}$.

We now briefly turn to observing that $\leqslant$ and $\approx$ satisfy some natural constraints.

Fact 4.1.5 *(Inclusion is a Preorder)*. $\leqslant$ is a preorder on languages.

Corollary 4.1.6 *(Expressive Equivalence is a Congruence Relation)*. $\approx$ is an equivalence relation on languages. Moreover, it is a congruence relation for $\leqslant$, i.e., if $\mathbf{L}_1 \approx \mathbf{L}_1'$ and $\mathbf{L}_2 \approx \mathbf{L}_2'$, then $\mathbf{L}_1 \leqslant \mathbf{L}_2$ iff $\mathbf{L}_1' \leqslant \mathbf{L}_2'$.

By Corollary 4.1.6, the following precisification of expressive power is well-defined:

Definition 4.1.7 *(Expressive Power)*. The ***expressive power of*** $\mathbf{L}$ is the equivalence class $[\mathbf{L}]_{\approx} := \{\mathbf{L}' \mid \mathbf{L} \approx \mathbf{L}'\}$. (We drop mention of $\approx$ throughout.) We define the relation $\leqslant$ on expressive powers so that $[\mathbf{L}_1] \leqslant [\mathbf{L}_2]$ iff $\mathbf{L}_1 \leqslant \mathbf{L}_2$. Likewise for $<$ and $\approx$.

Fact 4.1.8 *(Independence of Representative)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be C-languages. Then $[\mathbf{L}_1] \leqslant [\mathbf{L}_2]$ iff for all $\mathbf{L}_1' \in [\mathbf{L}_1]$ and all $\mathbf{L}_2' \in [\mathbf{L}_2]$, $\mathbf{L}_1' \leqslant \mathbf{L}_2'$. Thus, $\leqslant$ on expressive powers is independent of the representative chosen.

Fact 4.1.9 *(Inclusion on Expressive Powers is a Partial Order)*. The relation $\leqslant$ on expressive powers is a weak partial order.

Fact 4.1.10 *(Fragments are Expressible in Their Extensions)*. For any C-languages, $\mathbf{L}_1$ and $\mathbf{L}_2$, if $\mathbf{L}_1 \subseteq \mathbf{L}_2$, then $\mathbf{L}_1 \leqslant \mathbf{L}_2$.

Hence, any language that is expressively equivalent to a fragment of $\mathbf{L}$ is (by transitivity) expressively included in $\mathbf{L}$. The converse also holds:

Proposition 4.1.11 *(Inclusion is Equivalence to a Fragment)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be C-languages. The following are equivalent:

(a) $\mathbf{L}_1 \leqslant \mathbf{L}_2$.

(b) There is a $\mathbf{L}_2' \subseteq \mathbf{L}_2$ such that $\mathbf{L}_1 \approx \mathbf{L}_2'$.

Proof: The proof from (b) to (a) is given by **Fact 4.1.10**. For (a) to (b), it suffices to show that $\mathbf{L}_1 \approx \mathtt{t}[\mathbf{L}_1]$ where $\mathtt{t}\colon \mathbf{L}_1 \leqslant \mathbf{L}_2$. Clearly, $\mathbf{L}_1 \leqslant \mathtt{t}[\mathbf{L}_1]$. For the converse, observe that while $\mathtt{t}^{-1}$ (the inverse of $\mathtt{t}$) might not necessarily be a function, we can always find a total function on $\mathtt{t}[\mathcal{L}_1]$ contained in $\mathtt{t}^{-1}$. Let $\mathtt{t}^* \subseteq \mathtt{t}^{-1}$ be a total function on $\mathtt{t}[\mathcal{L}_1]$. Since $\mathtt{t}^*$ is a right inverse of $\mathtt{t}$, $[\![\mathtt{t}^*(\mathtt{t}(\phi))]\!]_1 = [\![\mathtt{t}(\mathtt{t}^*(\mathtt{t}(\phi)))]\!]_2 = [\![\mathtt{t}(\phi)]\!]_2$. Hence, we have that $\mathtt{t}[\mathbf{L}_1] \leqslant \mathbf{L}_1$. ■

Fact 4.1.10 does not automatically generalize to expansions of languages. For one thing, $\leqslant$ is only defined for languages which share the same evaluation space. But even setting that aside, there is still a further complication, viz., that we might extend the content of ϕ when we extend the evaluation space to C_2, and that no sentence has a content in $\mathbf{L}_2$ that exactly matches $[\![\phi]\!]_1$. To fix this, it would be natural to consider a more relativized version of expressibility that can allow for fragments of this sort to be comparable.

Definition 4.1.12 *(Relative Expressibility)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let C be a class. We say that $\mathbf{L}_1$ is ***C-expressible in*** $\mathbf{L}_2$ (written "$\mathbf{L}_1 \leqslant_{\mathsf{C}} \mathbf{L}_2$") if for every $\mathcal{L}_1$-formula ϕ, there is an $\mathcal{L}_2$-formula ψ such that $[\![\phi]\!]_1 \cap \mathsf{C} = [\![\psi]\!]_2 \cap \mathsf{C}$. Strict C-inclusion ($<_{\mathsf{C}}$) and C-equivalence ($\approx_{\mathsf{C}}$) are similarly defined.

Thus, $\leqslant$ is the special case of $\leqslant_{\mathsf{C}}$ when $\mathsf{C} = \mathsf{C}_1 = \mathsf{C}_2$. If $\mathsf{C} = \varnothing$, then every language is trivially expressively equivalent to every other language. As another special case, we have the following:

Fact 4.1.13 *(Restrictions are Relatively Expressible in their Expansions)*. For any languages $\mathbf{L}_1$ and $\mathbf{L}_2$, if $\mathbf{L}_1 \Subset \mathbf{L}_2$, then $\mathbf{L}_1 \leqslant_{\mathsf{C}_1} \mathbf{L}_2$

We mostly focus on the unrelativized notions of expressibility below.

§ 4.1.2 Weak Expressibility

It is common (e.g., in the literature on the expressive power of modal languages) to adopt a more generous notion of expressibility that is strictly weaker than the notion of expressibility adopted here. In particular, the notion of expressibility from **Definition 4.1.2** assumes that for each formula in the source language, we need to find a *single* formula with the same semantic value in the target language for the source language to be expressible in the target language. One might instead be willing to allow that a formula from the source language is "expressible" in the target language if there is some *class* of formulas in the latter with the same semantic value.

Definition 4.1.14 *(Weak Expressibility)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be C-languages. We say that $\mathbf{L}_1$ is ***weakly expressible in*** $\mathbf{L}_2$ (written "$\mathbf{L}_1 \preceq \mathbf{L}_2$") if for every class of $\mathcal{L}_1$-formulas Γ, there is a class of $\mathcal{L}_2$-formulas Δ such that $[\![\Gamma]\!]_1 = [\![\Delta]\!]_2$. We use "$\prec$" for strict weak expressibility, and "$\sim$" for weak expressive equivalence. We define the relativized versions of weak expressibility analogously to **Definition 4.1.12**.

Lemma 4.1.15 *(Equivalent Definition of $\preceq_L$)*. For any languages $\mathbf{L}_1$ and $\mathbf{L}_2$ and any class C, we have that $\mathbf{L}_1 \preceq_{\mathsf{C}} \mathbf{L}_2$ iff for every $\mathcal{L}_1$-formula ϕ, there is a class of $\mathcal{L}_2$-formulas Δ such that $[\![\phi]\!]_1 \cap \mathsf{C} = [\![\Delta]\!]_2 \cap \mathsf{C}$.

Proof: Left-to-right is obvious, since we can consider singletons of the form $\{\phi\}$. For the right-to-left direction, let Γ be a set of $\mathcal{L}_1$-formulas. We know by hypothesis that for each $\psi \in \Gamma$, there is a class of $\mathcal{L}_2$-formulas Δ_ψ such that $[\![\psi]\!]_1 \cap \mathsf{C} = [\![\Delta_\psi]\!]_2 \cap \mathsf{C}$. Let $\Delta := \bigcup_{\psi \in \Gamma} \Delta_\psi$. Then $[\![\Gamma]\!]_1 \cap \mathsf{C} = [\![\Delta]\!]_2 \cap \mathsf{C}$. ■

Corollary 4.1.16 *(Expressibility Implies Weak Expressibility)*. For any C, and any languages $\mathbf{L}_1$ and $\mathbf{L}_2$, if $\mathbf{L}_1 \leqslant_{\mathsf{C}} \mathbf{L}_2$, then $\mathbf{L}_1 \preceq_{\mathsf{C}} \mathbf{L}_2$.

Proof: Immediate from **Lemma 4.1.15**. ■

It is straightforward to check the facts above regarding $\leqslant$ all apply to $\preceq$. However, the two notions of expressibility are not equivalent: weak expressibility is strictly weaker than expressibility.

Example 4.1.17 *(Weak Expressibility Does Not Imply Expressibility)*. Consider two simple languages $\mathbf{L}_1$ and $\mathbf{L}_2$, where:

- $\mathcal{L}_1 = \{p_1, p_2, p_3, p_4\}$
- $\mathcal{L}_2 = \{q, \neg q, r, \neg r\}$
- $\mathsf{C}_1 = \mathsf{C}_2 = \{x_1, x_2, x_3, x_4\}$
- $[\![p_i]\!]_1 = \{x_i\}$

- $[\![q]\!]_2 = \{x_1, x_2\}$, $[\![r]\!]_2 = \{x_1, x_3\}$, and $[\![\neg\phi]\!]_2 = \mathsf{C}_2 - [\![\phi]\!]_2$.

Now, $\mathbf{L}_1 \preceq \mathbf{L}_2$; for instance, $[\![p_2]\!]_1 = \{x_2\} = [\![q]\!]_2 \cap [\![\neg r]\!]_2 = [\![\{q, \neg r\}]\!]_2$. But $\mathbf{L}_1 \nleqslant \mathbf{L}_2$; in fact, none of the formulas in $\mathcal{L}_1$ pick out the same set of points as any of the formulas in $\mathcal{L}_2$.

Example 4.1.18 *(Restricted Infinite Quantification)*. Recall that $\mathcal{L}_{\mathtt{Pred}}$ is the syntax of standard first-order logic. Let $\mathcal{L}_{\mathtt{Pred}}(=)$ be the syntax of first-order logic with identity. Define the new syntax $\mathcal{L}^{\infty}_{\mathtt{Pred}}(=)$ as follows:

$$\mathcal{L}^{\infty}_{\mathtt{Pred}} = \mathcal{L}_{\mathtt{Pred}}(=) \cup \{\exists_{\infty} x_i\, \phi \mid \phi \in \mathcal{L}_{\mathtt{Pred}}(=) \;\&\; i \in \mathbb{N}\}\,.$$

Where **FOL** is the usual language of first-order logic, define **FOL**(=) as the usual language of first-order logic with identity and $\mathbf{FOL}^{\infty}(=)$ as the extension of **FOL**(=) with the following semantic clause:

$$\mathcal{M}, g \Vdash \exists_{\infty} x\, \phi \quad\Leftrightarrow\quad \left|\{a \in D^{\mathcal{M}} \mid \mathcal{M}, g^x_a \Vdash \phi\}\right| \geqslant \aleph_0.$$

Then $\mathbf{FOL}^{\infty}(=) \preceq \mathbf{FOL}(=)$. In particular, let $\exists_{\geqslant n} x\, \phi$ abbreviate the $\mathcal{L}_{\mathtt{Pred}}(=)$-formula that says "there are at least n ϕs". Then for any $\mathcal{L}^{\infty}_{\mathtt{Pred}}(=)$-formula of the form $\exists_{\infty} x\, \phi$, the set $\{\exists_{\geqslant n} x\, \phi \mid n \geqslant 1\}$ picks out the same class of models. But $\mathbf{FOL}^{\infty}(=) \nleqslant \mathbf{FOL}(=)$, since $\exists_{\infty} x\,(x = x)$ is already not expressible in **FOL**(=) by compactness.

These examples, while simple, are somewhat gerrymandered. I have not yet come up with a more natural example of some languages where one is only weakly expressible in the other.

Of course, in languages with infinitary conjunction, the differences between these two notions of expressive power collapse, since you can simply take your translating formula to be the conjunction of the translating class. And in fact, the converse is true as well: weak expressibility implies expressibility only if the target language has infinitary conjunction.

Proposition 4.1.19 *(Collapsing Expressibility to Weak Expressibility)*. Let $\mathbf{L}_2$ be a C-language. Then the following are equivalent:

(a) For any C-language $\mathbf{L}_1$, if $\mathbf{L}_1 \preceq \mathbf{L}_2$, then $\mathbf{L}_1 \leqslant \mathbf{L}_2$.

(b) $\mathbf{L}_2$ has infinite conjunction.

Proof: Showing (b) implies (a) is easy. To show that (a) implies (b), let $\Gamma \subseteq \mathcal{L}_2$. Define the language $\mathbf{L}_1 = \langle \mathcal{L}_2, \mathsf{C}, \Vdash_1 \rangle$ such that for all $\phi \in \mathcal{L}_2$ and $x \in \mathsf{C}$:

$$x \Vdash_1 \phi \quad \Leftrightarrow \quad x \Vdash_2 \Gamma \text{ and } x \Vdash_2 \phi.$$

Then if $\phi \in \mathcal{L}_2$, $[\![\phi]\!]_1 = [\![\Gamma \cup \{\phi\}]\!]_2$. Hence, $\mathbf{L}_1 \preceq \mathbf{L}_2$, and thus $\mathbf{L}_1 \leqslant \mathbf{L}_2$. Now, let $\psi \in \Gamma$. Then $[\![\psi]\!]_1 = [\![\Gamma]\!]_2$, so by expressibility, there is a $\bigwedge \Gamma \in \mathcal{L}_2$ such that $[\![\psi]\!]_2 = [\![\bigwedge \Gamma]\!]_2 = [\![\Gamma]\!]_2$. ■

The proof of **Proposition 4.1.19** shows how to generate a large number of counterexamples to the collapse of weak and strong expressibility when infinite conjunctions are absent. But this construction does not give us, for any $\mathbf{L}_2$ without infinite conjunction, a relatively natural language $\mathbf{L}_1$ such that $\mathbf{L}_1 \preceq \mathbf{L}_2$ but $\mathbf{L}_1 \not\leqslant \mathbf{L}_2$.

As far as I am aware, no one has seriously defended the claim that expressive power should be measured by weak expressibility rather than expressibility. For natural languages, it would be odd to use weak expressibility as a form of expressive power since one literally could not utter an infinite number of sentences. But for formal languages, whose purpose is more theoretical in nature, this consideration does not apply. In practice, most expressibility results which show that $\mathbf{L}_1 \preceq \mathbf{L}_2$ already show that $\mathbf{L}_1 \leqslant \mathbf{L}_2$; similarly, most inexpressibility results that show $\mathbf{L}_1 \not\leqslant \mathbf{L}_2$ already show $\mathbf{L}_1 \not\preceq \mathbf{L}_2$. So we will not concern ourselves here with the question of which is more fitting as a notion of genuine expressive power.

§ 4.1.3 Discernibility

The most common technique used for showing that one language is lacking in expressive power compared to another is to construct two models which are equivalent relative to the former but disagree one some sentence in the latter—that is, if the former cannot discern the difference between two possibilities discernible by the latter. For example, to show that the standard propositional modal language cannot express some formula in the correspondence language, one first constructs two modally equivalent (usually, bisimilar) models, and then demonstrates that these models disagree on the formula from the correspondence language. This suggests an alternative form of expressive power: two languages are expressively equivalent if the ways they carve logical space are equally fine-grained.

Definition 4.1.20 *(Discernibility)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be C-languages. We say $\mathbf{L}_1$ is ***discernible in*** $\mathbf{L}_2$ (written "$\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$") if for all $x, y \in \mathsf{C}$, if $\mathtt{Diag}_2(x) = \mathtt{Diag}_2(y)$, then $\mathtt{Diag}_1(x) = \mathtt{Diag}_1(y)$—alternatively, using the "$\equiv$" notation from **Definition 1.1.25**, if $x \equiv_2 y$, then $x \equiv_1 y$. We say $\mathbf{L}_1$ is ***strictly discernible in*** $\mathbf{L}_2$ (written "$\mathbf{L}_1 \sqsubset \mathbf{L}_2$") if $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ but $\mathbf{L}_2 \not\sqsubseteq \mathbf{L}_1$. We say that $\mathbf{L}_1$ is ***discernibly equivalent to*** $\mathbf{L}_2$ (written "$\mathbf{L}_1 \sqsupseteq\sqsubseteq \mathbf{L}_2$") if $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ and $\mathbf{L}_2 \sqsubseteq \mathbf{L}_1$.

Fact 4.1.21 *(Discernibility is a Preorder).* $\sqsubseteq$ is a preorder.

Proposition 4.1.22 *(Weak Expressibility Implies Discernibility).* Suppose $\mathbf{L}_1$ and $\mathbf{L}_2$ are both C-languages. If $\mathbf{L}_1 \preceq \mathbf{L}_2$, then $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$.

Proof: Suppose that $\mathtt{Diag}_2(x) = \mathtt{Diag}_2(y)$ but $\mathtt{Diag}_1(x) \neq \mathtt{Diag}_1(y)$ for some $x, y \in \mathsf{C}_{\mathbf{L}}$. Without loss of generality, assume there is a $\phi \in \mathcal{L}_1$ such that $x \Vdash_1 \phi$ and $y \nVdash_1 \phi$.[a] Let $\Delta \subseteq \mathcal{L}_2$. Since $\mathtt{Diag}_2(x) = \mathtt{Diag}_2(y)$, we have $x \Vdash_2 \Delta$ iff $y \Vdash_2 \Delta$. If $x \Vdash_2 \Delta$ and $y \Vdash_2 \Delta$, then $[\![\Delta]\!]_2 \nsubseteq [\![\phi]\!]_1$ since y is a counterexample. If $x \nVdash_2 \Delta$ and $y \nVdash_2 \Delta$, then $[\![\phi]\!]_1 \nsubseteq [\![\Delta]\!]_2$ since x is a counterexample. So either way, $[\![\phi]\!]_1 \neq [\![\Delta]\!]_2$. But then $\mathbf{L}_1 \npreceq \mathbf{L}_2$. ■

[a] It might be that $\mathtt{Diag}_1(x) = \varnothing$, in which case there is no such ϕ; but then there is a $\phi \in \mathcal{L}_1$ such that $x \nVdash_1 \phi$ and $y \Vdash_1 \phi$. The proof in this case is symmetric.

Corollary 4.1.23 *(Refuting Expressibility).* Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be C-languages. If there are some $x, y \in \mathsf{C}$ such that $x \equiv_2 y$ but $x \Vdash_1 \phi$ while $y \nVdash_1 \phi$ for some $\phi \in \mathcal{L}_1$, then $\mathbf{L}_1 \npreceq \mathbf{L}_2$.

In words, to show that $\mathbf{L}_1 \npreceq \mathbf{L}_2$, it suffices to find two points of evaluation that agree on $\mathbf{L}_2$ but disagree on $\mathbf{L}_1$. We briefly present an example of Corollary 4.1.23 in action.

Example 4.1.24 *(Universal Modality).* Let $\mathbf{K}$ be the minimal normal modal logic (as before) and let $\mathbf{K}^+$ be the extension of $\mathbf{K}$ with the unary operator ■ such that for all pointed Kripke models $\langle \mathcal{M}, w \rangle$ and all ϕ:

$$\mathcal{M}, w \Vdash^+ \blacksquare\phi \quad \Leftrightarrow \quad \forall v \in W^{\mathcal{M}}\colon\ \mathcal{M}, v \Vdash^+ \phi.$$

Obviously, $\mathbf{K} \preccurlyeq \mathbf{K}^+$. But $\mathbf{K}^+ \npreccurlyeq \mathbf{K}$, since $\mathbf{K}^+ \not\sqsubseteq \mathbf{K}$. For let $\mathcal{M} = \langle \{w, v\}, \varnothing, V \rangle$ be such that $V(p) = \{w\}$ and let $\mathcal{N} = \langle \{w\}, \varnothing, V \rangle$. Then it can be shown that $\mathcal{M}, w$ and $\mathcal{N}, w$ are bisimilar (and hence that $\mathcal{M}, w \equiv_{\mathbf{K}} \mathcal{N}, w$), even though $\mathcal{M}, w \nVdash^+ \blacksquare p$ while $\mathcal{N}, w \Vdash^+ \blacksquare p$.

It should also be noted that showing this is not *necessary* for showing failures of expressibility. That is, $\mathbf{L}_1$ might be discernible in $\mathbf{L}_2$ without being even weakly expressible in $\mathbf{L}_2$.

Example 4.1.25 *(Discernibility Does Not Imply Weak Expressibility)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be defined as follows:

- $\mathcal{L}_1 = \mathcal{L}_2 = \{p\}$
- $\mathsf{C}_1 = \mathsf{C}_2 = \{x\}$
- $x \nVdash_1 p$ but $x \Vdash_2 p$.

Then vacuously, $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ (in fact, $\mathbf{L}_1 \sqsupseteq\sqsubseteq \mathbf{L}_2$), but $\mathbf{L}_1 \not\leqslant \mathbf{L}_2$.

Discernibility is a rather weak notion of expressive power. To say that two languages are discernibly equivalent is just to say that each fractures logical space with equal granularity. But it says nothing about *how* the two languages fracture it. It could be that one language has a sentence for each class of equivalent points whereas another only achieves fracturing after applying many different cuts through logical space. Though discernibility is a way of measuring a language's ability to distinguish possibilities, it does not guarantee that there is a way of "preserving meaning" of formulas from one language into another.

One way to see this is that, unlike (weak) expressibility, discernibility does not imply discernible equivalence to a fragment:

Example 4.1.26 *(Discernibility Does Not Imply Discernible Equivalence to a Fragment)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be defined as follows:

- $\mathcal{L}_1 = \{p\}$, $\mathcal{L}_2 = \{q, r\}$
- $\mathsf{C}_1 = \mathsf{C}_2 = \{x_1, x_2, x_3, x_4\}$
- $\mathtt{Diag}_1(x_1) = \mathtt{Diag}_1(x_2) = \{p\}$, $\mathtt{Diag}_1(x_3) = \mathtt{Diag}_1(x_4) = \varnothing$
- $\mathtt{Diag}_2(x_1) = \{q\}$, $\mathtt{Diag}_2(x_2) = \{r\}$, $\mathtt{Diag}_2(x_3) = \{q, r\}$, $\mathtt{Diag}_2(x_4) = \varnothing$

Then $\mathbf{L}_1 \sqsubset \mathbf{L}_2$, even though there is no $\mathbf{L}_2' \subseteq \mathbf{L}_2$ such that $\mathbf{L}_1 \sqsupseteq\sqsubseteq \mathbf{L}_2'$.

The above example makes use of the fact that a language can discern a region of logical space not just by what formulas the points in that region satisfy but also by what formulas they do not satisfy. In languages with truth-functional negation, however, there is no difference between the two. Indeed, as one would suspect, this is essential. For we saw in **Proposition 4.1.11** that expressibility is expressive equivalence to a fragment. And indeed, there are special circumstances making use of truth-functional negation where $\preceq$ or $\leqslant$ collapse to $\sqsubseteq$, akin to **Proposition 4.1.19**.

Proposition 4.1.27 *(Collapsing Expressibility to Discernibility)*. Let $\mathbf{L}_2$ be a C-language. Then the following are equivalent:

(a) For any C-language $\mathbf{L}_1$, if $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, then $\mathbf{L}_1 \leqslant \mathbf{L}_2$.

(b) $\mathbf{L}_2$ has infinite conjunction and truth-functional negation.

Proof:

(a) ⇒ (b). Suppose $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ implies $\mathbf{L}_1 \leqslant \mathbf{L}_2$ for all C-languages $\mathbf{L}_1$. By **Corollary 4.1.16** and **Proposition 4.1.19**, $\mathbf{L}_2$ has infinite conjunction. Now, let $\phi \in \mathcal{L}_2$. Define the language $\mathbf{L}_1 = \langle \{p\}, \mathsf{C}, \Vdash_1 \rangle$ such that:

$$x \Vdash_1 p \quad \Leftrightarrow \quad x \nVdash_2 \phi.$$

If $x, y \in \mathsf{C}$ are such that $\mathtt{Diag}_1(x) \neq \mathtt{Diag}_1(y)$, then either $p \in \mathtt{Diag}_1(x)$ and $p \notin \mathtt{Diag}_1(y)$ or *vice versa*. Without loss of generality, suppose it is the former. Then $x \nVdash_2 \phi$ and $y \Vdash_2 \phi$, so $\mathtt{Diag}_2(x) \neq \mathtt{Diag}_2(y)$. Hence, $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, and so $\mathbf{L}_1 \leqslant \mathbf{L}_2$. So there is a $\neg\phi \in \mathcal{L}_2$ such that $[\![p]\!]_1 = [\![\neg\phi]\!]_2 = \mathsf{C} - [\![\phi]\!]_2$.

(b) ⇒ (a). Suppose $\mathbf{L}_2$ has infinite conjunction and negation and suppose $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$. Let $\phi \in \mathcal{L}_1$. Since infinite disjunction can be defined in terms of infinite conjunction and negation, we can define the formula $\phi^* := \phi_1^* \wedge \phi_2^*$, where:

$$\phi_1^* := \bigvee \left\{ \bigwedge \mathtt{Diag}_2(x) \;\middle|\; x \in \mathsf{C} \text{ and } x \Vdash_1 \phi \right\}$$
$$\phi_2^* := \bigwedge \left\{ \neg \bigwedge \mathtt{Diag}_2(y) \;\middle|\; y \in \mathsf{C} \text{ and } y \nVdash_1 \phi \right\}.$$

We now show that $[\![\phi]\!]_1 = [\![\phi^*]\!]_2$.

First, suppose $x \notin [\![\phi]\!]_1$. Then $x \nVdash_2 \phi_2^*$, so trivially $x \nVdash_2 \phi^*$. Hence, $x \notin [\![\phi^*]\!]_2$. Next, suppose $x \in [\![\phi]\!]_1$. Clearly, $x \Vdash_2 \phi_1^*$. Suppose for *reductio* that $x \nVdash_2 \phi_2^*$. That means that for some $y \in \mathsf{C}$ such that $y \nVdash_1 \phi$, $x \Vdash_2 \mathtt{Diag}_2(y)$. Since $x \Vdash_1 \phi$ and $y \nVdash_1 \phi$, $\mathtt{Diag}_1(x) \neq \mathtt{Diag}_1(y)$. So $\mathtt{Diag}_2(x) \neq \mathtt{Diag}_2(y)$ by discernibility. Hence, $\mathtt{Diag}_2(y) \subset \mathtt{Diag}_2(x)$. But by **Fact 1.3.24**, any language with negation must be opinionated, $\lightning$. Thus, $x \Vdash_2 \phi_2^*$, and so $x \in [\![\phi^*]\!]_2$. Therefore, $[\![\phi]\!]_1 = [\![\phi^*]\!]_2$. ∎

In light of this result, it would be natural to ask whether one can characterize the conditions under which weak expressibility and discernibility collapse. The answer is affirmative, though the conditions are somewhat less natural to state. Let us say that $\mathbf{L}$ has ***setwise negation*** if for any $\Gamma \subseteq \mathcal{L}$, there is a $\neg\Gamma \subseteq \mathcal{L}$ such that $[\![\neg\Gamma]\!] = \mathsf{C} - [\![\Gamma]\!]$. Let us also say that $\mathbf{L}$ has ***infinite setwise disjunction*** if for all $\Sigma \subseteq \wp(\mathcal{L})$, there is a $\bigvee\Sigma \subseteq \mathcal{L}$ such that $[\![\bigvee\Sigma]\!] = \bigcup_{\Gamma\in\Sigma}[\![\Gamma]\!]$.

Proposition 4.1.28 *(Collapsing Weak Expressibility to Discernibility)*. Let $\mathbf{L}_2$ be a C-language. Then the following are equivalent:

(a) For any C-language $\mathbf{L}_1$, if $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, then $\mathbf{L}_1 \preceq \mathbf{L}_2$.

(b) $\mathbf{L}_2$ has infinite setwise disjunction and setwise negation.

Proof:

(a) ⇒ (b). Where $\Sigma \subseteq \wp(\mathcal{L})$ and $\Delta \subseteq \mathcal{L}$, define $\mathbf{L}_1^{\vee} = \langle\{p\}, \mathsf{C}, \Vdash_1^{\vee}\rangle$ and $\mathbf{L}_1^{\neg} = \langle\{q\}, \mathsf{C}, \Vdash_1^{\neg}\rangle$ so that:

$$\begin{aligned} x \Vdash_1^{\vee} p \quad &\Leftrightarrow \quad \exists\Gamma \in \Sigma\colon\ x \Vdash_2 \Gamma \\ x \Vdash_1^{\neg} q \quad &\Leftrightarrow \quad x \nVdash_2 \Delta. \end{aligned}$$

Then the proof is as in **Proposition 4.1.27**.

(b) ⇒ (a). Define $\Phi^* := \Phi_1^* \wedge \Phi_2^*$, where:

$$\Phi_1^* := \bigvee\{\mathtt{Diag}_2(x) \mid x \in \mathsf{C} \text{ and } x \Vdash_1 \phi\}$$

$$\Phi_2^* := \neg\bigvee\{\mathtt{Diag}_2(y) \mid y \in \mathsf{C} \text{ and } y \nVdash_1 \phi\}.$$

Then the proof is as in **Proposition 4.1.27**.[a] ∎

[a]Note that a language is opinionated if it has even setwise negation.

These results say that a language $\mathbf{L}$ (weakly) expressively *includes* every language discernible in it just in case some truth-functional connectives are present. One might wonder whether there is a similar result stating the conditions under which a language $\mathbf{L}$ is (weakly) expressively *included* in every language that it is discernible in. The answer is either negative (in the case of expressibility) or as good as negative (in the case of weak expressibility).

First, some qualifications are needed to make the question interesting. For one thing, observe that if $\mathcal{L}_1 = \varnothing$, then vacuously $\mathbf{L}_1 \leqslant \mathbf{L}_2$ for any C-language $\mathbf{L}_2$. Thus, the question is only interesting if we restrict to nonempty languages. Moreover, if

$\mathcal{L}_1 \neq \varnothing$ but $\mathsf{C} = \varnothing$, then vacuously $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, though $\mathbf{L}_1 \leqslant \mathbf{L}_2$ iff $\mathcal{L}_2 \neq \varnothing$. So the answer to our question is trivial if $\mathsf{C} = \varnothing$. Finally, if $\mathcal{L}_2 = \varnothing$, then vacuously $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ iff $\mathtt{Diag}_1(x) = \mathtt{Diag}_1(y)$ for all $x, y \in \mathsf{C}$, while $\mathbf{L}_1 \leqslant \mathbf{L}_2$ iff $\mathcal{L}_1 = \varnothing$. Hence, the interesting question is when a nonempty language over a nonempty evaluation space is expressible in every nonempty language that it is discernible in. The answer to that question is never:

Proposition 4.1.29 *(No Non-Trivial Language is Expressible in Every Non-Trivial Language it is Discernible In)*. Let $\mathbf{L}_1$ be a C-language where $\mathcal{L}_1 \neq \varnothing$ and $\mathsf{C} \neq \varnothing$. Then there is a C-language $\mathbf{L}_2$ where $\mathcal{L}_2 \neq \varnothing$ such that $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ but $\mathbf{L}_1 \nleqslant \mathbf{L}_2$.

Proof: Suppose for all C-languages $\mathbf{L}_2$, if $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, then $\mathbf{L}_1 \leqslant \mathbf{L}_1^*$. Define $\mathbf{L}_1^* := \langle \mathsf{C}, \mathsf{C}, = \rangle$. Cleary, $\mathbf{L}_1 \sqsubseteq \mathbf{L}_1^*$, since $\mathtt{Diag}_1^*(x) = \mathtt{Diag}_1^*(y)$ iff $x = y$. Hence, for all $\phi \in \mathcal{L}_1$, there is an $x \in \mathsf{C}$ such that $[\![\phi]\!]_1 = [\![x]\!]_1^* = \{x\}$. That is, $|[\![\phi]\!]_1| = 1$ for all $\phi \in \mathcal{L}_1$ (i.e., $\mathbf{L}_1$ is already maximally fine-grained).

We now distinguish three cases, showing that in either case there is a C-language $\mathbf{L}_2$ (with $\mathcal{L}_2 \neq \varnothing$) such that $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ and $\mathbf{L}_1 \nleqslant \mathbf{L}_2$, $\lightning$.

Case 1: $|\mathsf{C}| = 1$**.** Then trivially $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ for any C-language $\mathbf{L}_2$. Define $\mathbf{L}_2 := \langle \{p\}, \mathsf{C}, \varnothing \rangle$. Then $\mathbf{L}_1 \nleqslant \mathbf{L}_2$, since for all $\phi \in \mathcal{L}_1$, $[\![\phi]\!]_1 = \mathsf{C}$, whereas $[\![p]\!]_2 = \varnothing \neq \mathsf{C}$.

Case 2: $|\mathsf{C}| = 2$**.** Let $\mathsf{C} = \{x, y\}$. Since $|[\![\phi]\!]_1| = 1$ for all $\phi \in \mathcal{L}$, it cannot be that $\mathtt{Diag}_1(x) = \mathtt{Diag}_1(y) = \mathcal{L}_1$. Without loss of generality, suppose $\mathtt{Diag}_1(x) \neq \mathcal{L}_1$. If $\mathtt{Diag}_1(x) = \varnothing$ (and so $\mathtt{Diag}_1(y) = \mathcal{L}_1$), then define:

$$\mathbf{L}_2 := \langle \{p\}, \mathsf{C}, \{\langle x, p \rangle\} \rangle .$$

Then $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, since $\mathtt{Diag}_2(x) \neq \mathtt{Diag}_2(y)$. But $\mathbf{L}_1 \nleqslant \mathbf{L}_2$, since $[\![\phi]\!]_1 = \{y\}$ for all $\phi \in \mathcal{L}_1$ while $[\![p]\!]_2 = \{x\}$.

If instead $\mathtt{Diag}_1(x) \neq \varnothing$ (and so $\mathtt{Diag}_1(y) \neq \mathcal{L}_1$[a]), define:

$$\mathbf{L}_2 := \langle \mathtt{Diag}_1(x), \mathsf{C}, \Vdash_1 \restriction_{\mathtt{Diag}_1(x)} \rangle .$$

Then once more $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$. But $\mathbf{L}_1 \nleqslant \mathbf{L}_2$, since if $\phi \in \mathtt{Diag}_1(x)$, then $[\![\phi]\!]_2 \neq \{y\}$, whereas there is a $\psi \in \mathcal{L}_1$ such that $[\![\psi]\!]_1 = \{y\}$.

Case 3: $|\mathsf{C}| \geqslant 3$**.** Define:

$$\mathbf{L}_2 := \langle \{\{x, y\} \mid x, y \in \mathsf{C} \text{ and } x \neq y\}, \mathsf{C}, \in \rangle .$$

Then $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ (since $\mathtt{Diag}_2(x) = \mathtt{Diag}_2(y)$ iff $x = y$). But $\mathbf{L}_1 \nleqslant \mathbf{L}_2$, since $|[\![\phi]\!]_2| = 2$ for all $\phi \in \mathcal{L}_2$. ∎

[a] Moreover, $\mathtt{Diag}_1(y) \neq \varnothing$, since $\mathtt{Diag}_1(x) \neq \mathcal{L}_1$.

Thus, there is no interesting collapse theorem of the form "**L** is expressible in any language it is discernible in iff…". The languages that can be expressed in a language **L** are much more confined than the possible languages that can express **L**.

Turning to weak expressibility, the situation is a little more complicated. Again, the question is only interesting if we consider nonempty languages over a nonempty evaluation space.[3] So: when is a nonempty language over a nonempty evaluation space weakly expressible in every nonempty language it is discernible in? The answer is exactly when every formula in the language is valid:

Proposition 4.1.30 *(Almost No Non-Trivial Language is Weakly Expressible in Every Non-Trivial Language it is Discernible In)*. Let $\mathbf{L}_1$ be a C-language where $\mathcal{L}_1 \neq \varnothing$ and $\mathsf{C} \neq \varnothing$. Then the following are equivalent:

(a) For all C-languages $\mathbf{L}_2$ such that $\mathcal{L}_1 \neq \varnothing$, if $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, then $\mathbf{L}_1 \preceq \mathbf{L}_2$.

(b) For all $\phi \in \mathcal{L}_1$, $[\![\phi]\!]_1 = \mathsf{C}$.

Proof: The implication from (b) to (a) is trivial, since $[\![\varnothing]\!]_2 = \mathsf{C}$ for any C-language $\mathbf{L}_2$. For the implication from (a) to (b), we first prove two lemmas:

(1) For all $\phi \in \mathcal{L}_1$, either:

  (i) $[\![\phi]\!]_1 = \varnothing$

  (ii) $[\![\phi]\!]_1 = \{x\}$ for some $x \in \mathsf{C}$

  (iii) $[\![\phi]\!]_1 = \mathsf{C}$.

(2) There is a $\phi \in \mathcal{L}_1$ such that $[\![\phi]\!]_1 \neq \varnothing$.

For the proof of (1), define $\mathbf{L}_1^* := \langle \mathsf{C}, \mathsf{C}, = \rangle$. As before, $\mathbf{L}_1 \sqsubseteq \mathbf{L}_1^*$, so $\mathbf{L}_1 \preceq \mathbf{L}_1^*$. Let $\phi \in \mathcal{L}_1$. Then there is a $\Gamma \subseteq \mathsf{C}$ where $[\![\phi]\!]_1 = [\![\Gamma]\!]_1^*$. There are three cases:

(i) $|\Gamma| > 1$. Then $[\![\Gamma]\!]_1^* = \varnothing$.

(ii) $|\Gamma| = 1$. Then $[\![\Gamma]\!]_1^* = \{x\}$ for some $x \in \mathsf{C}$.

(iii) $|\Gamma| = 0$. Then $[\![\Gamma]\!]_1^* = \mathsf{C}$.

This establishes (1).

For (2), suppose for *reductio* that $[\![\phi]\!]_1 = \varnothing$ for all $\phi \in \mathcal{L}_1$. Define $\mathbf{L}_2 := \langle \{p\}, \mathsf{C}, \mathsf{C} \times \{p\} \rangle$. Then $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, since $\mathtt{Diag}_1(x) = \mathtt{Diag}_1(y) = \varnothing$ for all $x, y \in \mathsf{C}$. But $\mathbf{L}_1 \not\preceq \mathbf{L}_2$, since $[\![p]\!]_2 = \mathsf{C}$ and $\mathsf{C} \neq \varnothing$, $\lightning$. This establishes (2).

Now, we distinguish two cases, showing that either way, (b) holds.

[3]We already saw above that if $\mathcal{L}_1 = \varnothing$, then vacuously $\mathbf{L}_1 \leqslant \mathbf{L}_2$. If $\mathcal{L}_1 \neq \varnothing$ but $\mathsf{C} = \varnothing$, then trivially $\mathbf{L}_1 \preceq \mathbf{L}_2$, since $[\![\varnothing]\!]_2 = \mathsf{C} = \varnothing$. And if $\mathcal{L}_2 = \varnothing$, $\mathbf{L}_1 \preceq \mathbf{L}_2$ iff $\mathtt{Diag}_1(x) = \mathtt{Diag}_1(y)$ for all $x, y \in \mathsf{C}$, so $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ iff $\mathbf{L}_1 \preceq \mathbf{L}_2$.

Case 1: $|\mathsf{C}| = 1$**.** Suppose for *reductio* that (b) fails. Thus, for some $\phi \in \mathcal{L}_1$, $[\![\phi]\!]_1 \neq \mathsf{C}$. Since $|\mathsf{C}| = 1$, that means $[\![\phi]\!]_1 = \varnothing$. Now, observe that $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ trivially for any C-language $\mathbf{L}_2$. So define:

$$\mathcal{L}_2 := \mathcal{L}_1 - \{\psi \mid [\![\psi]\!]_1 = \varnothing\}$$
$$\mathbf{L}_2 := \langle \mathcal{L}_2, \mathsf{C}, \Vdash_1 \restriction_{\mathcal{L}_2} \rangle .$$

By (2), $\mathcal{L}_2 \neq \varnothing$. But $\mathbf{L}_1 \not\leq \mathbf{L}_2$, since $[\![\Gamma]\!]_2 = \mathsf{C}$ for all $\Gamma \subseteq \mathcal{L}_2$, whereas $[\![\phi]\!]_1 = \varnothing$, $\lightning$.

Case 2: $|\mathsf{C}| > 1$**.** Suppose for *reductio* that (b) fails. Notice first that if every $\phi \in \mathcal{L}_1$ is such that either $[\![\phi]\!]_1 = \varnothing$ or $[\![\phi]\!]_1 = \mathsf{C}$, then defining $\mathbf{L}_2$ as in Case 1, $\mathbf{L}_1 \not\leq \mathbf{L}_2$, though $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, $\lightning$. Hence, there is a $\phi \in \mathcal{L}_1$ such that $[\![\phi]\!]_1 \neq \varnothing$ and $[\![\phi]\!]_1 \neq \mathsf{C}$, which by (1) implies that $[\![\phi]\!]_1 = \{x\}$ for some $x \in \mathsf{C}$. Fixing on such an x, define:

$$\mathbf{L}_1^{**} := \langle \mathsf{C} - \{x\}, \mathsf{C}, = \rangle .$$

Then vacuously $\mathbf{L}_1 \sqsubseteq \mathbf{L}_1^{**}$. But $\mathbf{L}_1 \not\leq \mathbf{L}_1^{**}$, since for no $\Gamma \subseteq \mathsf{C} - \{x\}$ is $[\![\Gamma]\!]_1^{**} = \{x\}$ (either $[\![\Gamma]\!]_1^{**} = \varnothing$, or $[\![\Gamma]\!]_1^{**} = \{y\}$ for some $y \neq x$, or $\Gamma = \varnothing$, in which case $[\![\Gamma]\!]_1^{**} = \mathsf{C}$), $\lightning$. ■

§ 4.1.4 Strong Discernibility

Just as we can generalize expressibility from being defined in terms of formulas to sets of formulas, so too we can generalize discernibility from being defined in terms of points of evaluations to sets of points.

Notation: Let $\mathbf{L}$ be a C-language. We define the diagram of a class of points $\mathsf{X} \subseteq \mathsf{C}$ so that $\mathtt{Diag}_{\mathbf{L}}(\mathsf{X}) := \bigcap_{x \in \mathsf{X}} \mathtt{Diag}_{\mathbf{L}}(x)$.

Definition 4.1.31 *(Strong Discernibility)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be C-languages. We say $\mathbf{L}_1$ is ***strongly discernible in*** $\mathbf{L}_2$ (written "$\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$") if for all $\mathsf{X}, \mathsf{Y} \subseteq \mathsf{C}$, if $\mathtt{Diag}_2(\mathsf{X}) = \mathtt{Diag}_2(\mathsf{Y})$, then $\mathtt{Diag}_1(\mathsf{X}) = \mathtt{Diag}_1(\mathsf{Y})$. We say $\mathbf{L}_1$ is ***strictly strongly discernible in*** $\mathbf{L}_2$ (written "$\mathbf{L}_1 \sqsubset_s \mathbf{L}_2$") if $\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$ but $\mathbf{L}_2 \not\sqsubseteq_s \mathbf{L}_1$. We say that $\mathbf{L}_1$ is ***strongly discernibly equivalent to*** $\mathbf{L}_2$ (written "$\mathbf{L}_1 \sqsupseteq\sqsubseteq_s \mathbf{L}_2$") if $\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$ and $\mathbf{L}_2 \sqsubseteq_s \mathbf{L}_1$.

As the name suggests, strong discernibility is strictly stronger than discernibility.

Fact 4.1.32 *(Strong Discernibility Implies Discernibility)*. For any C-languages $\mathbf{L}_1$ and $\mathbf{L}_2$, if $\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$, then $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$.

Example 4.1.33 *(Discernibility Does Not Imply Strong Discernibility)*. Consider two languages $\mathbf{L}_1$ and $\mathbf{L}_2$, where:

- $\mathcal{L}_1 = \{p\}$
- $\mathcal{L}_2 = \{p, q\}$
- $\mathsf{C}_1 = \mathsf{C}_2 = \{x, y\}$
- $x \Vdash_1 p$ while $y \nVdash_2 p$
- $x \Vdash_2 p$ and $x \nVdash_2 q$
- $y \Vdash_2 p$ and $y \Vdash_2 q$.

Clearly, $\mathbf{L}_1 \sqsupseteq\sqsubseteq \mathbf{L}_2$. But $\mathbf{L}_1 \not\sqsubseteq_s \mathbf{L}_2$. In particular, $\mathtt{Diag}_2(\{x, y\}) = \mathtt{Diag}_2(\{x\}) = \{p\}$. But $\mathtt{Diag}_1(\{x, y\}) = \varnothing$, while $\mathtt{Diag}_1(\{x\}) = \{p\}$.

One can use **Example 4.1.25** to show that strong discernibility does not imply weak expressibility. But the converse is true (using essentially the same proof as **Proposition 4.1.22**):

Fact 4.1.34 *(Weak Expressibility Implies Strong Discernibility)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be C-languages. If $\mathbf{L}_1 \preceq \mathbf{L}_2$, then $\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$.

As before, we may ask when strong discernibility collapses to other forms of expressive power.

Proposition 4.1.35 *(Collapsing Strong Discernibility and Expressibility)*. Let $\mathbf{L}_2$ be a C-language. Then the following are equivalent:

(a) For any C-language $\mathbf{L}_1$, if $\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$, then $\mathbf{L}_1 \leqslant \mathbf{L}_2$.

(b) $\mathbf{L}_2$ has infinite conjunction and infinite disjunction.

Proof:

(a) ⇒ (b) Let $\Gamma, \Delta \subseteq \mathcal{L}_2$. Define $\mathbf{L}_1^\wedge = \langle \{p\}, \mathsf{C}_2, \Vdash_1^\wedge \rangle$ and $\mathbf{L}_1^\vee = \langle \{q\}, \mathsf{C}_2, \Vdash_1^\vee \rangle$ so that:

$$\begin{aligned} x \Vdash_1^\wedge p \quad &\Leftrightarrow \quad \forall \phi \in \Gamma\colon\ x \Vdash_2 \phi \\ x \Vdash_1^\vee q \quad &\Leftrightarrow \quad \exists \phi \in \Delta\colon\ x \Vdash_2 \phi. \end{aligned}$$

First, observe that $\mathbf{L}_1^\wedge \sqsubseteq_s \mathbf{L}_2$ and $\mathbf{L}_1^\vee \sqsubseteq_s \mathbf{L}_2$. For instance, if $\mathsf{X} \equiv_2 \mathsf{Y}$, then $\mathsf{X} \Vdash_2 \Gamma$ iff $\mathsf{Y} \Vdash_2 \Gamma$, in which case $\mathsf{X} \Vdash_1^\wedge p$ iff $\mathsf{Y} \Vdash_1^\wedge p$, and so $\mathsf{X} \equiv_1^\wedge \mathsf{Y}$. Likewise for $\mathbf{L}_1^\vee$. Hence, $\mathbf{L}_1^\wedge \leqslant \mathbf{L}_2$ and $\mathbf{L}_1^\vee \leqslant \mathbf{L}_2$. Thus, there are $\phi, \psi \in \mathcal{L}_2$ such that $[\![\phi]\!]_2 = \bigcap_{\theta\in\Gamma} [\![\theta]\!]_1$ and $[\![\psi]\!]_2 = \bigcup_{\theta\in\Delta} [\![\theta]\!]_1$.

(b) ⇒ (a) Suppose $\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$. Let $\phi \in \mathcal{L}_1$. Define ψ as follows:

$$\psi := \bigvee \left\{ \bigwedge \mathtt{Diag}_2(x) \;\middle|\; x \in \mathsf{C}_2 \text{ and } x \Vdash_1 \phi \right\}.$$

Let $x \in \mathsf{C}$. Clearly, if $x \Vdash_1 \phi$, then $x \Vdash_2 \psi$. Conversely, if $x \Vdash_2 \psi$, then there is a $y \in [\![\phi]\!]_1$ such that $x \Vdash_2 \mathtt{Diag}_2(y)$. That means $\{x, y\} \equiv_2 \{y\}$. Hence, by strong discernibility, $\{x, y\} \equiv_1 \{y\}$, which means $x \Vdash_1 \phi$. Thus, $[\![\phi]\!]_1 = [\![\psi]\!]_2$, and so $\mathbf{L}_1 \leqslant \mathbf{L}_2$. ■

This provides an interesting measure of the difference in strength between discernibility and strong discernibility. To collapse strong discernibility to expressibility, less express material needs to be present in the language (infinite conjunction and disjunction) than to collapse discernibility to expressibility (infinite conjunction and negation).

Likewise, to collapse strong discernibility to weak expressibility, fewer expressive resources need to be present (infinite setwise disjunction) than to collapse discernibility to weak expressibility (infinite setwise disjunction and setwise negation).[4]

Proposition 4.1.36 *(Collapsing Strong Discernibility and Weak Expressibility)*. Let $\mathbf{L}_2$ be a C-language. Then the following are equivalent:

(a) For any C-language $\mathbf{L}_1$, if $\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$, then $\mathbf{L}_1 \preceq \mathbf{L}_2$.

(b) $\mathbf{L}_2$ has infinite setwise disjunction.

Proof:

(a) ⇒ (b) Let $\Sigma \subseteq \wp(L_2)$. Define $\mathbf{L}_1^* = \langle \{p\}, \mathsf{C}, \Vdash_1^* \rangle$ so that for all $x \in \mathsf{C}$:

$$x \Vdash_1^* p \quad\Leftrightarrow\quad \exists \Delta \in \Sigma\colon\; x \Vdash_2 \Delta.$$

First, observe that $\mathbf{L}_1^* \sqsubseteq_s \mathbf{L}_2$. For suppose $\mathtt{Diag}_2(\mathsf{X}) = \mathtt{Diag}_2(\mathsf{Y})$.

[4] Note that "infinite setwise conjunction" is, in a sense, present in every language: if $\Sigma \subseteq \wp(\mathcal{L})$, then we can let $\bigwedge \Sigma := \bigcup \Sigma$.

Then for all $\Delta \in \Sigma$, $\mathsf{X} \Vdash_2 \Delta$ iff $\mathsf{Y} \Vdash_2 \Delta$. Hence, $\mathsf{X} \Vdash_1^* p$ iff $\mathsf{Y} \Vdash_1^* p$, i.e., $\mathtt{Diag}_1(\mathsf{X}) = \mathtt{Diag}_1(\mathsf{Y})$.

Now, since $\mathbf{L}_1^* \sqsubseteq_s \mathbf{L}_2$, we have $\mathbf{L}_1^* \preceq \mathbf{L}_2$. So there is a $\Gamma \subseteq \mathcal{L}_2$ such that $[\![p]\!]_1^* = [\![\Gamma]\!]_2 = \{x \in \mathsf{C} \mid \exists \Delta \in \Sigma\colon\ x \Vdash_2 \Delta\}$.

(b) ⇒ (a) Suppose $\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$. Let $\phi \in \mathcal{L}_1$. Define:

$$\Delta := \bigvee \{\mathtt{Diag}_2(y) \mid y \in \mathsf{C} \text{ and } y \Vdash_1 \phi\}.$$

Clearly, if $x \Vdash_1 \phi$, then $x \Vdash_2 \Delta$. Conversely, suppose $x \nVdash_1 \phi$. It suffices to show that for all $y \in \mathsf{C}$ such that $y \Vdash_1 \phi$, $x \nVdash_2 \mathtt{Diag}_2(y)$.

Suppose for *reductio* $y \Vdash_1 \phi$ but $x \Vdash_2 \mathtt{Diag}_2(y)$. Since $x \nVdash_1 \phi$, it follows that $\mathtt{Diag}_1(\{y\}) \neq \mathtt{Diag}_1(\{x, y\})$. Hence, $\mathtt{Diag}_2(\{y\}) \neq \mathtt{Diag}_2(\{x, y\})$. Since $\mathtt{Diag}_2(\{y\}) \supseteq \mathtt{Diag}_2(\{x, y\})$, there must be a $\psi \in \mathtt{Diag}_2(\{y\}) - \mathtt{Diag}_2(\{x, y\})$. That is, $y \Vdash_2 \psi$ and $x \nVdash_2 \psi$. But $x \Vdash_2 \mathtt{Diag}_2(y)$, ↯. ■

Given **Propositions 4.1.28** and **4.1.36**, one might guess that strong discernibility collapses to discernibility exactly when setwise negation is present. This would make the circle complete, since setwise negation and infinite conjunction are sufficient to express regular negation ($\neg\phi := \bigwedge \neg\{\phi\}$). If that were the case, then we would have a nice hierarchy as presented below.

$$\begin{tikzcd}
\leqslant \arrow[r] & \preceq \arrow[r] \arrow[l, bend left, "\bigwedge"] & \sqsubseteq_s \arrow[r] \arrow[l, bend left, "\bigvee^{\text{set}}"] \arrow[ll, bend right, "{\bigwedge, \bigvee}"'] & \sqsubseteq \arrow[l, dashed, bend left, "\neg^{\text{set}}"] \arrow[ll, bend right, "{\bigvee^{\text{set}}, \neg^{\text{set}}}"'] \arrow[lll, bend right=60, "{\bigwedge, \neg}"']
\end{tikzcd}$$

Certainly, we can show that if it is present, then strong discernibility collapses into discernibility:

Proposition 4.1.37 *(Setwise Negation Collapses Strong Discernibility to Discernibility)*. Let $\mathbf{L}_2$ be a C-language with setwise negation. Then for any C-language $\mathbf{L}_1$, if $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, then $\mathbf{L}_1 \sqsubseteq_s \mathbf{L}_2$.

Proof: Suppose $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$. Let $\mathsf{X}, \mathsf{Y} \subseteq \mathsf{C}$. Suppose $\mathtt{Diag}_1(\mathsf{X}) \neq \mathtt{Diag}_1(\mathsf{Y})$. Without loss of generality, let's say that $\mathsf{X} \Vdash_1 \phi$ and $\mathsf{Y} \nVdash_1 \phi$. Thus, there is a $y \in \mathsf{Y}$ such that $y \nVdash_1 \phi$. But every $x \in \mathsf{X}$ is such that $x \Vdash_1 \phi$. So for every $x \in \mathsf{X}$, $x \not\equiv_1 y$. Since $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$, $x \not\equiv_2 y$ either. Hence, for every $x \in \mathsf{X}$, there is a $\psi_x \in \mathcal{L}_2$ such that either (i) $x \Vdash_2 \psi_x$ and $y \nVdash_2 \psi_x$ or (ii) $x \nVdash_2 \psi_x$ and $y \Vdash_2 \psi_x$. Since we have setwise negation, it follows that for every $x \in \mathsf{X}$, there is a $\Gamma_x \subseteq \mathcal{L}_2$ such that $x \nVdash_2 \Gamma_x$ and $y \Vdash_2 \Gamma_x$. Hence, $\mathsf{Y} \Vdash_2 \bigcup \{\Gamma_x \mid x \in \mathsf{X}\}$ whereas $\mathsf{X} \nVdash_2 \bigcup \{\Gamma_x \mid x \in \mathsf{X}\}$. So $\mathtt{Diag}_2(\mathsf{X}) \neq \mathtt{Diag}_2(\mathsf{Y})$. ∎

Unfortunately, I have yet to find a proof (or refutation) of the claim that the presence of setwise negation is necessary for strong discernibility to collapse to discernibility.[5] I suspect it is possible to close this gap, but the gap will have to remain for now.

§ 4.2 Transforming Logical Space

In the previous section, we assumed that whenever we were comparing the expressive power of some languages, the languages in question all shared the same conception of logical space. But what about languages which do not share logical space? Can we still settle how two languages are related in expressive power even when we cannot directly compare models from one language with models of the other? In this section, we look at several ways one might do this distinct from the method developed in Chapters 2–3.

§ 4.2.1 Model-Preservation

Often, when proving that one language is translatable into another, one does not just give a mapping from formulas to formulas. In addition, one usually specifies something like a model transformation, saying how models from one evaluation space generate "equivalent" models in the other model space. This suggests one way in which expressive power might extend beyond translatability: in addition to showing how the consequence relations of the source logic are preserved in the target logic, one must also show how the *models* of the former correspond to the models of the latter.

A standard way of making this precise was articulated by Epstein [1990]:

[5] In particular, the proof strategy employed in previous collapse results does not seem to work. Let $\Gamma \subseteq \mathcal{L}_2$. Suppose we construct an $\mathbf{L}_1^* = \langle \{p\}, \mathsf{C}, \Vdash_1^* \rangle$ such that for all $x \in \mathsf{C}$, $x \Vdash_1^* p$ iff $x \nVdash_2 \Gamma$. Then one can show that $\mathbf{L}_1^* \sqsubseteq \mathbf{L}_2$, in which case $\mathbf{L}_1^* \sqsubseteq_s \mathbf{L}_2$. But there does not seem to be any obvious way (at least to me) to show from this fact that there is a $\Delta \subseteq \mathcal{L}_2$ such that $[\![\Delta]\!]_2 = \mathsf{C} - [\![\Gamma]\!]_2$.

Definition 4.2.1 *(Model-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages, let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and let $\mathsf{c}\colon \mathsf{C}_2 \to \mathsf{C}_1$. Then $\mathtt{t}$ is ***model-preserving*** via c (written "$\mathtt{t};\mathsf{c}\colon \mathbf{L}_1 \propto \mathbf{L}_2$") if:

(i) for each $x \in \mathsf{C}_1$, there is a $y \in \mathsf{C}_2$ such that $x \equiv_1 \mathsf{c}(y)$

(ii) for each $y \in \mathsf{C}_2$ and $\phi \in \mathcal{L}_1$, we have $y \Vdash_2 \mathtt{t}(\phi)$ iff $\mathsf{c}(y) \Vdash_1 \phi$.

We say $\mathtt{t}$ is ***model-preserving*** (written "$\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$") if it is model-preserving via some $\mathsf{c}\colon \mathsf{C}_2 \to \mathsf{C}_1$. We call $\mathtt{t}$ a translation and c a ***correspondence map***.

The correspondence map can be thought of as a "coarsening" of the states or models in C_2. Put another way, if $\mathsf{c}(y) \equiv_1 x$, we can think of y as essentially a refinement of x into $\mathbf{L}_2$.

Model-preservation is the natural generalization of expressibility to languages with distinct evaluation spaces. For one thing, observe that expressibility just is model-preservation via identity:

Fact 4.2.2 *(From Expressibility to Model-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some C-languages and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathbf{L}_1 \leqslant \mathbf{L}_2$ iff $\mathtt{t};\mathsf{id}\colon \mathbf{L}_1 \propto \mathbf{L}_2$.

Moreover, model-preservation induces expressibility not for the original languages involved but for modified versions of them.

Fact 4.2.3 *(From Model-Preservation to Expressibility)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages. Suppose $\mathtt{t};\mathsf{c}\colon \mathbf{L}_1 \propto \mathbf{L}_2$. Define $\mathbf{L}_1^* := \langle \mathcal{L}_1, \mathsf{C}_2, \Vdash_1^* \rangle$, where $y \Vdash_1^* \phi$ iff $\mathsf{c}(y) \Vdash_1 \phi$ for all $y \in \mathsf{C}_2$. Then $\mathtt{t}\colon \mathbf{L}_1^* \leqslant \mathbf{L}_2$.

Example 4.2.4 *(S4)*. Recall from **Example 2.1.8** that **S4** is the restriction of **K** to pointed Kripke models whose accessibility relation is a preorder. We noted that **IPL**$\rightsquigarrow$**S4**. In fact, **IPL** $\propto$ **S4**. This can be established as follows (see, e.g., Chagrov and Zakharyaschev 1997, pp. 96–97). Let G be the Gödel translation from **Example 2.1.8**. Where $\mathcal{M} = \langle W, \twoheadrightarrow, V \rangle$ is an **S4**-model, let $\mathcal{M}^* = \langle W, \twoheadrightarrow, V^* \rangle$ be the **IPL**-model such that for all $p \in \mathtt{At}$:

$$V^*(p) = \{ w \in W \mid \forall v \in W\colon\ w \twoheadrightarrow v \ \Rightarrow\ v \in V(p) \}\,.$$

(Note this is an **IPL**-model since V^* is upward-closed, i.e., if $w \in V^*(p)$ and $w \twoheadrightarrow v$, then $v \in V^*(p)$.) Define $\langle \mathcal{M}, w \rangle^* = \langle \mathcal{M}^*, w \rangle$. Then $\mathsf{G}, *\colon$ **IPL** $\propto$ **S4**. To show (i), note that if $\mathcal{M}$ is an **IPL**-model, then it is an **S4**-model. Hence, $V^* = V$, and so vacuously $\mathcal{M}, w \equiv_{\mathbf{IPL}} \mathcal{M}^*, w$. The proof that (ii) proceeds by induction on the complexity of formulas.

One useful fact about model-preserving maps is that they preserve truth-functional connectives.

Fact 4.2.5 *(Model-Preservation Preserves Truth-Functional Operators)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages and let $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$. Then assuming each language has a truth-functional negation $\neg$, we have $\mathtt{t}(\neg\phi) \equiv_2 \neg\,\mathtt{t}(\phi)$. Likewise for all the other truth-functional operators.

Observe that the correspondence map points in the opposite direction of the translation, so to speak. We could have instead formulated model-preservation with correspondence maps going the other way:

Fact 4.2.6 *(Flipping Correspondence Maps Around)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$ iff there is a map $\mathsf{r}\colon \mathsf{C}_1 \to \mathsf{C}_2$ such that:

(i) for each $y \in \mathsf{C}_2$, there is an $x \in \mathsf{C}_1$ such that $y \equiv_{\mathtt{t}[1]} \mathsf{r}(x)$

(ii) for each $x \in \mathsf{C}_1$ and $\phi \in \mathcal{L}_1$, we have $x \Vdash_1 \phi$ iff $\mathsf{r}(x) \Vdash_{\mathtt{t}[1]} \mathtt{t}(\phi)$.

On this formulation of model-preservation, the correspondence map is thought of as a "refinement" of the states or models in the source language. Usually, it turns out to be slightly easier to verify $\mathbf{L}_1$-agreement than to verify $\mathtt{t}[\mathbf{L}_1]$-agreement. So we will stick with the original formulation of model-preservation in **Definition 4.2.1** for convenience.

There is a rather simple reformulation of model-preservation that does away with the correspondence map altogether. This formulation of model-preservation will be useful when proving abstractly the existence of model-preserving maps between languages.

Definition 4.2.7 *(Correlation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. The ***t-correlation relation*** from $\mathbf{L}_1$ to $\mathbf{L}_2$ is the relation $\propto_{\mathtt{t}} \subseteq \mathsf{C}_1 \times \mathsf{C}_2$ such that for all $x \in \mathsf{C}_1$ and $y \in \mathsf{C}_2$:

$$x \propto_{\mathtt{t}} y \quad \Leftrightarrow \quad \text{for all } \phi \in \mathcal{L}_1\colon x \Vdash_1 \phi \Leftrightarrow y \Vdash_2 \mathtt{t}(\phi).$$

Often, we drop mention of $\mathbf{L}_1$ and $\mathbf{L}_2$ when context makes them clear.

Definition 4.2.8 *(Totality and Surjectivity)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. We say $\propto_{\mathtt{t}}$ is ***total*** if for all $x \in \mathsf{C}_1$, there is a $y \in \mathsf{C}_2$ such that $x \propto_{\mathtt{t}} y$. We say $\propto_{\mathtt{t}}$ is ***surjective*** if for all $y \in \mathsf{C}_2$, there is an $x \in \mathsf{C}_1$ such that $x \propto_{\mathtt{t}} y$.

Lemma 4.2.9 *(Relation Between $\propto_{\mathtt{t}}$ and c)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t};\mathtt{c}\colon \mathbf{L}_1 \propto \mathbf{L}_2$. Then for all $x \in \mathsf{C}_1$ and $y \in \mathsf{C}_2$:

$$x \propto_{\mathtt{t}} y \quad \Leftrightarrow \quad x \equiv_1 \mathtt{c}(y).$$

Proof: Let $x \in \mathsf{C}_1$ and $y \in \mathsf{C}_2$. Then $x \propto_{\mathtt{t}} y$ just in case for all $\phi \in \mathcal{L}_1$: $x \Vdash_1 \phi$ iff $y \Vdash_2 \mathtt{t}(\phi)$. But since $y \Vdash_2 \mathtt{t}(\phi)$ iff $\mathtt{c}(y) \Vdash_1 \phi$, we have $x \propto_{\mathtt{t}} y$ just in case for all $\phi \in \mathcal{L}_1$: $x \Vdash_1 \phi$ iff $\mathtt{c}(y) \Vdash_1 \phi$, i.e., $x \equiv_1 \mathtt{c}(y)$. ■

Proposition 4.2.10 *(Alternative Formulation of Model-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$ iff $\propto_{\mathtt{t}}$ is total and surjective.

Proof:

(⇒) Suppose $\mathtt{t},\mathtt{c}\colon \mathbf{L}_1 \propto \mathbf{L}_2$. Let $x \in \mathsf{C}_1$. Then there is a $y \in \mathsf{C}_2$ such that $x \equiv_1 \mathtt{c}(y)$, i.e., such that $x \propto_{\mathtt{t}} y$ by **Lemma 4.2.9**. So $\propto_{\mathtt{t}}$ is total. Moreover, if $y \in \mathsf{C}_2$, then $\mathtt{c}(y) \propto_{\mathtt{t}} y$. So $\propto_{\mathtt{t}}$ is surjective.

(⇐) Suppose $\propto_{\mathtt{t}}$ is total and surjective. Pick any $\mathtt{c}\colon \mathsf{C}_2 \to \mathsf{C}_1$ so that $\mathtt{c}(y) \propto_{\mathtt{t}} y$ for all $y \in \mathsf{C}_2$ (we know one exists by surjectivity). We just need to verify that c satisfies (i) and (ii) of **Definition 4.2.1**. First, let $x \in \mathsf{C}_1$. By totality, there is a $y \in \mathsf{C}_2$ such that $x \propto_{\mathtt{t}} y$. Since $\mathtt{c}(y) \propto_{\mathtt{t}} y$, we have $x \equiv_1 \mathtt{c}(y)$. Next, let $y \in \mathsf{C}_2$ and $\phi \in \mathcal{L}_1$. Since $\mathtt{c}(y) \propto_{\mathtt{t}} y$, we have $\mathtt{c}(y) \Vdash_1 \phi$ iff $y \Vdash_2 \mathtt{t}(\phi)$. ■

We now turn to the relationship between translations and model-preserving maps. Using **Proposition 4.2.10**, we can more simply prove an observation from Epstein [1990, p. 391, Corollary 11] that relates model-preservation to translation:

Proposition 4.2.11 *(From Model-Preservation to Translation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages and let $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$. Then $\mathtt{t}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$.

Proof: Let $\Gamma \subseteq \mathcal{L}_1$ and $\phi \in \mathcal{L}_1$. Then:

$$\begin{aligned} \Gamma \vDash_1 \phi \quad &\Leftrightarrow \quad \text{for all } x \in \mathsf{C}_1\text{: } x \Vdash_1 \Gamma \Rightarrow x \Vdash_1 \phi \\ &\Leftrightarrow \quad \text{for all } x \in \mathsf{C}_1 \text{ and all } y \in \mathsf{C}_2\text{: } x \propto_{\mathtt{t}} y \ \&\ y \Vdash_2 \mathtt{t}[\Gamma] \Rightarrow y \Vdash_2 \mathtt{t}(\phi) \\ &\Leftrightarrow \quad \text{for all } y \in \mathsf{C}_2\text{: } y \Vdash_2 \mathtt{t}[\Gamma] \Rightarrow y \Vdash_2 \mathtt{t}(\phi) \end{aligned}$$

$$\Leftrightarrow \quad \mathtt{t}[\Gamma] \vDash_2 \mathtt{t}(\phi).$$

For the second step, the left-to-right direction is by definition of $\propto_{\mathtt{t}}$, and the right-to-left direction follows by totality. For the third step, the left-to-right direction follows by surjectivity, and the right-to-left direction is obvious. ∎

To say that $\propto_{\mathtt{t}}$ is total is to say that every counterexample to an inference in $\mathbf{L}_1$ can be matched with a counterexample to that inference (translated) in $\mathbf{L}_2$. Conversely, to say that $\propto_{\mathtt{t}}$ is surjective is to say that every counterexample in $\mathbf{L}_2$ to a translated inference can be matched with a counterexample already in $\mathbf{L}_1$ to that inference. Given this, one might wonder whether translatability implies $\propto_{\mathtt{t}}$ is total and surjective. The answer is negative, as these next examples show:

Example 4.2.12 *(Propositional Logic with an Absurd State)*. Let $\mathbf{CPL}_\perp$ be just like **CPL** except $\mathsf{V}_\perp = \mathsf{V} \cup \{\perp\}$, where $\perp \notin \mathsf{V}$, and $\perp \Vdash_\perp \phi$ for all $\phi \in \mathcal{L}_{\mathtt{Prop}}$. It is straightforward to check that $\mathtt{id}\colon \mathbf{CPL} \rightsquigarrow \mathbf{CPL}_\perp$. But clearly, $\propto_{\mathtt{id}}$ is not surjective given the presence of $\perp$. Likewise, $\mathtt{id}\colon \mathbf{CPL}_\perp \rightsquigarrow \mathbf{CPL}$, even though $\propto_{\mathtt{id}}$ is not total.

One might suspect that the problem in **Example 4.2.12** arises due to the presence of the state $\perp$, which satisfies every formula. But even without such states, counterexamples arise.

Example 4.2.13 *(Propositional Logic Plus a Partial Valuation Function)*. Let v_p be a partial function that just maps p to 1, and define $\mathbf{CPL}_p$ so that $\mathsf{V}_p = \mathsf{V} \cup \{v_p\}$ and $\Vdash_p$ works just like $\Vdash$ for each $v \in \mathsf{V}$, and $v_p \Vdash_p \phi$ iff $p \vDash_{\mathbf{CPL}} \phi$. Then it is straightforward to check that there is a translation from **CPL** into $\mathbf{CPL}_p$; but the resulting $\propto$ relation will not be surjective. Likewise for totality.

These examples illustrate a crucial point about the difference between translatability and model-preservability. The question of whether there exists a translation between some languages can be reduced to the question of whether there exists a translation between the logics of those languages. This is not so for model-preserving maps. It makes no sense, without further specification, to ask whether there exists a model-preserving map between some *logics* since the answer depends on which semantics we have in mind for these logics. The languages from **Examples 4.2.12–4.2.13** (**CPL**, $\mathbf{CPL}_\perp$, $\mathbf{CPL}_p$) all have the same underlying logic (CPL) but differ in their underlying semantics. This is not just a feature of these gerrymandered semantic frameworks: it is also a feature of natural and independently motivated semantic theories for **CPL** developed in the literature.

Example 4.2.14 *(Model-Preserving from Classical Logic into Intuitionistic Logic).* We noted in Example 2.2.8 that **CPL**$\rightsquigarrow$**IPL**. But it is not true that **CPL** $\propto$ **IPL**. For suppose $\mathtt{t}\colon$ **CPL** $\propto$ **IPL**. Let $\mathcal{M}_{\mathbf{IPL}}$ be the canonical model for **IPL** and let $\mathcal{M}^{+}_{\mathbf{IPL}}$ be result of adding a root r to $\mathcal{M}_{\mathbf{IPL}}$ that sees every state in $\mathcal{M}_{\mathbf{IPL}}$. It is easy to verify that $\mathcal{M}^{+}_{\mathbf{IPL}}, r \Vdash \phi$ iff $\vDash_{\mathbf{IPL}} \phi$. Hence, there must be a classical valuation v such that $v \Vdash \phi$ iff ϕ is a classical tautology, $\lightning$.

However, there is a natural semantics for **CPL** that can be mapped into **IPL** in a model-preserving manner, viz., the possibility semantics developed by Humberstone [1981] and Holliday [2018]. Let us write "$\leqslant$" for the accessibility relation in **IPL**-models. Let P be the class of pairs of the form $\langle \mathcal{M}, w\rangle$ where $\mathcal{M}$ is an **IPL**-model, $w \in W^{\mathcal{M}}$, and $\mathcal{M}$ satisfies the following "refinability" constraint:

$$\forall w\ [w \notin V(p) \Rightarrow \exists v \geqslant w\, \forall u \geqslant v\ (u \notin V(p))]$$

We call such models ***possibility models***. Define $\mathbf{P} = \langle \mathcal{L}_{\mathtt{Prop}}(\vee, \rightarrow), \mathsf{P}, \Vdash_{\mathbf{P}}\rangle$, where:

$$\begin{array}{lcl}
\mathcal{M}, w \Vdash_{\mathbf{P}} p & \Leftrightarrow & w \in V(p) \\
\mathcal{M}, w \Vdash_{\mathbf{P}} \neg\phi & \Leftrightarrow & \forall v \geqslant w\colon\ \mathcal{M}, v \nVdash_{\mathbf{P}} \phi \\
\mathcal{M}, w \Vdash_{\mathbf{P}} \phi \wedge \psi & \Leftrightarrow & \mathcal{M}, w \Vdash_{\mathbf{P}} \phi \text{ and } \mathcal{M}, w \Vdash_{\mathbf{P}} \psi \\
\mathcal{M}, w \Vdash_{\mathbf{P}} \phi \vee \psi & \Leftrightarrow & \forall v \geqslant w\, \exists u \geqslant v\colon\ \mathcal{M}, u \Vdash_{\mathbf{P}} \phi \text{ or } \mathcal{M}, u \Vdash_{\mathbf{P}} \psi \\
\mathcal{M}, w \Vdash_{\mathbf{P}} \phi \rightarrow \psi & \Leftrightarrow & \forall v \geqslant w\colon\ \mathcal{M}, v \Vdash_{\mathbf{P}} \phi \Rightarrow \mathcal{M}, v \Vdash_{\mathbf{P}} \psi.
\end{array}$$

It is straightforward to show that $\mathsf{L}_{\mathbf{P}} = \mathsf{CPL}$.

We now show that **P** $\propto$ **IPL**.[a] For instance, let $\mathtt{N}$ be the double-negation translation from Example 2.1.8. Where $\mathcal{M} = \langle W, \leqslant, V\rangle$ is an intuitionistic model, let $\mathcal{M}^* = \langle W, \leqslant, V^*\rangle$ where:

$$V^*(p) := \{w \in W \mid \mathcal{M}, w \Vdash \neg\neg p\}.$$

If $\mathcal{M}$ is an intuitionistic model, then $\mathcal{M}^*$ is a possibility model. For if $w \notin V^*(p)$, that means $\mathcal{M}, w \nVdash \neg\neg p$, i.e, there is a $v \geqslant w$ such that $\mathcal{M}, v \Vdash \neg p$. Thus for all $u \geqslant v$, $\mathcal{M}, u \Vdash \neg p$ and hence $\mathcal{M}, u \nVdash \neg\neg p$, i.e., $u \notin V^*(p)$. Moreover, if $\mathcal{M}$ is a possibility model, then $\mathcal{M}^* = \mathcal{M}$, since in that case $\mathcal{M}, w \Vdash \neg\neg p$ iff $\mathcal{M}, w \Vdash p$. This means (i) of Definition 4.2.1 is automatically satisfied, since every possibility model is an intuitionistic model. As for (ii), if $\mathcal{M}$ is an intuitionistic model, if $w \in W^{\mathcal{M}}$, and if $\phi \in \mathcal{L}_{\mathtt{Prop}}$, then by induction, $\mathcal{M}, w \Vdash \mathtt{t}(\phi)$ iff $\mathcal{M}^*, w \Vdash \phi$.

[a] I have Wes Holliday to thank for this observation.

What these examples seem to have in common is partiality: none of the alternative versions of **CPL** are opinionated (in the sense of **Definition 1.2.1**). It turns out that this is the only barrier to translatability implying model-preservability.

Proposition 4.2.15 *(From Translation to Model-Preservation)*. Suppose $\mathtt{t}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$ and suppose $\mathbf{L}_1$ and $\mathtt{t}[\mathbf{L}_1]$ are opinionated. Then $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$.

Proof: By **Lemma 1.2.4**, if C_1 contains an absurd state in $\mathbf{L}_1$, then every point in C_1 is an absurd state in $\mathbf{L}_1$. Thus, $\vDash_1 \mathcal{L}_1$, and so $\vDash_{\mathtt{t}[1]} \mathtt{t}[\mathcal{L}_1]$. So every point in C_2 is an absurd state in $\mathtt{t}[\mathbf{L}_1]$ as well. So for all $x \in \mathsf{C}_1$ and $y \in \mathsf{C}_2$, $x \propto_{\mathtt{t}} y$. Likewise if C_2 contains an absurd state in $\mathtt{t}[\mathbf{L}_1]$. Hence, if either $\mathbf{L}_1$ or $\mathtt{t}[\mathbf{L}_1]$ contain absurd states, $\propto_{\mathtt{t}}$ is trivially total and surjective. So we may assume throughout that neither $\mathbf{L}_1$ nor $\mathtt{t}[\mathbf{L}_1]$ contain absurd states.

First, totality. Let $x \in \mathsf{C}_1$. Suppose there is no $y \in \mathsf{C}_2$ such that $x \propto_{\mathtt{t}} y$.

Claim: For no $y \in \mathsf{C}_2$ does $y \Vdash_2 \mathtt{t}[\mathtt{Diag}_1(x)]$.

Subproof: Suppose for *reductio* that $y \Vdash_2 \mathtt{t}[\mathtt{Diag}_1(x)]$. If there were no $\phi \in \mathcal{L}_1$ such that $x \nVdash_1 \phi$ and $y \Vdash_2 \mathtt{t}(\phi)$, then $x \propto_{\mathtt{t}} y$. So we may assume that $y \Vdash_2 \mathtt{t}[\mathtt{Diag}_1(x) \cup \{\phi\}]$ for some $\phi \in \mathcal{L}_1$ such that $x \nVdash_1 \phi$. Since $\mathbf{L}_1$ is opinionated, and since $\vDash_1$ is explosive, $\mathtt{Diag}_1(x) \cup \{\phi\} \vDash_1 \mathcal{L}_1$. Thus, $\mathtt{t}[\mathtt{Diag}_1(x) \cup \{\phi\}] \vDash_2 \mathtt{t}[\mathcal{L}_1]$. But then $y \Vdash_2 \mathtt{t}[\mathcal{L}_1]$, $\lightning$. ■

So no $y \in \mathsf{C}_2$ satisfies all of $\mathtt{t}[\mathtt{Diag}_1(x)]$. But now let $\phi \in \mathcal{L}_1$ be such that $x \nVdash_1 \phi$. Since $\vDash_2$ is explosive, $\mathtt{t}[\mathtt{Diag}_1(x)] \vDash_2 \mathtt{t}(\phi)$. But $\mathtt{Diag}_1(x) \nvDash_1 \phi$ (since x is a counterexample), contrary to $\mathtt{t}$ being a translation, $\lightning$. Therefore, $\propto_{\mathtt{t}}$ must be total.

Second, surjectivity. Let $y \in \mathsf{C}_2$. Let $\mathtt{Diag}_2{}^{-1}(y) := \{\phi \in \mathcal{L}_1 \mid y \Vdash_2 \mathtt{t}(\phi)\}$ (so notice that $\mathtt{t}[\mathtt{Diag}_2{}^{-1}(y)] = \mathtt{Diag}_{\mathtt{t}[1]}(y)$). Suppose there is no $x \in \mathsf{C}_1$ such that $x \propto_{\mathtt{t}} y$.

Claim: For no $x \in \mathsf{C}_1$ does $x \Vdash_1 \mathtt{Diag}_2{}^{-1}(y)$.

Subproof: Suppose for *reductio* that there is such an $x \in \mathsf{C}_1$. By the same reasoning as above, there must be some $\psi \in \mathcal{L}_1$ such that $x \Vdash_1 \psi$ but $y \nVdash_2 \mathtt{t}(\psi)$. By totality of $\propto_{\mathtt{t}}$, there is a $y' \in \mathsf{C}_2$ such that $x \propto_{\mathtt{t}} y'$. But then $\mathtt{Diag}_{\mathtt{t}[1]}(y) \subset \mathtt{Diag}_{\mathtt{t}[1]}(y')$, contrary to $\mathtt{t}[\mathbf{L}_1]$ opinionation, $\lightning$. ■

Thus, $\mathtt{Diag}_2{}^{-1}(y) \vDash_1 \mathcal{L}_1$. But $y \Vdash_2 \mathtt{t}[\mathtt{Diag}_2{}^{-1}(y)]$ while $y \nVdash_2 \mathtt{t}[\mathcal{L}_1]$, since there are no absurd states in $\mathtt{t}[\mathbf{L}_1]$, $\lightning$. Therefore, $\propto_{\mathtt{t}}$ is surjective. ∎

The existence of model-preserving maps is not sufficient to guarantee that both languages are opinionated (since every non-opinionated language is trivially translatable into itself via a model-preserving map). It is sufficient to guarantee, however, that either both languages are opinionated or neither is.

Proposition 4.2.16 *(Opinionation Reflected and Preserved)*. Let $\mathtt{t}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$.

(a) If $\propto_{\mathtt{t}}$ is total and $\mathtt{t}[\mathbf{L}_1]$ is opinionated, then $\mathbf{L}_1$ is opinionated.

(b) If $\propto_{\mathtt{t}}$ is surjective and $\mathbf{L}_1$ is opinionated, then $\mathtt{t}[\mathbf{L}_1]$ is opinionated.

Proof: We just prove (a), since the proof of (b) is symmetric. Suppose $\propto_{\mathtt{t}}$ is total and $\mathtt{t}[\mathbf{L}_1]$ is opinionated but $\mathbf{L}_1$ is not opinionated. Let $x, x' \in \mathsf{C}_1$ be such that $\mathtt{Diag}_1(x) \subset \mathtt{Diag}_1(x')$. Then since $\propto_{\mathtt{t}}$ is total, there are $y, y' \in \mathsf{C}_2$ such that $x \propto_{\mathtt{t}} y$ and $x' \propto_{\mathtt{t}} y'$. But then that $\mathtt{Diag}_{\mathtt{t}[1]}(y) \subset \mathtt{Diag}_{\mathtt{t}[1]}(y')$, $\lightning$. ∎

Corollary 4.2.17 *(Model-Preservation Preserves (Non-)Opinionation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages and let $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$. Then $\mathbf{L}_1$ is opinionated iff $\mathtt{t}[\mathbf{L}_1]$ is.

This raises the question of whether translatability between two non-opinionated languages is sufficient to guarantee model-preservability. The answer is negative. In fact, $\mathbf{CPL}_\perp$ from **Example 4.2.12** and $\mathbf{CPL}_p$ from **Example 4.2.13** are translationally isomorphic via $\mathtt{id}$, but $\propto_{\mathtt{id}}$ is neither total nor surjective.

So translatability does not generally imply model-preservability when the languages involved are not opinionated. However, translatability does imply model-preservability for a special class of non-opinionated languages:

Proposition 4.2.18 *(Model-Preservability Collapses to Translatability for Canonical Languages)*. Let L_1 and L_2 be logics and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathbf{Can}(\mathsf{L}_1) \propto \mathbf{Can}(\mathsf{L}_2)$ iff $\mathtt{t}\colon \mathbf{Can}(\mathsf{L}_1) \rightsquigarrow \mathbf{Can}(\mathsf{L}_2)$.

Proof: The left-to-right direction is already given by **Proposition 4.2.11**. For the right-to-left direction, suppose $\mathtt{t}\colon \mathbf{Can}(\mathsf{L}_1) \rightsquigarrow \mathbf{Can}(\mathsf{L}_2)$. We show $\propto_{\mathtt{t}}$ is total and surjective.

First, totality. Let $\Gamma \in \mathtt{Th}(\mathsf{L}_1)$. Define $\Gamma^* := \mathtt{Cn}_2(\mathtt{t}[\Gamma])$. Note that:

$$\begin{aligned} \mathtt{t}(\phi) \in \Gamma^* &\Leftrightarrow \mathtt{t}[\Gamma] \vDash_2 \mathtt{t}(\phi) \\ &\Leftrightarrow \Gamma \vDash_1 \phi \\ &\Leftrightarrow \phi \in \Gamma \\ &\Leftrightarrow \mathtt{t}(\phi) \in \mathtt{t}[\Gamma]. \end{aligned}$$

The third step follows from the fact that Γ is an L_1-theory. Hence:

$$\begin{aligned} \Gamma \Vdash_1 \phi &\Leftrightarrow \phi \in \Gamma \\ &\Leftrightarrow \mathtt{t}(\phi) \in \mathtt{t}[\Gamma] \\ &\Leftrightarrow \mathtt{t}(\phi) \in \Gamma^* \\ &\Leftrightarrow \Gamma^* \Vdash_2 \mathtt{t}(\phi). \end{aligned}$$

So $\Gamma \propto_{\mathtt{t}} \Gamma^*$.

Next, surjectivity. Let $\Delta \in \mathtt{Th}(\mathsf{L}_2)$. Define $\Delta^* := \{\phi \in \mathcal{L}_1 \mid \mathtt{t}(\phi) \in \Delta\}$. Note that $\Delta^* \in \mathtt{Th}(\mathsf{L}_1)$. For suppose $\Delta^* \vDash_1 \phi$. Then $\mathtt{t}[\Delta^*] = \Delta \vDash_2 \mathtt{t}(\phi)$. And since Δ is an L_2-theory, $\mathtt{t}(\phi) \in \Delta$. So by definition of Δ^*, $\phi \in \Delta^*$. Now:

$$\begin{aligned} \Delta^* \Vdash_1 \phi &\Leftrightarrow \phi \in \Delta^* \\ &\Leftrightarrow \mathtt{t}(\phi) \in \Delta \\ &\Leftrightarrow \Delta \Vdash_2 \mathtt{t}(\phi). \end{aligned}$$

So $\Delta^* \propto_{\mathtt{t}} \Delta$. ■

Fact 4.2.19 *(Translatability Elsewhere)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$ iff $\mathtt{t}\colon \mathbf{Can}(\mathsf{L}_1) \rightsquigarrow \mathbf{Can}(\mathsf{L}_2)$.

Corollary 4.2.20 *(Translatibility Implies Model-Preservability Elsewhere)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages. Suppose $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$ iff $\mathtt{t}\colon \mathbf{Can}(\mathsf{L}_1) \propto \mathbf{Can}(\mathsf{L}_2)$.

Just as we saw that expressibility, weak expressibility, and discernibility collapse under certain circumstances, one might wonder when translatability and model-preservability collapse. The answer is almost never: the collapse happens only for languages with trivial logics.

Proposition 4.2.21 *(Translatability and Model-Preservability Almost Never Collapse)*. Let $\mathbf{L}_2$ be a nonempty language. Suppose that for every nonempty language $\mathbf{L}_1$ and every $\mathtt{t}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$, we have $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$. Then $\vDash_2 \mathcal{L}_2$.

Proof: Clearly, $\mathtt{id}\colon \mathbf{Can}(\mathsf{L}_2) \rightsquigarrow \mathbf{L}_2$. Hence, $\mathtt{id}\colon \mathbf{Can}(\mathsf{L}_2) \propto \mathbf{L}_2$. Since $\mathtt{Cn}_2(\mathcal{L})$ is an absurd state in $\mathbf{Can}(\mathsf{L}_2)$, there must also be an absurd state in $\mathbf{L}_2$. Now, suppose $\nvDash_2 \mathcal{L}_2$. That means there are some non-absurd states in $\mathbf{L}_2$. By **Fact 1.2.3**, removing the absurd states does not affect consequence. Thus, where $\mathbf{L}_2^-$ is just like $\mathbf{L}_2$ except $\mathsf{C}^- = \mathsf{C} - \{x \in \mathsf{C} \mid x \Vdash_2 \mathcal{L}\}$, then $\mathtt{id}\colon \mathbf{L}_2^- \rightsquigarrow \mathbf{L}_2$. By hypothesis, C^- is nonempty. So $\mathtt{id}\colon \mathbf{L}_2^- \propto \mathbf{L}_2$. But $\mathtt{id}$ cannot be model-preserving, since $\mathbf{L}_2$ has absurd states and $\mathbf{L}_2^-$ doesn't, so $\propto_{\mathtt{t}}$ cannot be surjective, $\lightning$. ∎

In short, for any nontrivial language, we can find another language that is translatable into it but not in a model-preserving way. Either our target language has absurd states (in which case, a language removing the absurd states would work) or else it does not (in which case the canonical language for its logic would work).

Note that **Proposition 4.2.21** fails if we allow $\mathbf{L}_2$ to be empty (i.e., $\mathcal{L}_2 = \varnothing$). For then $\mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$ only if $\mathcal{L}_1 = \varnothing$, so no nonempty $\mathbf{L}_1$ is translatable into $\mathbf{L}_2$. If, however, we drop the requirement that $\mathbf{L}_1$ be nonempty, then the result is trivial (regardless of whether we require $\mathbf{L}_2$ to be nonempty): if $\mathbf{L}_2$ is such that $\mathsf{C}_2 \neq \varnothing$, then where $\mathbf{L}_1 = \langle \varnothing, \varnothing, \varnothing \rangle$, we have $\varnothing\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$ even though $\varnothing$ is not model-preserving here; and if $\mathsf{C}_2 = \varnothing$, then where $\mathbf{L}_1 = \langle \mathcal{L}_2, \{x\}, \{x\} \times \mathcal{L} \rangle$, we have that $\mathtt{id}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$ even though $\mathtt{id}$ is not model-preserving here.

§ 4.2.2 Model-Coarsening

We noted that model-preservation is the natural generalization of expressibility to languages with distinct evaluation spaces. It is easy to generalize the notion of weak expressibility in a similar manner (if we make $\mathtt{t}\colon \mathcal{L}_1 \to \wp(\mathcal{L}_2)$), so we will not discuss it further. Instead, we discuss the natural generalization of discernibility:

Definition 4.2.22 *(Model-Coarsening)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages and let $\mathsf{c}\colon \mathsf{C}_2 \to \mathsf{C}_1$. Then c is a ***model-coarsening*** of $\mathbf{L}_2$ to $\mathbf{L}_1$ if:

(i) for each $x \in \mathsf{C}_1$, there is a $y \in \mathsf{C}_2$ such that $x \equiv_1 \mathsf{c}(y)$

(ii) for each $x, y \in \mathsf{C}_2$, if $x \equiv_2 y$, then $\mathsf{c}(x) \equiv_1 \mathsf{c}(y)$.

Fact 4.2.23 *(From Discernibility to Model-Coarsening)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some C-languages. Then $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ iff id is a model-coarsening of $\mathbf{L}_2$ to $\mathbf{L}_1$.

Fact 4.2.24 *(From Model-Coarsening to Discernibility)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages. Suppose $\mathsf{c}\colon \mathsf{C}_2 \to \mathsf{C}_1$ is model-coarsening of $\mathbf{L}_2$ to $\mathbf{L}_1$. Define $\mathbf{L}_1^* := \langle \mathcal{L}_1, \mathsf{C}_2, \Vdash_1^* \rangle$, where $y \Vdash_1^* \phi$ iff $\mathsf{c}(y) \Vdash_1 \phi$ for all $y \in \mathsf{C}_2$. Then $\mathbf{L}_1^* \sqsubseteq \mathbf{L}_2$.

Like in the case of model-preservation, where we could use either coarsenings or refinements as the correspondence maps, we can define a notion of model-refinement as follows:

Definition 4.2.25 *(Model-Refinement)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages and let $\mathsf{r}\colon \mathsf{C}_1 \to \wp(\mathsf{C}_2) - \{\varnothing\}$. Then r is a ***model-refinement*** of $\mathbf{L}_1$ to $\mathbf{L}_2$ if:

(i) for each $y \in \mathsf{C}_2$, there is an $x \in \mathsf{C}_1$ such that $y \in \mathsf{r}(x)$

(ii) for each $x, y \in \mathsf{C}_1$, either $\mathsf{r}(x) \cap \mathsf{r}(y) = \varnothing$ or $\mathsf{r}(x) = \mathsf{r}(y)$

(iii) for each $z \in \mathsf{C}_1$, $\mathsf{r}(z)$ is closed under $\equiv_2$, i.e., for all $x, y \in \mathsf{C}_2$, if $x \equiv_2 y$ and $x \in \mathsf{r}(z)$, then $y \in \mathsf{r}(z)$

(iv) for each $x, y \in \mathsf{C}_1$, if $\mathsf{r}(x) = \mathsf{r}(y)$, then $x \equiv_1 y$.

Proposition 4.2.26 *(Converting a Model-Coarsening into a Model-Refinement)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages. Let $\mathsf{c}\colon \mathsf{C}_2 \to \mathsf{C}_1$. Define $\mathsf{r}\colon \mathsf{C}_1 \to \wp(\mathsf{C}_2)$ so that:

$$\mathsf{r}(x) = \{ y \in \mathsf{C}_2 \mid \mathsf{c}(y) \equiv_1 x \}.$$

Then c is a model-coarsening iff r is a model-refinement.

Proof:

($\Rightarrow$) Observe that $\mathsf{r}(x) \neq \varnothing$, since for each $x \in \mathsf{C}_1$, there is a $y \in \mathsf{C}_2$ such that $x \equiv_1 \mathsf{c}(y)$. So r is a potential candidate for being a model-refinement. We verify that r satisfies the conditions for being a model-refinement.

First, observe that for all $y \in \mathsf{C}_2$, $y \in \mathsf{r}(\mathsf{c}(y))$, since y is trivially such a $z \in \mathsf{C}_2$ such that $\mathsf{c}(z) \equiv_1 \mathsf{c}(y)$. Hence, (i) is satisfied.

Second, if $u \in \mathsf{r}(x) \cap \mathsf{r}(y) \neq \varnothing$, and if $z \in \mathsf{r}(x)$, then $\mathsf{c}(z) \equiv_1 x \equiv_1 \mathsf{c}(u) \equiv_1 y$. Hence, $z \in \mathsf{r}(y)$. Via a symmetric argument, if $z \in \mathsf{r}(y)$, then $z \in \mathsf{r}(x)$. So $\mathsf{r}(x) = \mathsf{r}(y)$, i.e., (ii) is satisfied.

Third, suppose $x \equiv_2 y$ and $x \in \mathsf{r}(z)$ for some $z \in \mathsf{C}_1$. Then $\mathsf{c}(x) \equiv_1 z$ by definition of r. But since c is a model-coarsening, $\mathsf{c}(x) \equiv_1 \mathsf{c}(y)$. Hence, $\mathsf{c}(y) \equiv_1 z$, so by definition of r, $y \in \mathsf{r}(z)$. Thus, (iii) is satisfied.

Finally, suppose $x, y \in \mathsf{C}_1$ are such that $\mathsf{r}(x) = \mathsf{r}(y)$. Since $\mathsf{r}(x) \neq \varnothing$, letting $z \in \mathsf{r}(x)$, $y \equiv_1 \mathsf{c}(z) \equiv_1 x$. So (iv) is satisfied.

($\Leftarrow$) We verify that c satisfies the conditions for being a model-coarsening. First, let $x \in \mathsf{C}_1$. Since $\mathsf{r}(x)$ is non-empty, there is a $y \in \mathsf{C}_2$ such that $\mathsf{c}(y) \equiv_1 x$.

Next, let $x, y \in \mathsf{C}_2$ be such that $x \equiv_2 y$. By the closure condition on model-refinements, and by the fact that $x \in \mathsf{r}(\mathsf{c}(x))$, $y \in \mathsf{r}(\mathsf{c}(x))$. Hence, by definition of r, $\mathsf{c}(y) \equiv_1 \mathsf{c}(x)$. ■

One might wonder what the relationship between translatability and model-coarsening is. It turns out that one language can even be discernible in another without being translatable into it.

Example 4.2.27 *(Discernibility without Translatability)*. Let $\mathbf{CPL}^-$ be just like $\mathbf{CPL}$ except for the valuation function v such that $v(p) = 1$ for all $p \in \mathtt{At}$, we set $v \nVdash^- p_1$ (keeping the semantic clauses for $\neg$ and $\wedge$ the same). Clearly, $\mathbf{CPL}^- \sqsubseteq \mathbf{CPL}$. But $\mathbf{CPL}^- \not\rightsquigarrow \mathbf{CPL}$, since $\mathbf{CPL}$ is compact and $\mathbf{CPL}^-$ is not. In particular, $\mathtt{At} \vDash^- \bot$ even though no finite subset of $\mathtt{At}$ is $\mathbf{CPL}^-$-inconsistent.

Of course, for opinionated languages, translatability implies model-preservation, which in turn implies model-coarsening.

§ 4.2.3 Model-Corroboration

So model-preservation is the semantic correlate of translatability at least roughly. Is there a semantic correlate of translational equivalence? The answer is affirmative.

Definition 4.2.28 *(Model-Corroborating)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages, let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$, and let $\mathsf{c}\colon \mathsf{C}_2 \to \mathsf{C}_1$ and $\mathsf{d}\colon \mathsf{C}_1 \to \mathsf{C}_2$. We say that $\mathtt{t}$ and $\mathtt{s}$ are ***model-corroborating*** via c and d (written "$\mathtt{t}, \mathtt{s}; \mathsf{c}, \mathsf{d}\colon \mathbf{L}_1 \infty \mathbf{L}_2$") if:

(i) $\mathtt{t}; \mathsf{c}\colon \mathbf{L}_1 \propto \mathbf{L}_2$

(ii) $\mathtt{s}; \mathsf{d}\colon \mathbf{L}_2 \propto \mathbf{L}_1$

(iii) for all $x \in \mathsf{C}_1$, $x \equiv_1 \mathsf{c}(\mathsf{d}(x))$

(iv) for all $y \in \mathsf{C}_2$, $y \equiv_2 \mathsf{d}(\mathsf{c}(y))$.

We say $\mathtt{t}$ and $\mathtt{s}$ are ***model-corroborating*** (written "$\mathtt{t}, \mathtt{s}\colon \mathbf{L}_1 \infty \mathbf{L}_2$") if they are model-corroborating via some $\mathsf{c}\colon \mathsf{C}_2 \to \mathsf{C}_1$ and $\mathsf{d}\colon \mathsf{C}_1 \to \mathsf{C}_2$.

Just as with model-preservation, there is a relatively nice statement of when two maps $\mathtt{t}$ and $\mathtt{s}$ are model-corroborating in terms of correlation relations $\propto_{\mathtt{t}}$ and $\propto_{\mathtt{s}}$, though it requires a couple of intermediate steps.

Lemma 4.2.29 *(Model-Preservation Reflects Equivalence on States)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages and let $\mathtt{t};\mathtt{c}\colon \mathcal{L}_1 \propto \mathcal{L}_2$. Then for all $y, y' \in \mathsf{C}_2$, $\mathtt{c}(y) \equiv_1 \mathtt{c}(y')$ iff $y \equiv_{\mathtt{t}[1]} y'$.

Proof: For the right-to-left direction, since $\mathtt{c}(y) \propto_{\mathtt{t}} y$ and $\mathtt{c}(y') \propto_{\mathtt{t}} y'$ by **Lemma 4.2.9**, and since $y \Vdash_2 \mathtt{t}(\phi)$ iff $y' \Vdash_2 \mathtt{t}(\phi)$, we have $\mathtt{c}(y) \Vdash_1 \phi$ iff $\mathtt{c}(y') \Vdash_1 \phi$. The left-to-right direction follows easily from **Lemma 4.2.9** since $\mathtt{t}[\mathbf{L}_1]$ is restricted to $\mathtt{t}[\mathcal{L}_1]$. ∎

Lemma 4.2.30 *(Model-Corroboration Implies Correlation Relations are Inverses)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}, \mathtt{s}\colon \mathbf{L}_1 \infty \mathbf{L}_2$. Then we have $\propto_{\mathtt{t}} = {\propto_{\mathtt{s}}}^{-1}$.

Proof: Suppose $\mathtt{t}, \mathtt{s}; \mathtt{c}, \mathtt{d}\colon \mathbf{L}_1 \infty \mathbf{L}_2$. It suffices to show that $x \propto_{\mathtt{t}} y$ implies $y \propto_{\mathtt{s}} x$, since the proof of the converse is symmetric.

Let $x \propto_{\mathtt{t}} y$. By **Proposition 4.2.10**, $\propto_{\mathtt{s}}$ is surjective, so there is an $x' \in \mathsf{C}_1$ such that $y \propto_{\mathtt{s}} x'$. By **Lemma 4.2.9**, $\mathtt{d}(x') \equiv_2 y$. By **Lemma 4.2.29**, $\mathtt{c}(\mathtt{d}(x')) \equiv_1 \mathtt{c}(y)$. But $\mathtt{c}(\mathtt{d}(x')) \equiv_1 x'$ by **Definition 4.2.28**, and $\mathtt{c}(y) \equiv_1 x$ by **Lemma 4.2.9**. So $x \equiv_1 x'$, and thus, $y \propto_{\mathtt{s}} x$. ∎

Proposition 4.2.31 *(Alternative Formulations of Model-Corroboration)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$, and let $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. Then the following are equivalent:

(a) $\mathtt{t}, \mathtt{s}\colon \mathbf{L}_1 \infty \mathbf{L}_2$

(b) $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$ and $\mathtt{s}\colon \mathbf{L}_2 \propto \mathbf{L}_1$ so that $\propto_{\mathtt{s}} = {\propto_{\mathtt{t}}}^{-1}$

(c) The following two properties hold:

 (i) for all $x \in \mathsf{C}_1$, there is a $y \in \mathsf{C}_2$ such that $x \propto_{\mathtt{t}} y$ and for all $x' \in \mathsf{C}_1$, if $y \propto_{\mathtt{s}} x'$, then $x \equiv_1 x'$

 (ii) for all $y \in \mathsf{C}_2$, there is a $x \in \mathsf{C}_1$ such that $y \propto_{\mathtt{s}} x$ and for all $y' \in \mathsf{C}_2$, if $x \propto_{\mathtt{t}} y'$, then $y \equiv_2 y'$.

Proof:

(a) ⇒ (b) By **Lemma 4.2.30**.

(b) ⇒ (a) Suppose $\propto_{\mathtt{s}} = {\propto_{\mathtt{t}}}^{-1}$. It suffices to show **Definition 4.2.28**(iii), since the proof of (iv) is symmetric. First, by **Lemma 4.2.9**, $\mathtt{d}(x) \propto_{\mathtt{s}} x$ since trivially $\mathtt{d}(x) \equiv_2 \mathtt{d}(x)$. But since $\propto_{\mathtt{s}} = {\propto_{\mathtt{t}}}^{-1}$, $x \propto_{\mathtt{t}} \mathtt{d}(x)$. Again by **Lemma 4.2.9**, that means $x \equiv_1 \mathtt{c}(\mathtt{d}(x))$.

(b) ⇒ (c) We just prove (i), since the proof of (ii) is symmetric. Since $\mathtt{t}$ is model-preserving, $\propto_{\mathtt{t}}$ is total by **Proposition 4.2.10**. So there is a $y \in \mathsf{C}_2$ such that $x \propto_{\mathtt{t}} y$. Pick any such y. Then if $y \propto_{\mathtt{s}} x'$ where $x' \in \mathsf{C}_1$, then since $\propto_{\mathtt{s}} = {\propto_{\mathtt{t}}}^{-1}$, $x' \propto_{\mathtt{t}} y$. So $x \Vdash_1 \phi$ iff $y \Vdash_2 \mathtt{t}(\phi)$ iff $x' \Vdash_1 \phi$, i.e., $x \equiv_1 x'$.

(c) ⇒ (b) Suppose $\mathtt{t}$ and $\mathtt{s}$ satisfy (i) and (ii). We first need to establish that $\mathtt{t}$ and $\mathtt{s}$ are model-preserving. By **Proposition 4.2.10**, it suffices to show that $\propto_{\mathtt{t}}$ and $\propto_{\mathtt{s}}$ are total and surjective. We just show $\propto_{\mathtt{t}}$ is total and surjective, noting the proof that $\propto_{\mathtt{s}}$ is total and surjective is symmetric. Totality follows by (i) automatically. For surjectivity, let $y \in \mathsf{C}_2$. By (ii), there is an $x \in \mathsf{C}_1$ such that $y \propto_{\mathtt{s}} x$ and whenever $x \propto_{\mathtt{t}} y'$, we have $y \equiv_2 y'$. By the totality of $\propto_{\mathtt{t}}$, there is a $z \in \mathsf{C}_2$ such that $x \propto_{\mathtt{t}} z$. Picking such a z, we have $y \equiv_2 z$. That means $x \Vdash_1 \phi$ iff $z \Vdash_2 \mathtt{t}(\phi)$, iff $y \Vdash_2 \mathtt{t}(\phi)$, i.e., $x \propto_{\mathtt{t}} y$. So $\propto_{\mathtt{t}}$ is surjective.

Next, we need to establish that $\propto_{\mathtt{s}} = {\propto_{\mathtt{t}}}^{-1}$. We just show that $x \propto_{\mathtt{t}} y$ implies $y \propto_{\mathtt{s}} x$. Suppose $x \propto_{\mathtt{t}} y$. By (ii), there is a $z \in \mathsf{C}_1$ such that $y \propto_{\mathtt{s}} z$ and for all $y' \in \mathsf{C}_2$, if $z \propto_{\mathtt{t}} y'$, then $y \equiv_2 y'$. Since $\propto_{\mathtt{t}}$ is total, there is at least one $y' \in \mathsf{C}_2$ such that $z \propto_{\mathtt{t}} y'$. Since $y \equiv_2 y'$, we have $x \propto_{\mathtt{t}} y'$. So $z \Vdash_1 \phi$ iff $y' \Vdash_2 \mathtt{t}(\phi)$ iff $x \Vdash_1 \phi$. Thus, $x \equiv_1 z$, and so $y \propto_{\mathtt{s}} x$. ■

Just as model-preservation implies translation but not *vice versa* unless the languages are both opinionated, model-corroboration implies translational equivalence, though not *vice versa* unless the languages involved are both opinionated. We now verify this point.

Proposition 4.2.32 *(From Corroboration to Translational Equivalence)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages. Suppose $\mathtt{t}, \mathtt{s} \colon \mathbf{L}_1 \infty \mathbf{L}_2$. Then $\mathtt{t}, \mathtt{s} \colon \mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$.

Proof: Since $\mathtt{t}$ and $\mathtt{s}$ are model-corroborating, they are model-preserving, which implies $\mathtt{t}, \mathtt{s} \colon \mathbf{L}_1 \overset{\leftsquigarrow}{\rightsquigarrow} \mathbf{L}_2$ by **Proposition 4.2.11**. So it suffices to show that

$\phi \equiv_1 \mathtt{s}(\mathtt{t}(\phi))$ for all $\phi \in \mathcal{L}_1$ (the proof that $\psi \equiv_2 \mathtt{t}(\mathtt{s}(\psi))$ for $\psi \in \mathcal{L}_2$ is symmetric). Let $x \in C_1$. By **Proposition 4.2.10**, there is a $y \in C_2$ such that $x \propto_{\mathtt{t}} y$. By **Lemma 4.2.30**, we also have $y \propto_{\mathtt{s}} x$. Thus, $x \Vdash_1 \phi$ iff $y \Vdash_2 \mathtt{t}(\phi)$ iff $x \Vdash_1 \mathtt{s}(\mathtt{t}(\phi))$. Hence, $\phi \equiv_1 \mathtt{s}(\mathtt{t}(\phi))$. ■

Lemma 4.2.33 *(Translational Equivalence Preserves Opinionatedness in Images)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be opinionated, and let $\mathtt{t}, \mathtt{s} \colon \mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$. Then $\mathtt{t}[\mathbf{L}_1]$ and $\mathtt{s}[\mathbf{L}_2]$ are opinionated.

Proof: We will just show $\mathtt{t}[\mathbf{L}_1]$ is opinionated. Suppose for *reductio* that $\mathtt{Diag}_{\mathtt{t}[1]}(y) \subset \mathtt{Diag}_{\mathtt{t}[1]}(y')$ for some $y, y' \in C_2$. That means there is a $\mathtt{t}(\phi) \in \mathtt{Diag}_{\mathtt{t}[1]}(y') - \mathtt{Diag}_{\mathtt{t}[1]}(y)$, and thus, $\mathtt{t}(\phi) \in \mathtt{Diag}_2(y') - \mathtt{Diag}_2(y)$. That in turn implies $\mathtt{Diag}_2(y) \neq \mathtt{Diag}_2(y')$. Moreover:

$$\begin{aligned}
\psi \in \mathtt{Diag}_2(y) \quad &\Leftrightarrow \quad \mathtt{t}(\mathtt{s}(\psi)) \in \mathtt{Diag}_2(y) \\
&\Leftrightarrow \quad \mathtt{t}(\mathtt{s}(\psi)) \in \mathtt{Diag}_{\mathtt{t}[1]}(y) \\
&\Rightarrow \quad \mathtt{t}(\mathtt{s}(\psi)) \in \mathtt{Diag}_{\mathtt{t}[1]}(y') \\
&\Leftrightarrow \quad \mathtt{t}(\mathtt{s}(\psi)) \in \mathtt{Diag}_2(y') \\
&\Leftrightarrow \quad \psi \in \mathtt{Diag}_2(y').
\end{aligned}$$

So $\mathtt{Diag}_2(y) \subset \mathtt{Diag}_2(y')$, $\lightning$. ■

Proposition 4.2.34 *(From Translational Equivalence to Corroboration)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}, \mathtt{s} \colon \mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$. Suppose $\mathbf{L}_1$ and $\mathbf{L}_2$ are opinionated. Then $\mathtt{t}, \mathtt{s} \colon \mathbf{L}_1 \propto \mathbf{L}_2$.

Proof: The fact that $\mathtt{t}$ and $\mathtt{s}$ are model-preserving follows from **Proposition 4.2.15** and **Lemma 4.2.33**. Thus, by **Proposition 4.2.31**, it suffices to show the following:

(i′) for all $x, x' \in C_1$ and all $y \in C_2$, if $x \propto_{\mathtt{t}} y \propto_{\mathtt{s}} x'$, then $x \equiv_1 x'$

(ii′) for all $y, y' \in C_2$ and all $x \in C_1$, if $y \propto_{\mathtt{s}} x \propto_{\mathtt{t}} y'$, then $y \equiv_2 y'$.

For (i′), let $x, x' \in C_1$ and $y \in C_2$ be such that $x \propto_{\mathtt{t}} y \propto_{\mathtt{s}} x'$. Then $x \Vdash_1 \phi$ iff $y \Vdash_2 \mathtt{t}(\phi)$ iff $x' \Vdash_1 \mathtt{s}(\mathtt{t}(\phi))$ iff $x' \Vdash_1 \phi$. So $x \equiv_1 x'$. Similarly for (ii′). ■

And again, just as model-preservability collapses to translatability for canonical languages, model-corroboration collapses to translational equivalence for canonical languages.

Proposition 4.2.35 *(Model-Corroboration Collapses to Translational Equivalence for Canonical Languages)*. Let L_1 and L_2 be logics, let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$, and let $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. Then $\mathtt{t}, \mathtt{s}\colon \mathbf{Can}(\mathsf{L}_1) \propto \mathbf{Can}(\mathsf{L}_2)$ iff $\mathtt{t}, \mathtt{s}\colon \mathbf{Can}(\mathsf{L}_1) \leftrightsquigarrow \mathbf{Can}(\mathsf{L}_2)$.

Proof: The left-to-right direction is given by **Proposition 4.2.32**. For the right-to-left direction, it suffices to show by **Proposition 4.2.31** that $\propto_{\mathtt{t}} = \propto_{\mathtt{s}}{}^{-1}$. We just show that $\propto_{\mathtt{t}} \subseteq \propto_{\mathtt{s}}{}^{-1}$. Suppose $\Gamma \propto_{\mathtt{t}} \Delta$. Thus, $\phi \in \Gamma$ iff $\mathtt{t}(\phi) \in \Delta$. But then $\psi \in \Delta$ iff $\mathtt{t}(\mathtt{s}(\psi)) \in \Delta$ iff $\mathtt{s}(\psi) \in \Gamma$. So $\Delta \propto_{\mathtt{s}} \Gamma$. ■

Fact 4.2.36 *(Translational Equivalence Elsewhere)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages, let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$, and let $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. Then we have $\mathtt{t}, \mathtt{s}\colon \mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$ iff $\mathtt{t}, \mathtt{s}\colon \mathbf{Can}(\mathsf{L}_1) \leftrightsquigarrow \mathbf{Can}(\mathsf{L}_2)$.

Corollary 4.2.37 *(Translational Equivalence Implies Model-Corroboration Elsewhere)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$, and let $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. Then $\mathtt{t}, \mathtt{s}\colon \mathbf{L}_1 \leftrightsquigarrow \mathbf{L}_2$ iff $\mathtt{t}, \mathtt{s}\colon \mathbf{Can}(\mathsf{L}_1) \propto \mathbf{Can}(\mathsf{L}_2)$.

§ 4.2.4 Value-Preservation

For a translation to be adequate, it should not only preserve the inferential relations between sentences, but it should also preserve meaning. When two languages do not share the same concept of logical space, it is difficult to determine how to compare the meaning of a sentence in one language to that of a sentence in the other language. But perhaps we could still talk about ways in which the semantic values of the source language are mirrored in the target language.

Definition 4.2.38 *(Value-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages. A map $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is ***value-preserving*** if for all $\phi \in \mathcal{L}_1$:

(i) $\propto_{\mathtt{t}} [\llbracket \phi \rrbracket_1] = \llbracket \mathtt{t}(\phi) \rrbracket_2$

(ii) $\propto_{\mathtt{t}}{}^{-1}[\llbracket \mathtt{t}(\phi) \rrbracket_2] = \llbracket \phi \rrbracket_1$.

Fact 4.2.39 *(Maps are Always Half Value-Preserving)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then for all $\phi \in \mathcal{L}_1$:

(a) $\propto_{\mathtt{t}} [\llbracket \phi \rrbracket_1] \subseteq \llbracket \mathtt{t}(\phi) \rrbracket_2$

(b) $\propto_{\mathtt{t}}{}^{-1}[\llbracket \mathtt{t}(\phi) \rrbracket_2] \subseteq \llbracket \phi \rrbracket_1$.

Proposition 4.2.40 *(Alternative Formulation of Value-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}$ is value-preserving iff for all $\phi \in \mathcal{L}_1$:

(i) $[\![\phi]\!]_1 = {\propto_{\mathtt{t}}}^{-1}[\propto_{\mathtt{t}} [[\![\phi]\!]_1]]$.

(ii) $[\![\mathtt{t}(\phi)]\!]_2 = \propto_{\mathtt{t}} [{\propto_{\mathtt{t}}}^{-1}[[\![\mathtt{t}(\phi)]\!]_2]]$.

Proof: The left-to-right direction is trivial. For the right-to-left direction, by **Fact 4.2.39**, $\propto_{\mathtt{t}} [[\![\phi]\!]_1] \subseteq [\![\mathtt{t}(\phi)]\!]_2$. Conversely, since ${\propto_{\mathtt{t}}}^{-1}[[\![\mathtt{t}(\phi)]\!]_2] \subseteq [\![\phi]\!]_1$:

$$\begin{aligned} [\![\mathtt{t}(\phi)]\!]_2 &= \propto_{\mathtt{t}} [{\propto_{\mathtt{t}}}^{-1}[[\![\mathtt{t}(\phi)]\!]_2]] \\ &\subseteq \propto_{\mathtt{t}} [[\![\phi]\!]_1]. \end{aligned}$$

This completes the proof. ■

Epstein [1990, pp. 390–392] argues that in order for a translation to be meaning-preserving, it ought to be model-preserving. There is a sense in which we can verify this conviction if we understand meaning as semantic value.

Proposition 4.2.41 *(Model-Preservation Implies Value-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$. Then $\mathtt{t}$ is value-preserving.

Proof: If $y \in [\![\mathtt{t}(\phi)]\!]_2$, there is a $x \in \mathsf{C}_2$ such that $x \propto_{\mathtt{t}} y$, and thus, $x \Vdash_1 \phi$. Hence, $y \in \propto_{\mathtt{t}} [[\![\phi]\!]_1]$. Likewise, if $x \in [\![\phi]\!]_1$, then $x \in {\propto_{\mathtt{t}}}^{-1}[[\![\mathtt{t}(\phi)]\!]_2]$. ■

Thus, when $\mathtt{t}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$ is model-preserving, it makes sense to talk about "the" semantic value of an $\mathcal{L}_1$-formula within $\mathbf{L}_2$. The converse holds as long as we ignore "empty" states (**Definition 1.2.5**).

Fact 4.2.42 *(Model-Preservation Implies Agreement on Empty States)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$. Then $\mathbf{L}_1$ has empty states iff $\mathtt{t}[\mathbf{L}_1]$ does too.

Proposition 4.2.43 *(Value-Preservation Almost Implies Model-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$.

(a) Suppose $\propto_{\mathtt{t}} [[\![\phi]\!]_1] = [\![\mathtt{t}(\phi)]\!]_2$ for all $\phi \in \mathcal{L}_1$ and suppose that $\mathtt{t}[\mathbf{L}_1]$ has an empty state only if $\mathbf{L}_1$ does. Then $\propto_{\mathtt{t}}$ is surjective.

(b) Suppose $\propto_{\mathtt{t}}{}^{-1}[[\![\mathtt{t}(\phi)]\!]_2] = [\![\phi]\!]_1$ for all $\phi \in \mathcal{L}_1$ and suppose that $\mathbf{L}_1$ has an empty state only if $\mathtt{t}[\mathbf{L}_1]$ does. Then $\propto_{\mathtt{t}}$ is total.

(c) Suppose $\mathtt{t}$ is value-preserving and $\mathbf{L}_1$ has empty states iff $\mathtt{t}[\mathbf{L}_1]$ does. Then $\mathtt{t}$ is model-preserving.

Proof: We simply show (a), since the proof of (b) is symmetric and (c) follows from (a) and (b) via **Proposition 4.2.10**. Let $y \in \mathsf{C}_2$. If $\mathtt{Diag}_{\mathtt{t}[1]}(y)$ is empty, then there is an $x \in \mathsf{C}_1$ such that $\mathtt{Diag}_1(x)$ is empty, and so $x \propto_{\mathtt{t}} y$. Otherwise, let $y \Vdash_2 \mathtt{t}(\phi)$. Then $y \in [\![\mathtt{t}(\phi)]\!]_2 = \propto_{\mathtt{t}} [[\![\phi]\!]_1]$, in which case there is a $x \in \mathsf{C}_1$ such that $x \propto_{\mathtt{t}} y$. So $\propto_{\mathtt{t}}$ is surjective. ∎

Corollary 4.2.44 *(From Value-Preservation to Model-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Suppose $\mathbf{L}_1$ and $\mathtt{t}[\mathbf{L}_1]$ are opinionated. Then if $\mathtt{t}$ is value-preserving, it is also model-preserving.

We cannot remove the condition that $\mathbf{L}_1$ and $\mathtt{t}[\mathbf{L}_1]$ agree on there being empty states in **Proposition 4.2.43**. As noted in **Fact 4.2.42**, a model-preserving map can exist between $\mathbf{L}_1$ and $\mathbf{L}_2$ only if $\mathbf{L}_1$ and $\mathtt{t}[\mathbf{L}_1]$ agree on whether there are empty states. By contrast, a map can still be value-preserving without even being a translation if the two disagree on empty states.

Example 4.2.45 *(Value-Preservability without Translatability)*. Take any $\mathbf{L}$ with some validities and add an empty state to it to obtain $\mathbf{L}_0$. Then the identity map $\mathtt{id}$ on $\mathcal{L}$ is value-preserving, but not a translation.

Model-preservation requires agreement on empty states. Value-preservation does not. However, this seems to be the only significant difference between the two notions. This suggests that there is a more general notion that might be able to secure the match with model-preservation. Indeed, that is the case:

Definition 4.2.46 *(Strong Value-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages. A map $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ is ***strongly value-preserving*** if for all $\Gamma \subseteq \mathcal{L}_1$:

(i) $\propto_{\mathtt{t}} [[\![\Gamma]\!]_1] = [\![\mathtt{t}[\Gamma]]\!]_2$

(ii) $\propto_{\mathtt{t}}{}^{-1}[[\![\mathtt{t}[\Gamma]]\!]_2] = [\![\Gamma]\!]_1$.

Proposition 4.2.47 *(Strong Value-Preservation is Model-Preservation)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. Then $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$ iff $\mathtt{t}$ is strongly value-preserving.

Proof: The left-to-right direction is exactly analogous to **Proposition 4.2.41**. For the right-to-left direction, $\propto_{\mathtt{t}} [\mathsf{C}_1] = \propto_{\mathtt{t}} [[\![\varnothing]\!]_1] = [\![\mathtt{t}[\varnothing]]\!]_2 = [\![\varnothing]\!]_2 = \mathsf{C}_2$. Hence, $\propto_{\mathtt{t}}$ is surjective. Likewise for totality. ■

Corollary 4.2.48 *(Semantic Formulations of Translatability)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$. If $\mathbf{L}_1$ and $\mathtt{t}[\mathbf{L}_1]$ are opinionated, then the following are equivalent:

(a) $\mathtt{t}\colon \mathbf{L}_1 \rightsquigarrow \mathbf{L}_2$.

(b) $\mathtt{t}\colon \mathbf{L}_1 \propto \mathbf{L}_2$.

(c) $\mathtt{t}$ is value-preserving.

(d) $\mathtt{t}$ is strongly value-preserving.

§ 4.2.5 Value-Corroboration

As before, one may ask whether there is a nice semantic-value correlate of translational equivalence, given that value-preservation is the rough semantic-value correlate of translation. Again, the answer is affirmative. The key essentially invokes **Lemma 4.2.30**.

Definition 4.2.49 *(Value-Corroboration)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. We say $\mathtt{t}$ and $\mathtt{s}$ are ***value-corroborating*** if for all $\phi \in \mathcal{L}_1$ and $\psi \in \mathcal{L}_2$:

(i) $[\![\phi]\!]_1 = [\![\mathtt{s}(\mathtt{t}(\phi))]\!]_1 = {\propto_{\mathtt{t}}}^{-1}[{\propto_{\mathtt{s}}}^{-1}[[\![\phi]\!]_1]]$.

(ii) $[\![\psi]\!]_2 = [\![\mathtt{t}(\mathtt{s}(\psi))]\!]_2 = {\propto_{\mathtt{s}}}^{-1}[{\propto_{\mathtt{t}}}^{-1}[[\![\psi]\!]_2]]$.

Proposition 4.2.50 *(Alternative Formulation of Value-Corroboration)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. Then $\mathtt{t}$ and $\mathtt{s}$ are value-corroborating iff $\mathtt{t}$ and $\mathtt{s}$ are both value-preserving and $\propto_{\mathtt{s}} = {\propto_{\mathtt{t}}}^{-1}$.

Proof:

(⇒) Suppose $\mathtt{t}$ and $\mathtt{s}$ are value-corroborating. It suffices to show that $\propto_{\mathtt{s}} = {\propto_{\mathtt{t}}}^{-1}$ by **Proposition 4.2.40**. Suppose $x \propto_{\mathtt{t}} y$. Then for all $\phi \in \mathcal{L}_1$, $x \Vdash_1 \phi$ iff $y \Vdash_2 \mathtt{t}(\phi)$. So if $\psi \in \mathcal{L}_2$, $y \Vdash_2 \psi$ iff $y \Vdash_2 \mathtt{t}(\mathtt{s}(\psi))$ iff $x \Vdash_1 \mathtt{s}(\psi)$. Hence, $y \propto_{\mathtt{s}} x$. Likewise for the converse.

(⇐) Suppose $\mathtt{t}$ and $\mathtt{s}$ are both value-preserving and $\propto_{\mathtt{s}} = {\propto_{\mathtt{t}}}^{-1}$. It suffices to show that $[\![\phi]\!]_1 = [\![\mathtt{s}(\mathtt{t}(\phi))]\!]_1$ for all $\phi \in \mathcal{L}_1$ (showing (ii) is similar). For any $\phi \in \mathcal{L}_1$ and any $x \in \mathsf{C}_1$:

$$\begin{array}{rcll} x \Vdash_1 \phi & \Leftrightarrow & \exists y \in \mathsf{C}_2\colon\ x \propto_{\mathtt{t}} y \ \&\ y \Vdash_2 \mathtt{t}(\phi) & \text{(LTR: totality)} \\ & \Leftrightarrow & \exists y \in \mathsf{C}_2\colon\ y \propto_{\mathtt{s}} x \ \&\ y \Vdash_2 \mathtt{t}(\phi) & \\ & \Leftrightarrow & x \Vdash_1 \mathtt{s}(\mathtt{t}(\phi)) & \text{(RTL: surjectivity).} \end{array}$$

This completes the proof. ∎

Just as there is a "strong" version of value-preservation, there is a correspondingly "strong" version of value-corroboration that behaves as one would expect by analogy with value-preservation.

Definition 4.2.51 *(Strong Value-Corroboration)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. We say that $\mathtt{t}$ and $\mathtt{s}$ are ***strongly value-corroborating*** if for all $\Gamma \subseteq \mathcal{L}_1$ and $\Delta \subseteq \mathcal{L}_2$:

(i) $[\![\Gamma]\!]_1 = [\![\mathtt{s}[\mathtt{t}[\Gamma]]]\!]_1 = {\propto_{\mathtt{t}}}^{-1}[{\propto_{\mathtt{s}}}^{-1}[[\![\Gamma]\!]_1]]$.

(ii) $[\![\Delta]\!]_2 = [\![\mathtt{t}[\mathtt{s}[\Delta]]]\!]_2 = {\propto_{\mathtt{s}}}^{-1}[{\propto_{\mathtt{t}}}^{-1}[[\![\Delta]\!]_2]]$.

Proposition 4.2.52 *(Alternative Formulation of Strong Value-Corroboration)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages, and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$. Then $\mathtt{t}$ and $\mathtt{s}$ are strongly value-corroborating iff $\mathtt{t}$ and $\mathtt{s}$ are both strongly value-preserving and $\propto_{\mathtt{s}} = {\propto_{\mathtt{t}}}^{-1}$.

Proof: Same as **Proposition 4.2.50**. ∎

Corollary 4.2.53 *(From Model-Corroboration to Value-Corroboration)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be some languages and let $\mathtt{t}, \mathtt{s}\colon \mathbf{L}_1 \infty \mathbf{L}_2$. Then $\mathtt{t}$ and $\mathtt{s}$ are strongly value-corroborating.

Proof: Immediate from **Lemma 4.2.30** and **Propositions 4.2.47** and **4.2.50**. ■

Proposition 4.2.54 *(From Value-Corroboration to Model-Corroboration)*. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be languages and let $\mathtt{t}\colon \mathcal{L}_1 \to \mathcal{L}_2$ and $\mathtt{s}\colon \mathcal{L}_2 \to \mathcal{L}_1$.

(a) If $\mathtt{t}$ and $\mathtt{s}$ are value-corroborating and if $\mathbf{L}_1$ and $\mathbf{L}_2$ agree on there being empty states, then $\mathtt{t}, \mathtt{s}\colon \mathbf{L}_1 \propto \mathbf{L}_2$.

(b) If $\mathtt{t}$ and $\mathtt{s}$ are strongly value-corroborating, then $\mathtt{t}, \mathtt{s}\colon \mathbf{L}_1 \propto \mathbf{L}_2$.

Proof: The proof is similar to that of **Proposition 4.2.34** (part (b) uses reasoning similar to **Proposition 4.2.47**). ■

As one would expect, value-corroborating maps are not guaranteed to even be translations if the two logics do not agree on empty states. In fact, the same counterexample from **Example 4.2.45** illustrating that value-preserving maps need not be translations works to show that value-corroborating maps need not be translations. Still, as before, we get a nice equivalence between translational equivalence, model-corroboration, and (strong) value-preservation when the languages in question are opinionated.

Chapter 5

Metametaphysics

We have all come across disputes that seem impossible to resolve. Philosophy is full of them. Here are some familiar examples of such disputes in several different areas of philosophy:

- Ethics: consequentialism vs. deontology
- Decision theory: causal decision theory vs. evidential decision theory
- Language: absolutism vs. contextualism vs. relativism (about any number of expressions)
- Logic: classical vs. non-classical logic.

At some point, these disputes became standoffs. Each side has found both plenty of arguments in support of their position and plenty of problems for their opposition. But neither side has found a convincing argument that uses only assumptions that all sides of the dispute are willing to grant, and there is a sense that none will ever be found.

There are two common attitudes to have towards apparently irresolvable disputes. On the one hand, there are the ***realists*** about a dispute, who hold that there is an objective fact of the matter as to which side of the dispute (if any) is correct. For the realist, irresolvability arises out of no fault of ours. The question surrounding the dispute is legitimate; it is just really difficult (maybe even impossible) to resolve the dispute in an objective manner. On the other hand, there are the ***anti-realists*** about a dispute, who hold that there is no objective fact of the matter as to which side of the dispute is correct. For the anti-realist, we are to blame for the irresolvability. The project of answering the question being disputed is somehow fundamentally misguided. Misguided not because the question is uninteresting or unimportant (such as the dispute over whether the number of grains of sand in the world is even or odd) but because both sides are making some false presupposition about the question and its answer in carrying out their dispute.

These opposing attitudes have lead philosophers to elevate the debate up a level: rather than engage in the first-order dispute over which side is right, we enter a "metadispute" over whether there is even a fact of the matter as to which side of the first-order dispute is right. And these disputes themselves have started to become standoffs. To be sure, each side faces its own unique challenges. In general, the realist must defend the importance of having these debates (given we cannot know which side is right) and explain why we can be so confident in our judgments about which side is right. By contrast, the anti-realist has to explain why people are led to thinking there is a debate in the first place. But these are obstacles, not barriers: over time, each side has found responses to these challenges that have suited them. Yet no one has found a compelling argument in favor of the realist or the anti-realist that can be made on neutral turf.

Metadisputes of this sort generally concern whether or not the two sides of the first-order dispute are speaking the same language. The realist insists that when one side asserts p and the other side asserts $\neg p$, the two sides of the dispute are using the same language but are asserting incompatible propositions. By contrast, the anti-realist will reject this presupposition, arguing instead that the best account of the dispute is one on which the two sides are merely speaking different languages.

To illustrate this point, we can imagine a first-order dispute over whether or not Pluto is a planet. One disputant asserts "Pluto is a planet" while the other disputant asserts "Pluto is not a planet." A realist interpretation of this debate holds that each disputant is using the term 'planet' in the same way as the other. What the two sides are disagreeing over, then, is whether Pluto actually is a planet in the sense the disputants agree upon. By contrast, an anti-realist interpretation of this debate holds that each side is actually using 'planet' to mean different things. Even though on the surface it looks as though the two sides are disagreeing with one another, really the propositions they assert are compatible.

It is generally easy to figure out on a case-by-case basis whether or not a realist or anti-realist interpretation of a dispute is appropriate. For instance, if the disputants in the debate over Pluto's status as planet were debating whether or not Pluto clears its orbital neighborhood, that would suggest they are both using the term 'planet' as the IAU proposes to use the term. If, instead, they were debating whether clearing its orbital neighborhood was necessary to count as a planet, that would suggest that they are debating how to use the term 'planet'.

In philosophy, however, it is far from clear which interpretation is appropriate. Recently, a whole branch of metaphysics, known as "metametaphysics", has been developed to try to sort out the question of whether we should be realists or anti-realists about certain metaphysical disputes. As of now, there seems to be little consensus as to how to adjudicate metadisputes in a satisfactory way.

A common example metametaphysicians pick on, which we will examine more closely in § 5.1, is the dispute over mereological composition. The question is: are there composite objects (i.e., objects with proper parts)? The common sense answer says that there are: objects like tables, chairs, computers, people, and so forth all have proper parts. A radical answer says that there are not: all there are are fundamental particles arranged in various ways *like* a table, *like* a chair, and so on (some prefer to say "arranged table-wise" in place of "arranged like a table").

What are we to make of the dispute over mereological composition? Must there be a fact of the matter as to whether the common sense answer or the radical answer (if any) is correct? The realist about this dispute would say there is a fact of the matter, even if there are no *a priori* means by which we can come to know which answer (if any) is right. Each side uses terms like "part" and "composite" in the same exact way, but merely disagrees over what is true of these concepts. The anti-realist would say that there is no fact of the matter as to which side is right, as these are just two different ways of talking. That is, the common-senser and the radical do not use the terms "part" and "composite" in the same exact way, and so they are disagreeing over how to use words rather than what reality is like.

I have intentionally used the word "attitude" in describing realism and anti-realism, rather than "theory". The reason is that it is unclear that either side must be objectively right or wrong in their characterization of the first-order dispute. Must there be a fact of the matter as to whether we are using words in the same exact way as another? Could one not be an anti-realist about this metaquestion? After all, it seems like these metadisputes are just as irresolvable as the first-order disputes they are meta-ing. One might not unreasonably conclude, then, that the choice between the realist and anti-realist is a theoretical one: we can choose to be realists or anti-realists about any of these sorts of disputes. The question is not which attitude *is correct*, but rather which attitude *we should take*.

It is commonly thought that anti-realism entails an end to a dispute. Anti-realists about a dispute, it is thought, will hold that there simply is no point in having the dispute in the first place. You can choose to talk however you like. If you want to talk like a common-senser about composition, you may do so. If you want to talk like a radical, you may do so. In choosing a language to speak, you are not thereby misdescribing the world. It is only once you choose a language and start attempting to describe the world that you can be subject to a charge of inaccuracy.

Much of the chapter that follows is inspired by the following, arguably obvious, thought: that the language we choose to speak *does* matter. The choice between speaking a particular language does have bearing on our abilities to describe the world. You may not be speaking inaccurately solely in virtue of speaking the language you speak; but speaking the language you speak can harbor tangible consequences on your ability to thereafter describe the world. Thus, even if you have an anti-realist attitude towards a first-order dispute, it need not follow that that dispute is pointless. A theoretical dispute over which language to speak can be as important as a theoretical dispute over what the world is like.

Considerations of expressive power make this point quite salient. You can choose to speak a propositional language with just conjunction and disjunction, or a propositional language with just conjunction and negation. You are not thereby getting things wrong in choosing to speak the former language, but your ability to describe the world will be significantly hindered by that choice.

This chapter goes beyond this simple point, however. For one thing, in the case of choosing between speaking a propositional language with just conjunction and disjunction or one with just conjunction and negation, it is fairly obvious that (all else being equal) you should choose the latter. The additional expressive power allows you to make genuine distinctions that would otherwise be ineffable in the weaker language. But sometimes, it can be a substantive question whether the addition of expressive power is one that marks a genuine distinction. In § 5.1, we will see that even though some theories about mereological composition are more expressive than others, the increase in expressive power afforded by these theories need not be recognized as legitimate by the others. Thus, an anti-realist can view such a dispute over mereological composition as a dispute over which possibilities one ought to recognize.

What's more, even if two languages each have a sufficient amount of expressive power to interpret one another charitably, it need not follow that the choice between which language to speak is unimportant. We saw in § 2.5 that there is an important difference between expressive equivalence and notational variance. There may be plenty of reasons to argue over which notational variant of a language we should speak (perhaps one is simpler or easier to use, for instance). But even if we assume that disputes over which notational variant to use are uninteresting in the relevant sense, it need not be that disputes over which of several expressively equivalent languages to speak is uninteresting. The fact that two languages might differ in their global structure even if they have the same expressive capacities gives the anti-realist sufficient grounds for carrying out a dispute over which language we ought to speak, as differences in global structure can affect one's ability to extend the language in various ways. We will see in § 5.2 that this is precisely the situation characterizing disputes over persistence.

§ 5.1 Mereology

Mereology is the study of parthood. A central question in mereology concerns the existence of composite objects. Say an object is an ***atom*** or a ***simple*** if it has no proper parts (I will use the terms interchangeably throughout). Say an object is ***composite*** if it is not an atom. Are there any composite objects? According to ***mereological nihilism***, the answer is negative: no object has proper parts. The only things that exist, according to the nihilist, are simples. According to ***mereological universalism***, the answer is affirmative to an extreme: *every* collection of objects has a "mereological fusion", i.e., an object such that anything that overlaps with it overlaps with some object in the collection. So unless there is only one atom, there will in general be lots of composite objects.

Sometimes, philosophers get the sense that there is not really a substantive debate between these different positions. Whenever the universalist, for instance, talks about a composite object (say, a table), the nihilist talks about some simples arranged in a certain way (say, some simples arranged table-wise). These two ways of talking seem completely interchangeable, and the difference, therefore, does not seem very important. Thus, to understand the sense in which this debate is or is not substantive, it is important to explore the translational and interpretational connections between them. But even setting that aside, such an investigation could prove beneficial for getting a clearer picture of the structure of these mereological theories, just as the standard translation from modal logic into first-order logic can help us see more clearly the structure of modal logic. In this section, we will make some first steps towards that investigation.

§ 5.1.1 Defining the Views

Throughout this section, let $\mathcal{L}^1_{\mathtt{Pred}}(\leqslant)$ be the following simple first-order syntax:

$$\phi ::= P^n(y_1, \ldots, y_n) \mid (x = y) \mid (x \leqslant y) \mid \neg\,\phi \mid (\phi \wedge \phi) \mid \forall x\, \phi,$$

where P^n is an n-place predicate and $x, y, y_1, \ldots, y_n$ are variables. As usual, we may drop parentheses for readability. Here, '$\leqslant$' stands for the parthood relation. We use the standard abbreviations for classical first-order logic, including those for restricted quantification (e.g., $\forall x \leqslant y\, \phi = \forall x\ (x \leqslant y \to \phi)$). We also use the following abbreviations throughout:

$$\begin{array}{rcll} (x < y) & := & (x \leqslant y \wedge \neg(y \leqslant x)) & (\text{“}x \text{ is a proper part of } y\text{”}) \\ \mathsf{Atom}(x) & := & \forall y\ (y \leqslant x \to x \leqslant y) & (\text{“}x \text{ is an atom”}) \\ (x \circ y) & := & \exists z\ (z \leqslant x \wedge z \leqslant y) & (\text{“}x \text{ overlaps with } y\text{”}) \\ (x \perp y) & := & \neg(x \circ y) & (\text{“}x \text{ is disjoint from } y\text{”}). \end{array}$$

Moreover, we use $\mathcal{L}^1_{\leqslant}$ for the restriction of $\mathcal{L}^1_{\mathtt{Pred}}(\leqslant)$ that removes all non-mereological non-logical predicates, i.e., the first-order syntax defined as follows:

$$\phi ::= (x = y) \mid (x \leqslant y) \mid \neg\,\phi \mid (\phi \wedge \phi) \mid \forall x\, \phi.$$

Before defining nihilism and universalism explicitly, we need some mereological principles that both sides can agree to. Everyone in this debate agrees on the following mereological principles:

Reflexivity. $\forall x\ (x \leqslant x)$.

Antisymmetry. $\forall x\, \forall y\ (x \leqslant y \wedge y \leqslant x \to x = y)$.

Transitivity. $\forall x\, \forall y\, \forall z\ (x \leqslant y \wedge y \leqslant z \to x \leqslant z)$.

In the next section, we may have to drop antisymmetry to accommodate certain views of persistence. But for now, we simply assume parthood is antisymmetric without further question.

Let M be the class of pairs $\langle \mathcal{M}, g\rangle$ such that $\mathcal{M} = \langle D, I\rangle$ is a first-order model where $I(\leqslant)$ is a partial order and g is a variable assignment on $\mathcal{M}$. Let $\Vdash$ be the usual satisfaction relation for first-order logic. Then we define $\mathbf{M} = \langle \mathcal{L}^1_{\texttt{Pred}}(\leqslant), \mathsf{M}, \Vdash\rangle$ to be the language of core mereology. We likewise define $\mathbf{M}_{\leqslant} = \langle \mathcal{L}^1_{\leqslant}, \mathsf{C}, \Vdash\rangle$ to be the language of pure core mereology. In either case, if $\boldsymbol{Ax}$ is a set of axioms, we will use "$\mathbf{M} + \boldsymbol{Ax}$" for the language obtained by restricting to the class of model-assignment pairs satisfying those axioms. If we define $\mathbf{L} \coloneqq \mathbf{M} + \boldsymbol{Ax}$, then $\mathbf{L}_{\leqslant} \coloneqq \mathbf{M}_{\leqslant} + \boldsymbol{Ax}$.

Universalism and nihilism can essentially be formulated in terms of an axiom (schema) as follows:

Universalism. $\forall \overline{u}\ (\exists x\, \phi(x, \overline{u}) \to \exists y\, \forall z\ (y \circ z \leftrightarrow \exists x\ (\phi(x, \overline{u}) \wedge x \circ z)))$.

Nihilism. $\forall x\, \mathsf{Atom}(x)$.

In words, ***Universalism*** says that whenever some things satisfy ϕ, there is an object that consists of all the ϕs and "nothing more". And of course, ***Nihilism*** says that nothing is composite: the world is full of simples. We will define the languages of nihilism and universalism accordingly:

- $\mathbf{N} = \mathbf{M} +$ ***Nihilism***
- $\mathbf{U} = \mathbf{M} +$ ***Universalism***.

For brevity, we will let U be the class of first-order model-assignment pairs satisfying ***Universalism*** and let N be the class of first-order model-assignment pairs satisfying ***Nihilism***. Observe that U and N are not completely disjoint. ***Universalism*** and ***Nihilism*** can both be true if there is exactly one object. But U and N are otherwise disjoint: $\langle \mathcal{M}, g\rangle \in \mathsf{U} \cap \mathsf{N}$ if *and only if* $|D^{\mathcal{M}}| = 1$.

Here is another way to write ***Universalism*** that will be helpful. Define the following abbreviation (where '$\imath$' is the definite description operator):

$$\sigma x\, \phi(x) \coloneqq\ \imath x \forall y\ (y \circ x \leftrightarrow \exists z\ (\phi(z) \wedge y \circ z))\,.$$

$\sigma x\, \phi(x)$ denotes the "fusion" of all the $\phi(x)$s. Then ***Universalism*** is equivalent to the following schema:

$$\forall \overline{u}\ (\exists x\, \phi(x, \overline{u}) \to \exists y\ (y = \sigma x\, \phi(x, \overline{u})))\,.$$

There are other extensions of $\mathbf{M}$ worth mentioning. Consider the following potential axioms:

Atomicity. $\forall x\, \exists y\ (\mathsf{Atom}(y) \wedge y \leqslant x)$.

Atomlessness. $\forall x\, \exists y\ (y < x)$.

Supplementation. $\forall x\, \forall y\ (\neg(y \leqslant x) \to \exists z\ (z \leqslant y \wedge z \perp x))$.[1]

Let A be the class of model-assignment pairs satisfying ***Atomicity***, $\tilde{\mathrm{A}}$ the class of model-assignment pairs satisfying ***Atomlessness***, and S the class of model-assignment pairs satisfying ***Supplementation***. We define the following mereological theories as restrictions of ones previously defined:[2]

- **EM** = **M** + ***Supplementation*** ("extensional mereology")
- **GEM** = **U** + ***Supplementation*** ("general extensional mereology")
- **AGEM** = **GEM** + ***Atomicity*** ("atomic general extensional mereology")
- $\tilde{\mathbf{A}}$**GEM** = **GEM** + ***Atomlessness*** ("atomless general extensional mereology")

§ 5.1.2 Trivial Translation

Given **Theorem 2.6.5**, there is a (nearly trivial) sense in which universalism and nihilism are equivalent:

Proposition 5.1.1 *(Translational Equivalence Between Mereological Theories)*. $\mathbf{N} \cong \mathbf{U}$ (and also $\mathbf{N}_{\leqslant} \cong \mathbf{U}_{\leqslant}$).

Proof (Sketch): The Lindenbaum-Tarski algebras of **N** and **U** are isomorphic, and each only has countably many formulas. So by **Theorem 2.6.5**, $\mathbf{N} \cong \mathbf{U}$. A similar proof works to show that $\mathbf{N}_{\leqslant} \cong \mathbf{U}_{\leqslant}$.[a] ■

[a]The Lindenbaum-Tarski algebras of $\mathbf{N}_{\leqslant}$ and $\mathbf{U}_{\leqslant}$ are not atomless. In fact, each has exactly one atom, which has as a member the formula $\forall x\, \forall y\ (x = y)$. But the proof that any two countable atomless Boolean algebras are isomorphic can be extended relatively easily to show that the Lindenbaum-Tarksi algebras of these languages are isomorphic.

There are a number of reasons why the existence of such a translation is unsatisfying, however. For one thing, **N** and **U** are equivalent only in the same sense that first-order logic and propositional logic are equivalent. Thus, the sense of equivalence here is arguably too weak to be of interest for metaphysical endeavors. More to the point, this translation scheme is not guaranteed to be schematic in general. In this sense, it does not seem to really be a *usable* translation scheme. Finally, whatever this translation looks like, it arguably will not capture important features of the debate between universalism and nihilism.

[1]This is essentially what Jech [2003, p. 204] calls "separativity". There are at least two versions of supplementation discussed in mereology, viz., a strong version and a weak version. This is the strong version. The weak version is as follows:

Weak Supplementation. $\forall x\, \forall y\ (x < y \to \exists z\ (z \leqslant y \wedge z \perp x))$.

Adding ***Weak Supplementation*** to **M** yields what is known as "minimal mereology".

[2]These conventions follow those from Varzi 2016.

Thus, we assume throughout that for universalism and nihilism to be deemed equivalent, there must be at least a schematic equivalence between them. But what kind of equivalence? For now, I leave this question open. Arguably, to show that the two sides are equivalent, we need to at the very least establish that they are schematically mutually interpretable. Moreover, if we establish that they are schematically translationally equivalent, that is as close as we can get to showing that they are equivalent. But I do not want to take a stand on what is both necessary and sufficient to establish equivalence. Rather, I just want to explore the more concrete/precise questions about what senses of equivalence hold between these theories. As we will see, the answer seems to be very few.

§ 5.1.3 The "Barthood" Strategy

For reasons I will return to in a moment, I will start by making a fairly obvious point. There is a very cheap way to interpret **U** in **N** and *vice versa*: take the other side as not talking about the *parthood* relation but rather some other relation—call it "barthood". Essentially, one can do this by simply reinterpreting the other predicates of the language systematically so that we free up a predicate to be interpreted exactly as the other side would have parthood be interpreted. This is proven more precisely below (I skip the inductive steps for brevity).[3]

Proposition 5.1.2 *(The Barthood Interpretation from Nihilism into Universalism)*. There is a schematic $\mathtt{t}\colon \mathbf{N} \propto_i \mathbf{U}$.

Proof: Define $\mathtt{t}$ as follows:

$$\begin{array}{rcl}
\mathtt{t}(P^n(y_1,\ldots,y_n)) & = & P^n(y_1,\ldots,y_n) \quad \text{for } n \neq 2 \\
\mathtt{t}(P^2_i(x,y)) & = & P^2_{i+1}(x,y) \\
\mathtt{t}(x = y) & = & x = y \\
\mathtt{t}(x \leqslant y) & = & P^2_0(x,y) \\
\mathtt{t}(\neg\,\phi) & = & \neg\,\mathtt{t}(\phi) \\
\mathtt{t}(\phi \wedge \psi) & = & \mathtt{t}(\phi) \wedge \mathtt{t}(\psi) \\
\mathtt{t}(\forall x\, \phi) & = & \forall x\ (\mathsf{Atom}(x) \rightarrow \mathtt{t}(\phi))\,.
\end{array}$$

We define the following abbreviations for convenience:

$$\begin{array}{rcl}
B(x,y) & := & P^2_0(x,y) \\
\mathsf{Atom}^B(x) & := & \forall y\ (B(y,x) \rightarrow B(x,y))\,.
\end{array}$$

[3]The results below hold even if we replace **U** with **AGEM**. The proofs can also be modified to replace **U** with $\tilde{\mathbf{A}}$**GEM**.

Define Γ to consist of the following formulas:

$$\begin{array}{rl} & \mathsf{Atom}(x), \quad \text{where } x \text{ is any variable} \\ \forall x & (\mathsf{Atom}(x) \to B(x,x)) \\ \forall x\,\forall y & (\mathsf{Atom}(x) \wedge \mathsf{Atom}(y) \wedge B(x,y) \wedge B(y,x) \to x = y) \\ \forall x\,\forall y\,\forall z & (\mathsf{Atom}(x) \wedge \mathsf{Atom}(y) \wedge \mathsf{Atom}(z) \wedge B(x,y) \wedge B(y,z) \to B(x,z)) \\ \forall x & (\mathsf{Atom}(x) \to \mathsf{Atom}^B(x))\,. \end{array}$$

We will show $\propto_{\mathtt{t}}$ is total and surjective with respect to $[\![\Gamma]\!]_{\mathbf{U}}$.

First, let $\langle \mathcal{N}, h \rangle \in \mathsf{N}$. Define $h^{\mathsf{U}}(x) = \{h(x)\}$ and define $\mathcal{N}^{\mathsf{U}} = \langle D^{\mathsf{U}}, I^{\mathsf{U}} \rangle$, where:

$$\begin{array}{rcl} D^{\mathsf{U}} & = & \wp(D) - \{\varnothing\} \\ I^{\mathsf{U}}(\leqslant) & = & \{\langle A, B \rangle \mid \varnothing \neq A \subseteq B \subseteq D\} \\ I^{\mathsf{U}}(B) & = & I(\leqslant) \\ I^{\mathsf{U}}(P^2_{i+1}) & = & I(P^2_i) \\ I^{\mathsf{U}}(P^n) & = & I(P^n) \quad \text{for } n \neq 2. \end{array}$$

Then by induction, $\mathcal{N}, h \Vdash_{\mathbf{N}} \phi$ iff $\mathcal{N}^{\mathsf{U}}, h^{\mathsf{U}} \Vdash_{\mathbf{U}} \mathtt{t}(\phi)$. Moreover, $\mathcal{N}^{\mathsf{U}}, h^{\mathsf{U}} \Vdash_{\mathbf{U}} \Gamma$.

Next, let $\langle \mathcal{M}, g \rangle \in [\![\Gamma]\!]_{\mathbf{U}}$. Define $\mathcal{M}^{\mathsf{N}} = \langle D^{\mathsf{N}}, I^{\mathsf{N}} \rangle$, where:

$$\begin{array}{rcl} D^{\mathsf{N}} & = & \{a \in D \mid \mathcal{M}, g^x_a \Vdash_{\mathbf{U}} \mathsf{Atom}(x)\} \\ I^{\mathsf{N}}(\leqslant) & = & I(B) \\ I^{\mathsf{N}}(P^2_i) & = & I(P^2_{i+1}) \cap (D^{\mathsf{N}})^2 \quad \text{for } i > 0 \\ I^{\mathsf{N}}(P^n) & = & I(P^n) \cap (D^{\mathsf{N}})^n \quad \text{for } n \neq 2. \end{array}$$

Then by induction, if $\mathcal{M}, g \Vdash_{\mathbf{U}} \Gamma$, then $\mathcal{M}, g \Vdash_{\mathbf{U}} \mathtt{t}(\phi)$ iff $\mathcal{M}^{\mathsf{N}}, g \Vdash_{\mathbf{N}} \phi$. ■

Proposition 5.1.3 *(The Barthood Interpretation from Universalism into Nihilism).* There is a schematic $\mathtt{t}\colon \mathbf{U} \propto_i \mathbf{N}$.

Proof: Define s as follows:

$$\begin{array}{rcl} \mathsf{s}(P^n(y_1, \dots, y_n)) & = & P^n(y_1, \dots, y_n) \quad \text{for } n \neq 2 \\ \mathsf{s}(P^2_i(x,y)) & = & P^2_{i+1}(x,y) \\ \mathsf{s}(x = y) & = & x = y \end{array}$$

$$\begin{aligned}
\mathsf{s}(x \leqslant y) &= P^2_0(x, y) \\
\mathsf{s}(\neg\, \phi) &= \neg\, \mathsf{s}(\phi) \\
\mathsf{s}(\phi \wedge \psi) &= \mathsf{s}(\phi) \wedge \mathsf{s}(\psi) \\
\mathsf{s}(\forall x\, \phi) &= \forall x\, \mathsf{s}(\phi).
\end{aligned}$$

We define the following abbreviations for convenience:

$$\begin{aligned}
B(x, y) &:= P^2_0(x, y) \\
\mathsf{Atom}^B(x) &:= \forall y\ (B(y, x) \to B(x, y)) \\
x \circ^B y &:= \exists z\ (B(z, x) \wedge B(z, y))\,.
\end{aligned}$$

Define Δ to consist of the following formulas:

$$\begin{gathered}
\forall x\, B(x, x) \\
\forall x\, \forall y\ (B(x, y) \wedge B(y, x) \to x = y) \\
\forall x\, \forall y\, \forall z\ (B(x, y) \wedge B(y, z) \to B(x, z)) \\
\forall \overline{u}\ (\exists x\, \mathsf{s}(\phi(x, \overline{u})) \to \exists y\, \forall z\ \left(y \circ^B z \leftrightarrow \exists x\ \left(\mathsf{s}(\phi(x, \overline{u})) \wedge x \circ^B z\right)\right)).
\end{gathered}$$

We now show s is total and surjective with respect to $[\![\Delta]\!]_{\mathbf{N}}$.

First, totality. Let $\langle \mathcal{M}, g\rangle \in \mathsf{U}$. Define $\mathcal{M}^{\mathsf{N}} = \langle D^{\mathsf{N}}, I^{\mathsf{N}}\rangle$, where:

$$\begin{aligned}
D^{\mathsf{N}} &= D \\
I^{\mathsf{N}}(\leqslant) &= \{\langle a, a\rangle \mid a \in D\} \\
I^{\mathsf{N}}(B) &= I(\leqslant) \\
I^{\mathsf{N}}(P^2_{i+1}) &= I(P^2_i) \\
I^{\mathsf{N}}(P^n) &= I(P^n) \quad \text{for } n \neq 2.
\end{aligned}$$

By induction, $\mathcal{M}, g \Vdash_{\mathbf{U}} \phi$ iff $\mathcal{M}^{\mathsf{N}}, g \Vdash_{\mathbf{N}} \mathsf{s}(\phi)$. Moreover, $\mathcal{M}^{\mathsf{N}}, g \Vdash_{\mathbf{N}} \Delta$.

Next, surjectivity. Let $\langle \mathcal{N}, h\rangle \in \mathsf{N}$. Define $\mathcal{N}^{\mathsf{U}} = \langle D^{\mathsf{U}}, I^{\mathsf{U}}\rangle$, where:

$$\begin{aligned}
D^{\mathsf{U}} &= D \\
I^{\mathsf{U}}(\leqslant) &= I(B) \\
I^{\mathsf{U}}(P^2_i) &= I(P^2_{i+1}) \\
I^{\mathsf{U}}(P^n) &= I(P^n) \quad \text{for } n \neq 2.
\end{aligned}$$

Then by induction, if $\mathcal{N}, h \Vdash_{\mathbf{N}} \Delta$, then $\mathcal{N}, h \Vdash_{\mathbf{N}} \mathsf{s}(\phi)$ iff $\mathcal{N}^{\mathsf{U}}, h \Vdash_{\mathbf{U}} \phi$. ■

The situation, then, is that each side can, in principle, interpret the other side completely as talking about a relation other than the parthood relation. Insofar as an interpretation exists in both directions, this is worth noting. However, I do not think this really shows very much.

For one thing, this seems to be a completely general sort of strategy for reinterpreting any philosophical debate. If whenever someone has a debate over the rules governing, say, some property P, one can (mis)interpret them as really talking about a different property P'. So if this was enough to show that the universalism-nihilism debate was merely verbal, it would prove too much.

Moreover, this kind of interpretation does not really seem to respect either side of the debate. It is not just that the universalist and nihilist take themselves to be talking about the same kind of relation—they could, after all, be mistaken about this. It is that it does not really respect how, for instance, nihilists in practice try to reinterpret claims apparently involving composite objects. They do not say, "Oh, when you say there are composite objects, you must mean there are combosite objects instead (i.e., objects with proper barts)." Rather, they try to reinterpret these claims in terms of collections of objects arranged composite-wise. Similarly, universalists do not simply say "Oh, when you say there are only simples, you must mean there are only simple simbles (i.e., objects with no proper barts)." Rather, they interpret the nihilist as having restricted quantification.

There is another odd feature of these interpretations. Suppose we fix on a model whose domain partly consisting of what the universalist would call composite objects (i.e., a member of U). According to the barthood interpretation above, the corresponding nihilist model is one whose domain is the same. But the nihilist only admits of simples according to their preferred relation of parthood. So according to this interpretation, the nihilist would deem as simples all the objects such as tables, planets, etc. that are ordinarily thought of as composite. While this view is compatible with nihilism, it does not seem to be the one nihilists themselves adopt. It is not that they deny that tables and chairs have no parts; it is that they deny that there *are* any tables to begin with.

§ 5.1.4 Restricted Quantification

Given how unsatisfying the barthood strategy is, it is natural to ask whether there are other interpretational strategies available to either side. To answer this question, it helps to shift the discussion from $\mathbf{N}$ and $\mathbf{U}$ to $\mathbf{N}_{\leqslant}$ and $\mathbf{U}_{\leqslant}$, which do not have access to non-logical predicates other than '$\leqslant$' to reinterpret the parthood relation of the other side.[4]

In the case of universalism, there is a very natural way to interpret the nihilist in terms of restricted quantification:

[4]The translations below can all be easily modified to allow non-logical predicates back in.

Proposition 5.1.4 *(Pure Interpretation of Nihilism into Universalism)*. There is a schematic $\mathtt{t}\colon \mathbf{N}_{\leqslant} \propto_i \mathbf{U}_{\leqslant}$.

Proof: Define $\mathtt{t}$ as follows:

$$
\begin{aligned}
\mathtt{t}(x = y) &= x = y \\
\mathtt{t}(x \leqslant y) &= x = y \\
\mathtt{t}(\neg\,\phi) &= \neg\,\mathtt{t}(\phi) \\
\mathtt{t}(\phi \wedge \psi) &= \mathtt{t}(\phi) \wedge \mathtt{t}(\psi) \\
\mathtt{t}(\forall x\,\phi) &= \forall x\ (\mathsf{Atom}(x) \to \mathtt{t}(\phi))\,.
\end{aligned}
$$

Define $\Gamma := \{\mathsf{Atom}(x) \mid x \text{ is a variable}\}$. Then the proof is almost exactly as before. ∎

In other words, the universalist is free to interpret the nihilist as employing a restrictive way of speaking. Even when the nihilist bangs their fist on the table, saying "No, when I say 'everything' to range over *everything*", the universalist can still interpret '*everything*' as ranging only over atoms. Then the universalist can point out that their theory is more general, since they can (a) allow names (here, free variables) to range over composite objects and (b) talk about objects that are not composed entirely (or even partly) of atoms (assuming the universalist rejects full atomicity).

§ 5.1.5 Limitations on Nihilist Expressivity

On the other hand, it is less than clear that nihilism can do the same. For one thing, we will definitely not find a $\mathtt{s}$ that produces a coherent translation scheme with $\mathtt{t}$:

Proposition 5.1.5 *(The Irreversibility of the Pure Interpretation of Nihilism into Universalism)*. Define $\mathtt{t}$ and Γ as in **Proposition 5.1.4**. Then there is no $\Delta \subseteq \mathcal{L}^1_{\leqslant}$ and $\mathtt{s}\colon \mathcal{L}^1_{\leqslant} \to \mathcal{L}^1_{\leqslant}$ such that $\mathtt{t},\mathtt{s}\colon \mathbf{N}_{\leqslant} \leftrightsquigarrow_{\Gamma,\Delta} \mathbf{U}_{\leqslant}$.

Proof: If there were such a $\mathtt{s}$, then $\vDash_{\mathbf{U}_{\leqslant}} \Gamma$, since:

$$
\begin{aligned}
\vDash_{\mathbf{U}_{\leqslant}} \Gamma &\Leftrightarrow \Delta \vDash_{\mathbf{N}_{\leqslant}} \mathtt{s}[\Gamma] \\
&\Leftrightarrow \Gamma, \mathtt{t}[\Delta] \vDash_{\mathbf{U}_{\leqslant}} \mathtt{t}[\mathtt{s}[\Gamma]] \\
&\Leftrightarrow \Gamma, \mathtt{t}[\Delta] \vDash_{\mathbf{U}_{\leqslant}} \Gamma.
\end{aligned}
$$

But clearly $\nvDash_{\mathbf{U}_{\leqslant}} \mathsf{Atom}(x)$. ∎

But there is a deeper concern regarding the prospects of adequately interpreting universalism into nihilism. The following observation follows from a basic fact from model theory:

Fact 5.1.6 *(Nihilistic Equivalence)*. For all $\langle \mathcal{N}_1, h_1 \rangle, \langle \mathcal{N}_2, h_2 \rangle \in \mathsf{N}$:

$$\mathcal{N}_1 \equiv_{\mathbf{N}_\leqslant} \mathcal{N}_2 \quad \Leftrightarrow \quad \text{either } |D_1| = |D_2| < \aleph_0 \text{ or both } |D_1| \geqslant \aleph_0 \text{ and } |D_2| \geqslant \aleph_0.$$

Moreover:

$$\mathcal{N}_1, h_1 \equiv_{\mathbf{N}_\leqslant} \mathcal{N}_2, h_2 \quad \Leftrightarrow \quad \begin{array}{l} \mathcal{N}_1 \equiv_{\mathbf{N}_\leqslant} \mathcal{N}_2 \text{ and for all } x \text{ and } y\text{:} \\ h_1(x) = h_1(y) \text{ iff } h_2(x) = h_2(y). \end{array}$$

That is, the pure nihilist language cannot say anything other than how many objects there are. This is not so for the pure universalist language. For one thing, because universalism does not require atomicity or atomlessness, infinite models can have very different structures depending on how many atoms there are. This leads to a fairly strong limitative result:

Proposition 5.1.7 *(No Reasonable Pure Interpretation from Universalism into Nihilism)*. There is no $\mathsf{s}\colon \mathbf{U}_\leqslant \rightsquigarrow_\Delta \mathbf{N}_\leqslant$ such that for all closed $\phi \in \mathcal{L}^1_\leqslant$:

(i) $\mathsf{s}(\neg\,\phi) \equiv_{\mathbf{N}_\leqslant} \neg\,\mathsf{s}(\phi)$, and

(ii) $\mathsf{s}(\phi)$ is equivalent to a closed $\mathcal{L}^1_\leqslant$-formula.

Proof: Suppose there were such an s and Δ. Let L_n be the formula stating that there are at least n-many things and let E_n be the formula stating that there are exactly n-many things. Define $\tilde{\alpha} :=$ ***Atomlessness***.

We start with some simple observations that are easy to verify. First, $\tilde{\alpha} \vDash_{\mathbf{U}_\leqslant} \neg\, E_n$—that is, $\tilde{\alpha}$ is $\mathbf{U}_\leqslant$-inconsistent with finite models. Hence, $\Delta, \mathsf{s}(\tilde{\alpha}) \vDash_{\mathbf{N}_\leqslant} \mathsf{s}(\neg\, E_n)$. Second, by (i) and **Fact 1.3.24**, $\mathsf{s}[\mathbf{U}_\leqslant]$ is opinionated, and thus s must be model-preserving. Third, let $\Delta' := \{\psi \in \mathcal{L} \mid \Delta \vDash_{\mathbf{N}_\leqslant} \psi \text{ and } \psi \text{ is closed}\}$, and let ψ be a closed formula. Then $\Delta \vDash_{\mathbf{N}_\leqslant} \psi$ iff $\Delta' \vDash_{\mathbf{N}_\leqslant} \psi$.

The proof strategy is as follows. First, we show that $\Delta' \cup \{\neg\, \mathsf{s}(\tilde{\alpha})\}$ has an infinite model in N. So by **Fact 5.1.6**, every infinite model in N satisfies $\Delta' \cup \{\neg\, \mathsf{s}(\tilde{\alpha})\}$. Hence, $\Delta', L_2, L_3, L_4, \ldots \vDash_{\mathbf{N}_\leqslant} \neg\, \mathsf{s}(\tilde{\alpha})$. By compactness, there is an n such that $\Delta', L_n \vDash_{\mathbf{N}_\leqslant} \neg\, \mathsf{s}(\tilde{\alpha})$. So, up to elementary equivalence, there are only finitely many models of $\Delta' \cup \{\mathsf{s}(\tilde{\alpha})\}$ in N. However, we then show that there are infinitely many non-equivalent models of $\tilde{\alpha}$ in U. Thus, s is not model-preserving, contrary to what was verified above, $\lightning$.

We first show that $\Delta' \cup \{\neg\, \mathsf{s}(\tilde{\alpha})\}$ has arbitrarily large finite models in $\mathbf{N}_{\leqslant}$. Suppose $\Delta', L_k \vDash_{\mathbf{N}_{\leqslant}} \mathsf{s}(\tilde{\alpha})$. Then $\Delta', L_k \vDash_{\mathbf{N}_{\leqslant}} \mathsf{s}(\neg\, E_n)$ for all $n \geqslant 1$. Clearly, $\Delta' \nvDash_{\mathbf{N}_{\leqslant}} \mathsf{s}(\neg\, E_n)$ for all $n \geqslant 1$, since $\nvDash_{\mathbf{U}_{\leqslant}} \neg\, E_n$. Moreover, up to elementary equivalence, L_k only rules out finitely many models of Δ'. But since each $\mathsf{s}(\neg\, E_n)$ is equivalent to a closed formula, by the pigeon-hole principle, there must be some model of Δ' satisfying $\neg\, \mathsf{s}(\neg\, E_n)$ and $\neg\, \mathsf{s}(\neg\, E_m)$ for $n \neq m$. Thus, $\Delta' \nvDash_{\mathbf{N}_{\leqslant}} \neg(\neg\, \mathsf{s}(\neg\, E_n) \wedge \neg\, \mathsf{s}(\neg\, E_m)) \equiv_{\mathbf{N}_{\leqslant}} \mathsf{s}(\neg(E_n \wedge E_m))$. And yet $\vDash_{\mathbf{U}_{\leqslant}} \neg(E_n \wedge E_m)$, $\lightning$. Hence, for no k does $\Delta, L_k \vDash_{\mathbf{N}_{\leqslant}} \mathsf{s}(\tilde{\alpha})$. That is, $\Delta' \cup \{\neg\, \mathsf{s}(\tilde{\alpha})\}$ has arbitrarily large finite models. By compactness, $\Delta' \cup \{\neg\, \mathsf{s}(\tilde{\alpha})\}$ has an infinite model.

Next, we show that $\tilde{\alpha}$ has infinitely many models in U. Let θ_n be defined to be the following formula (we explain what it means intuitively below):

$$\begin{aligned}
&\forall x\, \exists! y\ (y < x \wedge \forall z\ (z < x \rightarrow z \leqslant y)) \wedge \\
&\exists x_1 \cdots \exists x_{2n} \left(\bigwedge_{i \neq j} x_i \neq x_j \wedge \bigwedge_{i=1}^{n} \forall z\ (z < x_{2i-1} \leftrightarrow z < x_{2i}) \wedge \right. \\
&\forall x\, \forall y \left(x \neq y \wedge \forall z\ (z < x \leftrightarrow z < y) \rightarrow \bigvee_{i=1}^{n} ((x = x_{2i-1} \wedge y = x_{2i}) \vee (x = x_{2i} \wedge y = x_{2i-1})) \right) \wedge \\
&\left. \forall x\, \forall y \left(\bigwedge_{i=1}^{n} ((x = x_{2i-1} \wedge y = x_{2i}) \vee (x = x_{2i} \wedge y = x_{2i-1})) \rightarrow (x \leqslant y \vee y \leqslant x) \right) \right).
\end{aligned}$$

In the case where $n = 0$, we define θ_0 to be the following formula:

$$\forall x\, \exists! y\ (y < x \wedge \forall z\ (z < x \rightarrow z \leqslant y)) \wedge \forall x\, \forall y\ (x \leqslant y \vee y \leqslant x)\,.$$

Intuitively, here is what θ_n says. It says there are exactly n distinct pairs of composite objects x and y such that x and y share all their proper parts, and apart from these pairs, the parthood relation is total. Each θ_n is incompatible with any other θ_m where $m \neq n$. Moreover, each θ_n is compatible with $\tilde{\alpha}$, since we can just take our parthood relation to be an infinite discrete linear order with a top end point but replacing n-many elements (not including the top element) with a pair of incomparable elements sharing the same proper parts. So $\tilde{\alpha}$ has infinitely many models up to elementary equivalence. ■

The proof made use of ***Atomlessness***; but this was not essential.

Proposition 5.1.8 *(No Reasonable Pure Interpretation from Enhanced Universalism into Nihilism)*. Let $\mathbf{L}$ be one of the following languages:

(a) $\mathbf{U}_{\leqslant}$ + ***Atomlessness***

(b) $\mathbf{U}_{\leqslant}$ + $\neg$ ***Atomlessness***

(c) $\mathbf{U}_{\leqslant} + \boldsymbol{Atomicity}$

(d) $\mathbf{U}_{\leqslant} + \neg\, \boldsymbol{Atomicity}$.

Then there is no $\mathsf{s}\colon \mathbf{U}_{\leqslant} \rightsquigarrow_{\Delta} \mathbf{N}_{\leqslant}$ such that for all closed $\phi \in \mathcal{L}^1_{\leqslant}$:

(i) $\mathsf{s}(\neg\,\phi) \equiv_{\mathbf{N}_{\leqslant}} \neg\,\mathsf{s}(\phi)$, and

(ii) $\mathsf{s}(\phi)$ is equivalent to a closed $\mathcal{L}^1_{\leqslant}$-formula.

Proof: For (a) and (d), the proof is exactly as in **Proposition 5.1.7**. For (b) and (c), replace $\tilde{\alpha}$ and θ_n with the following formulas:

$$\begin{aligned} \tilde{\alpha}' &:= \exists x\ (\exists y\ (\sigma u\, \mathsf{Atom}(u) < y < x) \wedge \\ &\qquad\quad \forall y\ (\sigma u\, \mathsf{Atom}(u) < y < x) \to \exists z\ (y < z < x)) \\ \theta'_n &:= \exists_{=n} x\, \mathsf{Atom}(x). \end{aligned}$$

Then as before, $\tilde{\alpha}' \vDash \neg\, E_n$ for each n and the formulas θ'_n are able to distinguish infinitely many models of $\tilde{\alpha}'$. ■

The proof of **Proposition 5.1.7** also relied on the fact that we did not require supplementation (since θ_n is incompatible with it). This assumption is more essential. To illustrate, it is well-known that the models of $\mathbf{GEM}_{\leqslant}$ are all Boolean algebras with the bottom element removed. In particular, the models of $\tilde{\mathbf{A}}\mathbf{GEM}_{\leqslant}$ are all atomless Boolean algebras minus the bottom element. Incidentally, $\tilde{\mathbf{A}}\mathbf{GEM}_{\leqslant}$ is model-complete, which implies by Vaught's test that $\tilde{\mathbf{A}}\mathbf{GEM}_{\leqslant}$ is complete—meaning that if ϕ is closed, then either $\vDash_{\tilde{\mathbf{A}}\mathbf{GEM}_{\leqslant}} \phi$ or $\vDash_{\tilde{\mathbf{A}}\mathbf{GEM}_{\leqslant}} \neg\,\phi$. So the proof above will not work when we replace $\mathbf{U}_{\leqslant}$ with $\tilde{\mathbf{A}}\mathbf{GEM}_{\leqslant}$. Moreover, it will not work if we replace $\mathbf{U}_{\leqslant}$ with $\mathbf{AGEM}_{\leqslant}$, since the infinite models of $\mathbf{AGEM}_{\leqslant}$ are also elementarily equivalent. Still, we do get a closely related limitative result from all this:

Proposition 5.1.9 *(No Reasonable Pure Interpretation from Supplemented Universalism into Nihilism)*. There is no $\mathsf{s}\colon \mathbf{GEM}_{\leqslant} \rightsquigarrow_{\Delta} \mathbf{N}_{\leqslant}$ such that for all closed $\phi \in \mathcal{L}^1_{\leqslant}$:

(i) $\mathsf{s}(\neg\,\phi) \equiv_{\mathbf{N}_{\leqslant}} \neg\,\mathsf{s}(\phi)$,

(ii) $\mathsf{s}(\phi)$ is equivalent to a closed $\mathcal{L}^1_{\leqslant}$-formula, and

(iii) $\mathsf{s}(L_n) \vDash_{\mathbf{N}_{\leqslant}} L_n$ for all $n \geqslant 1$.

Likewise if we replace $\mathbf{GEM}_{\leqslant}$ with $\mathbf{AGEM}_{\leqslant}$, $\tilde{\mathbf{A}}\mathbf{GEM}_{\leqslant}$, $\mathbf{GEM}_{\leqslant} + \neg\, \boldsymbol{Atomicity}$, or $\mathbf{GEM}_{\leqslant} + \neg\, \boldsymbol{Atomlessness}$.

As for the requirement that (a) holds (i.e., that $\mathsf{s}(\neg\,\phi) \equiv_{\mathbf{N}_\leqslant} \neg\,\mathsf{s}(\phi)$), dropping this constraint is not entirely unreasonable. After all, it might be that the nihilist will want to embed negated formulas under some more complicated construction in defining their translation. I have not yet found a proof that allows me to drop (a). (Both directions are needed in the proof above.) Still, if one were looking for an equivalence between universalism and nihilism, it does seem troubling that there is no way to translate universalism into nihilism so as to preserve the boolean operations and so as to map closed formulas to (essentially) closed formulas. For that would mean that the universalist could say something general about the universe as a whole that the nihilist would interpret as being about some particular atoms.

There are other limitative results of this sort worth mentioning. One natural constraint one might want to impose on an adequate interpretation of universalism within nihilism is the constraint that the nihilist and the universalist must "agree" on whatever is said about just the atoms. So for instance, if the universalist says "There are two atoms", the nihilist better also say (something equivalent to) "There are two atoms". But achieving this is quite difficult.

Proposition 5.1.10 *(More Expressive Limitations on Nihilism).*

(a) There is no $\mathsf{s}\colon \mathbf{U}_\leqslant \rightsquigarrow_\Delta \mathbf{N}_\leqslant$ such that for any closed formula $\phi \in \mathcal{L}^1_\leqslant$:

$$\Delta, \phi \equiv_{\mathbf{N}_\leqslant} \Delta, \mathsf{s}(\phi^{\mathsf{Atom}}),$$

where ϕ^{Atom} is the result of restricting every quantifier in ϕ with Atom.

(b) Let $\mathbf{U}^a_\leqslant := \mathbf{U}_\leqslant + \exists x\, \mathsf{Atom}(x)$. Then there is no $\mathsf{s}\colon \mathbf{U}^a_\leqslant \rightsquigarrow_\Delta \mathbf{N}_\leqslant$ such that $\phi \in \mathcal{L}^1_\leqslant$ is closed iff $\mathsf{s}(\phi)$ is closed and for any closed formula $\phi \in \mathcal{L}^1_\leqslant$:

$$\Delta, \phi \equiv_{\mathbf{N}_\leqslant} \Delta, \mathsf{s}(\phi^{\mathsf{Atom}}).$$

(c) Let $\mathbf{U}^{\mathsf{A}}_\leqslant = \mathbf{U}_\leqslant + \textbf{\textit{Atomicity}}$. Then there is no $\mathsf{s}\colon \mathbf{U}^{\mathsf{A}}_\leqslant \rightsquigarrow_\Delta \mathbf{N}_\leqslant$ such that for all $\phi(y_1, \ldots, y_n) \in \mathcal{L}^1_\leqslant$:

$$\Delta, \phi(y_1, \ldots, y_n) \equiv_{\mathbf{N}_\leqslant} \Delta, \mathsf{s}(\bigwedge_{1 \leqslant i \leqslant n} \mathsf{Atom}(y_i) \wedge \phi^{\mathsf{Atom}}(y_1, \ldots, y_n)).$$

Proof:

(a) If there were such a s, then:

$$\Delta \equiv_{\mathbf{N}_\leqslant} \Delta, \exists x\,(x = x)$$

$$\begin{aligned} &\equiv_{\mathbf{N}_\leqslant} \Delta, \mathsf{s}(\exists x\, (\mathsf{Atom}(x) \wedge x = x)) \\ &\equiv_{\mathbf{N}_\leqslant} \Delta, \mathsf{s}(\exists x\, \mathsf{Atom}(x)). \end{aligned}$$

So $\Delta \vDash_{\mathbf{N}_\leqslant} \mathsf{s}(\exists x\, \mathsf{Atom}(x))$. But then $\vDash_{\mathbf{U}_\leqslant} \exists x\, \mathsf{Atom}(x)$, $\lightning$.

(b) Because we are restricting ourselves to $\mathbf{U}^a_\leqslant$, the proof above will not work, since $\vDash_{\mathbf{U}^a_\leqslant} \exists x\, \mathsf{Atom}(x)$. But we will show that the existence of such a s still leads to a contradiction in this case.

First, we observe the following: if $\phi \in \mathcal{L}^1_\leqslant$ is closed, then $\phi \equiv_{\mathbf{U}^a_\leqslant} \mathsf{s}(\phi)^{\mathsf{Atom}}$. For since ϕ is closed, $\mathsf{s}(\phi)$ is closed too. So $\Delta, \mathsf{s}(\phi) \equiv_{\mathbf{N}_\leqslant} \Delta, \mathsf{s}(\mathsf{s}(\phi)^{\mathsf{Atom}})$. Hence, by interpretability, $\phi \equiv_{\mathbf{U}^a_\leqslant} \mathsf{s}(\phi)^{\mathsf{Atom}}$.

But observe that there is no closed $\psi \in \mathcal{L}^1_\leqslant$ where $\exists x\, \neg \mathsf{Atom}(x) \equiv_{\mathbf{U}^a_\leqslant} \psi^{\mathsf{Atom}}$. For let $\mathcal{M}$ be a universalist model with at least one atom, and let $\mathcal{M}^{\mathsf{A}} = \langle D^{\mathsf{A}}, I^{\mathsf{A}} \rangle$, where:

$$\begin{aligned} D^{\mathsf{A}} &= \left\{a \in D \,\middle|\, \mathcal{M}, g^x_a \Vdash_{\mathbf{U}^a_\leqslant} \exists y\, (\mathsf{Atom}(y) \wedge y \leqslant x)\right\} \\ I^{\mathsf{A}}(\leqslant) &= I(\leqslant) \cap (D^{\mathsf{A}})^2. \end{aligned}$$

Then by induction, for all closed $\psi \in \mathcal{L}^1_\leqslant$, $\mathcal{M} \Vdash_{\mathbf{U}^a_\leqslant} \psi^{\mathsf{Atom}}$ iff $\mathcal{M}^{\mathsf{A}} \Vdash_{\mathbf{U}^a_\leqslant} \psi^{\mathsf{Atom}}$. But then where $\mathcal{M} \Vdash_{\mathbf{U}^a_\leqslant} \exists x\, \neg \mathsf{Atom}(x)$, $\mathcal{M}^{\mathsf{Atom}} \nVdash_{\mathbf{U}^a_\leqslant} \exists x\, \neg \mathsf{Atom}(x)$. So $\exists x\, \neg \mathsf{Atom}(x) \not\equiv_{\mathbf{U}^a_\leqslant} \mathsf{s}(\exists x\, \neg \mathsf{Atom}(x))^{\mathsf{Atom}}$, $\lightning$.

(c) If there were such an s, then we would have:

$$\begin{aligned} \Delta, \mathsf{s}(\neg \mathsf{Atom}(x)) &\equiv_{\mathbf{N}_\leqslant} \Delta, \mathsf{s}(\mathsf{Atom}(x) \wedge \mathsf{s}(\neg \mathsf{Atom}(x))^{\mathsf{Atom}}) \\ &\vDash_{\mathbf{N}_\leqslant} \mathsf{s}(\mathsf{Atom}(x)). \end{aligned}$$

Hence, we would have $\neg \mathsf{Atom}(x) \vDash_{\mathbf{U}^{\mathsf{A}}_\leqslant} \mathsf{Atom}(x)$, $\lightning$. ∎

In other words, we cannot find a way of interpreting $\mathbf{U}_\leqslant$ inside $\mathbf{N}_\leqslant$ so that the nihilist can agree that they can say whatever the universalist says solely about atoms.[5] Of course, there could well be some other way of schematically interpreting $\mathbf{U}_\leqslant$ within $\mathbf{N}_\leqslant$. But since such an interpretation will necessarily lack the properties from the result above, one can easily question whether such an interpretation does justice to the debate between the universalist and the nihilist, just as one may question whether the barthood interpretations do. So if we restrict to $\mathbf{U}_\leqslant$ and $\mathbf{N}_\leqslant$, it does seem at least that $\mathbf{U}_\leqslant$ is more powerful in a number of different (but relevant) senses.

[5] Turner [2011a] proves a related result.

None of this constitutes a proof that there is no $\mathtt{s}\colon \mathbf{U}_{\leqslant} \multimap \mathbf{N}_{\leqslant}$. In general, showing that there is no schematic translation from one language to another is very challenging when non-schematic translations exist. Even in very simple cases, such as showing there is no schematic translation from intuitionistic logic to classical logic, the proofs are rather syntactic. Still, the limitative results show that even if there is a schematic translation from $\mathbf{U}_{\leqslant}$ to $\mathbf{N}_{\leqslant}$, its existence will not be terribly convincing evidence that universalism and nihilism are on a par.

§ 5.1.6 Second-Order Mereology

A common way that nihilists interpret the univeralist's language is via plural quantification. So it makes sense to look at how the picture changes, if at all, when we move to higher-order mereology. Let $\mathcal{L}^2_{\mathtt{Pred}}(\leqslant)$ be the second-order language defined as follows:

$$\phi ::= P^n(y_1, \ldots, y_n) \mid X(y) \mid x = y \mid x \leqslant y \mid \neg\phi \mid (\phi \wedge \phi) \mid \forall x\, \phi \mid \forall X\, \phi.$$

For brevity, we will use the following abbreviations throughout:

$$\begin{aligned} X \subseteq Y \quad &:= \quad \forall z\ (X(z) \rightarrow Y(z)) \\ X = Y \quad &:= \quad X \subseteq Y \wedge Y \subseteq X. \end{aligned}$$

As before, we will let $\mathcal{L}^2_{\leqslant}$ denote the restriction of $\mathcal{L}^2_{\mathtt{Pred}}(\leqslant)$ to the purely mereological vocabulary.

We will use the standard (full) semantics for second-order logic with the exception that we will not include the empty set in the domain of quantification.[6] Thus, let M^2 be the class of pairs $\langle \mathcal{M}, g\rangle$ where $\mathcal{M} = \langle D, I\rangle$ is a first-order model, where $I(\leqslant)$ is a partial order, and where g is a *second-order* variable assignment on $\mathcal{M}$ (meaning g also maps second-order variables to nonempty subsets of D). Let $\Vdash$ be the standard (i.e., full) second-order satisfaction relation except we disallow second-order quantification to include the empty set. Then we define $\mathbf{M}^2 = \langle \mathcal{L}^2_{\mathtt{Pred}}(\leqslant), \mathsf{M}^2, \Vdash\rangle$ to be the language of second-order core mereology. As before, $\mathbf{M}^2_{\leqslant}$ will be the pure version of $\mathbf{M}^2$.

As before, we can define $\mathbf{N}^2 = \mathbf{M}^2 +$ ***Nihilism***. We define $\mathbf{U}^2$ analogously, except we interpret ***Universalism*** so that ϕ can be a second-order formula as well. Given that we are working in the standard semantics, such a second-order schema turns out to be equivalent to the second-order formula:

Universalism Simplified. $\forall X\, \exists y\, \forall z\ (y \circ z \leftrightarrow \exists x\ (X(x) \wedge x \circ z))$.

[6] The reason for this restriction is that the higher-order quantification is meant to be interpreted as "plural quantification", and there is reluctance in calling the empty set a "plurality". It turns out that this restriction of the second-order quantifier to nonempty subsets of the domain is nontrivial. If we allowed quantification over every subset of the domain, then some of the proofs that follow would break down.

Now, by going higher-order, nothing changes in terms of the barthood strategy: we still have $\mathbf{N}^2 \mathrel{\substack{\multimapinv \\ \multimap}} \mathbf{U}^2$. (Of course, the universalist will interpret the second-order nihilist quantifier as a restricted second-order quantifier, viz., a second-order quantifier restricted to sets of atoms.) So the interesting question is whether the relationship between the pure versions of the nihilist and universalist language changes when going second-order. The answer is mixed.

First, we observe that first-order universalism is already strong enough to interpret second-order nihilism.[7]

Proposition 5.1.11 *(Pure Interpretation of Second-Order Nihilism into Universalism)*. There is a schematic $\mathtt{t}\colon \mathbf{N}^2_{\leqslant} \propto_i \mathbf{U}_{\leqslant}$.

[7]The translation in **Proposition 5.1.11** was essentially proposed by Warren [2015], though the translation he actually proposed contains a slight error. The translation he gives is presented below (in our notation):

$$\begin{aligned}
\mathtt{t}(x \leqslant y) &= \mathsf{Atom}(x) \wedge \mathsf{Atom}(y) \wedge x = y \\
\mathtt{t}(x = y) &= \mathsf{Atom}(x) \wedge \mathsf{Atom}(y) \wedge x = y \\
\mathtt{t}(X(y)) &= y \leqslant x \\
\mathtt{t}(\neg\,\phi) &= \neg\,\mathtt{t}(\phi) \\
\mathtt{t}(\phi \wedge \psi) &= \mathtt{t}(\phi) \wedge \mathtt{t}(\psi) \\
\mathtt{t}(\forall x\,\phi) &= \forall x\ (\mathsf{Atom}(x) \to \mathtt{t}(\phi)) \\
\mathtt{t}(\forall X\,\phi) &= \forall x\,\mathtt{t}(\phi).
\end{aligned}$$

Warren claims this $\mathtt{t}$ is what we call a quasitranslation—that is, if ϕ is valid in $\mathbf{N}^2$, then $\mathtt{t}(\phi)$ is valid in $\mathbf{U}$. The claim is not quite correct, however. The problem is that because both X and x get mapped to the same first-order variable (viz., x), we can get unwanted conflicts. Here is an example. Consider the formula:

$$\forall x\,\forall y\ (x \neq y \to \exists Y\ (Y(x) \wedge \neg\,Y(y)))\,.$$

This is valid in $\mathbf{N}^2$, since if $g(x) \neq g(y)$, we can assign Y to $\{g(x)\}$. But its translation is not valid in $\mathbf{U}$:

$$\begin{aligned}
& \mathtt{t}(\forall x\,\forall y\ (x \neq y \to \exists Y\ (Y(x) \wedge \neg\,Y(y)))) \\
={} & \forall x\ (\mathsf{Atom}(x) \to \forall y\ (\mathsf{Atom}(y) \to \\
& \quad (\neg(\mathsf{Atom}(x) \wedge \mathsf{Atom}(y) \wedge x = y) \to \exists y\ (x \leqslant y \wedge \neg(y \leqslant y))))) \\
\equiv_{\mathbf{U}}{} & \forall x\,\forall y\ (\mathsf{Atom}(x) \wedge \mathsf{Atom}(y) \to (x \neq y \to \exists y\ (x \leqslant y \wedge \bot))) \\
\equiv_{\mathbf{U}}{} & \forall x\,\forall y\ (\mathsf{Atom}(x) \wedge \mathsf{Atom}(y) \to x = y)\,.
\end{aligned}$$

So Warren's translation is incorrect. Nevertheless, it is easily modified to form a schematic interpretation of $\mathbf{N}^2$ into $\mathbf{U}$.

Proof: The proof is similar to before, except we need to slightly modify $\mathtt{t}$ as follows:

$$\begin{array}{rcl} \mathtt{t}(x_i) & = & x_{2i} \\ \mathtt{t}(X_i) & = & x_{2i+1} \\ \mathtt{t}(X(y)) & = & \mathtt{t}(y) \leqslant \mathtt{t}(X) \\ \mathtt{t}(x = y) & = & \mathtt{t}(x) = \mathtt{t}(y) \\ \mathtt{t}(x \leqslant y) & = & \mathtt{t}(x) = \mathtt{t}(y) \\ \mathtt{t}(\neg\phi) & = & \neg\,\mathtt{t}(\phi) \\ \mathtt{t}(\phi \wedge \psi) & = & \mathtt{t}(\phi) \wedge \mathtt{t}(\psi) \\ \mathtt{t}(\forall x\,\phi) & = & \forall \mathtt{t}(x)\ (\mathsf{Atom}(\mathtt{t}(x)) \rightarrow \mathtt{t}(\phi)) \\ \mathtt{t}(\forall X\,\phi) & = & \forall \mathtt{t}(X)\,\mathtt{t}(\phi). \end{array}$$

Define $\Gamma := \{\mathsf{Atom}(\mathtt{t}(x)) \mid x \text{ is a variable}\}$. Then the proof is as before. ■

This result might make it seem as though $\mathbf{U}_{\leqslant}$ is quite a powerful theory. Actually, the result is not surprising, given that every $\mathcal{L}^2_{\leqslant}$-formula (without free second-order variables) is $\mathbf{N}_{\leqslant}$-equivalent to some $\mathcal{L}^1_{\leqslant}$-formula.[8] This means we get the following result for free:

Corollary 5.1.12 *(No Reasonable Pure Interpretation from Universalism into Second-Order Nihilism)*. There is no $\mathtt{s}\colon \mathbf{U}_{\leqslant} \rightsquigarrow_{\Delta} \mathbf{N}^2_{\leqslant}$ such that for all closed $\phi \in \mathcal{L}^1_{\leqslant}$:

(i) $\mathtt{s}(\neg\phi) \equiv_{\mathbf{N}^2_{\leqslant}} \neg\,\mathtt{s}(\phi)$, and

(ii) $\mathtt{s}(\phi)$ is equivalent to a closed $\mathcal{L}^2_{\leqslant}$-formula.

Likewise for the other languages (without supplementation).

Warren [2015] proposed the following quasitranslation from $\mathbf{AGEM}_{\leqslant}$ into $\mathbf{N}^2_{\leqslant}$ (again in our notation):

$$\begin{array}{rcl} \mathtt{s}(x_i) & = & X_i \\ \mathtt{s}(x = y) & = & \mathtt{s}(x) = \mathtt{s}(y) \\ \mathtt{s}(x \leqslant y) & = & \mathtt{s}(x) \subseteq \mathtt{s}(y) \\ \mathtt{s}(\neg\phi) & = & \neg\,\mathtt{s}(\phi) \\ \mathtt{s}(\phi \wedge \psi) & = & \mathtt{s}(\phi) \wedge \mathtt{s}(\psi) \\ \mathtt{s}(\forall x\,\phi) & = & \forall \mathtt{s}(x)\,\mathtt{s}(\phi). \end{array}$$

[8]This follows from a result due to Ackermann [1954] that every monadic second-order language with no non-logical predicates is equivalent to a first-order formula with identity.

As noted previously, proving the existence of a quasitranslation is not very informative. So it is natural to ask whether s is a full translation. The answer is affirmative *if* we continue to assume full $\mathbf{AGEM}_{\leqslant}$, but negative if we drop either ***Atomicity*** or ***Supplementation***.

Proposition 5.1.13 *(Warren's Translation of Atomic Universalism into Nihilism).* Let s be defined as follows:

$$\begin{array}{rcl} \mathsf{s}(x_i) & = & X_i \\ \mathsf{s}(x = y) & = & \mathsf{s}(x) = \mathsf{s}(y) \\ \mathsf{s}(x \leqslant y) & = & \mathsf{s}(x) \subseteq \mathsf{s}(y) \\ \mathsf{s}(\neg \phi) & = & \neg\, \mathsf{s}(\phi) \\ \mathsf{s}(\phi \wedge \psi) & = & \mathsf{s}(\phi) \wedge \mathsf{s}(\psi) \\ \mathsf{s}(\forall x\, \phi) & = & \forall \mathsf{s}(x)\, \mathsf{s}(\phi). \end{array}$$

Then $\mathsf{s}\colon \mathbf{AGEM}_{\leqslant} \propto \mathbf{N}_{\leqslant}$.

Proof: The proof that $\propto_{\mathsf{s}}$ is surjective goes through as before (just drop mention of Δ), so it suffices to show totality. Let $\langle \mathcal{M}, g\rangle \in \mathsf{A} \cap \mathsf{S}$. Define $\mathcal{M}^{\mathsf{N}} = \langle D^{\mathsf{N}}, I^{\mathsf{N}}\rangle$, where:

$$\begin{array}{rcl} D^{\mathsf{N}} & = & \{a \in D \mid \mathcal{M}, g^x_a \Vdash \mathsf{Atom}(x)\} \\ I^{\mathsf{N}}(\leqslant) & = & \{\langle a, a\rangle \mid a \in D^{\mathsf{N}}\}. \end{array}$$

Define $g^{\mathsf{N}}(X_i) = \{a \in D^{\mathsf{N}} \mid \langle a, g(x_i)\rangle \in I(\leqslant)\}$. Then by induction, $\mathcal{M}, g \Vdash_{\mathbf{AGEM}_{\leqslant}} \phi$ iff $\mathcal{M}^{\mathsf{N}}, g^{\mathsf{N}} \Vdash_{\mathbf{N}^2_{\leqslant}} \mathsf{s}(\phi)$. Hence, s is model-preserving. ∎

The translation cannot be extended to $\mathbf{GEM}_{\leqslant}$ or to $\mathbf{U}_{\leqslant}$+ ***Atomicity***. For $\mathbf{GEM}_{\leqslant}$, consider an equivalent formulation of ***Atomlessness***:

$$\forall x\, \exists y\ (y < x).$$

This is $\mathbf{GEM}_{\leqslant}$-consistent. But its s-translation is $\mathbf{N}^2_{\leqslant}$-inconsistent:

$$\forall X\, \exists Y\ (Y \subset X).$$

So $\mathsf{s}\colon \mathbf{GEM}_{\leqslant} \not\propto \mathbf{N}^2_{\leqslant}$. For $\mathbf{U}_{\leqslant}$+ ***Atomicity***, Consider the $\mathbf{U}_{\leqslant}$-formula:

$$\exists x\, \exists y\ (x \neq y) \wedge \forall x\, \forall y\ (x \leqslant y \vee y \leqslant x).$$

This formula is $\mathbf{U}^{\mathsf{A}}_{\leqslant}$-consistent. But its s-translation is $\mathbf{N}^2_{\leqslant}$-inconsistent:

$$\exists X\, \exists Y\ (X \neq Y) \wedge \forall X\, \forall Y\ (X \subseteq Y \vee Y \subseteq X).$$

So $\mathsf{s}\colon \mathbf{U}^{\mathsf{A}}_{\leqslant} \not\propto \mathbf{N}^2_{\leqslant}$.

§ 5.1.7 The Possibility of Gunk

The results of the previous section strongly suggest that universalism, in its most general form, is strictly more expressive than nihilism. For while nihilism can be easily compositionally interpreted into universalism, the converse is not true. We have not proven, beyond a doubt, that there is no compositional interpretation of universalism into nihilism; but if there is one, it will need to be so unnatural that one might question whether the interpretation can be said to be at all charitable.

Part of the reason the universalist enjoys this additional expressive power is that the universalist recognizes the possibility of gunk, i.e., the possibility that some object might not ultimately decompose into atoms. For the nihilist, there could not be such an object. So there are possibilities that the universalist recognizes that the nihilist cannot recognize.

Sider [1993] pointed out a version of this problem for nihilism. For nihilism entails that it is metaphysically impossible for there to be gunky objects. But these objects seem intuitively possible. As Schaffer [2003, p. 501] puts it, gunky objects have many of the telltale signs of metaphysical possibility: "infinite division passes excellent tests for being possible: (a) *conceivable*, (b) *logically consistent*, and (c) *physically serious*." Thus, one might think this difference in expressive power is enough to tip the scales in favor of universalism. If there is a seemingly genuine possibility that the universalist can recognize that the nihilist cannot, that suggests the latter is too weak for a base theory of parthood.

But the nihilist need not concede that gunky objects are genuinely possible. There are, of course, well known arguments from the history of philosophy against the possibility of infinite divisibility (e.g., Kant's antinomies). And one might reasonably be skeptical of the idea that an object could be decomposable "without end". One may ask: if the limit of decomposition was nothing, how could there be anything to compose the object to begin with. And there are other defenses of nihilism against the alleged possibility of gunk.[9]

This is not to say that the nihilist is correct that gunky objects are impossible (indeed, I am inclined to think otherwise). But what this illustrates is that even if we grant that universalism has more expressive power than nihilism, we can still maintain that a genuine dispute can be had between the universalist and the nihilist. The universalist will maintain that their theory is to be preferred since they have greater expressive power. The nihilist will deny that the additional expressive power afforded by the universalist carves out a genuine distinction. Thus, the debate can be said to be genuine insofar as the theories disagree over what possibilities ought to be recognized.

[9] Williams [2006] defends a version of nihilism called *emergence nihilism*, according to which the world contains macroscopic simples that can be (partially) colocated. Williams claims that it is this possibility that seems to give rise to the illusion that gunky objects are possible.

§ 5.2 Persistence

One of the most central questions in metaphysics concerns the identity of persons over time: what makes one person at one time the same person as a person at a different time? Or, more importantly, how do people survive through time? This question can be generalized from persons to any kind of object: what makes one material object at one time the same as a material object at a different time? How do material objects *persist* through time?

There have been two influential answers to this question in contemporary metaphysics. One theory of persistence is ***four-dimensionalism***, or ***perdurantism***, according to which objects persist by having temporal parts located at different times. An alternative theory is ***three-dimensionalism***, or ***endurantism***, according to which objects persist by being "wholly present" at different times. We can put the difference between the two views as follows. Let us say that an object ***persists*** through time if it exists at multiple times. Let us say an object ***perdures*** across time if it has parts at multiple times. Let us say an object ***endures*** across time if it is wholly present at multiple times. Thus, perdurantism and endurantism propose different accounts of how objects persist through time: perdurantism says persistence is perdurance, while endurantism says persistence in endurance.

This characterization of the debate is notoriously fraught with difficulties. Many ways of formulating the debate trivialize it. In particular, it is not really clear how to spell out what it means to be "wholly present" at multiple times in such a way that endurantism does not become trivialized or absurd. If being "wholly present" means that every part of them at every time at which they exist also exists at that time, then the perdurantist does not deny this. If it means that all of the object's parts across all time exist at every time at which the object exists, then endurantism seems plainly false (or else implies some kind of mereological essentialism, according to which objects cannot survive change of their parts). Thus, unlike the mereological debate, it becomes tricky to use the method of translations to assess the status of this debate.

In this section, we will understand the debate between perdurantism and endurantism as one over whether to admit certain first-order models of persistence. Perdurantism, on this picture, corresponds to the addition of a certain constraint on temporal existence, viz., an object exists at a time only if it has a temporal part at that time. Any model violating this constraint is ruled out by the perdurantist as an unrealistic model of persistence. Endurantism, by contrast, corresponds to the lack of this constraint. That is, the endurantist will hold that models violating this constraint can be accurate representations of persistence.[10]

[10] It might be tempting to seek a formulation of endurantism that imposes an opposing constraint on the class of models that can accurately represent persistence. But this is difficult to do without leaning on differences in commitments to certain temporal mereological principles. For example, one might think that endurantism should be formulated as the view that no object has a proper tem-

An alternative approach to characterizing the debate between perdurantism and endurantism utilizes locational locutions to distinguish the two views. This approach is a more recent development due primarily to Parsons [2007]. The idea is to understand perdurantism as the view that objects persist by having exact temporal locations that contain the time through which they persist. Endurantism, then, is the view that objects persist by having all of their parts' exact locations overlap with every region of time through which they persist. While this approach is worth pursuing further, I do not take up the project of investigating the relative expressive power of endurantism and perdurantism in the locational framework here.

Throughout, we will work within the following two-sorted first-order syntax:

$$\begin{aligned}
\alpha &::= P^{n;m}(\overline{x};\overline{t}) \mid E(x;t) \mid (x = y) \mid (s = t) \mid (x \leqslant_t y) \mid (s \subseteq t) \\
\phi &::= \alpha \mid \neg\,\phi \mid (\phi \wedge \phi) \mid \forall x\, \phi \mid \forall t\, \phi.
\end{aligned}$$

Here, variables x, y, etc. stand for objects, while s, t, etc. stand for intervals of time. Intervals, as we are defining things here, need not be connected. Nor need they be defined in terms of collections of points of time. Nothing we say here will assume anything about the structure of intervals of time except that they satisfy the axioms of **GEM**. The interpretation of the special predicates of this language are given as follows:

$$\begin{array}{rl}
E(x;t) & (\text{“}x \text{ exists throughout } t\text{”}) \\
x \leqslant_t y & (\text{“}x \text{ is a part of } y \text{ throughout } t\text{”}) \\
s \subseteq t & (\text{“}s \text{ is a subinterval of } t\text{”}).
\end{array}$$

Object quantifiers range over all objects across all times. The following abbreviations will be used throughout:

$$\begin{array}{rcll}
\forall_t x\, \phi & := & \forall x\ (E(x;t) \to \phi) & (\text{“for all } x \text{ that exist throughout } t, \ldots\text{”}) \\
(x \equiv_t y) & := & (x \leqslant_t y) \wedge (y \leqslant_t x) & (\text{“}x \text{ perfectly overlaps with } y \text{ throughout } t\text{”}) \\
(x \leqslant y) & := & \forall t\ (E(x;t) \to (x \leqslant_t y)) & (\text{“}x \text{ is a part of } y\text{”}) \\
\lambda(x) & := & \sigma t\, E(x;t) & (\text{“the lifespan of } x\text{”}).
\end{array}$$

Models are of the form $\mathcal{M} = \langle D, T, I \rangle$, where D and T are disjoint nonempty sets and I is an interpretation function, so that:

poral part. The problem with this formulation is that it rules out the possibility of distinct objects that have all the same parts at some time (which might be motivated as a response to, e.g., the paradox of material constitution). Endurantism, as a view about persistence, ought not take a stand on whether multiple distinct objects can be exactly co-located, or have exactly the same parts, at some time. (At the very least, we wish to remain neutral on this issue in our formulation of endurantism for the sake of generality.) If the endurantist is willing to rule out this possibility, it will turn out that their view will automatically be incompatible with perdurantism (unless objects only exist at a single time), and so no additional constraint is needed to characterize endurantism in any case.

- $I(P^{n;m}) \subseteq D^n \times T^m$
- $I(E) \subseteq D \times T$
- $I(\leqslant) \subseteq D^2 \times T$
- $I(\subseteq) \subseteq T^2$.

We will use the following abbreviations throughout (where $t \in T$):

$$D_t := \{a \in D \mid \langle a; t\rangle \in I(E)\}$$
$$I(\leqslant)_t := \{\langle a, b\rangle \in D^2 \mid \langle a, b; t\rangle \in I(\leqslant)\}\,.$$

Variable assignments are defined as usual. The satisfaction relation for model-assignment pairs and formulas in this syntax is defined as the standard one.

A ***persistence*** model is a first-order model satisfying the following axioms:

$$\begin{aligned}
&\forall x\, \exists t\, E(x; t)\\
&\forall t\, \exists s \subseteq t\, \exists x\, E(x; s)\\
&\forall t\, \forall_t x\, \forall s \subseteq t\, E(x; s)\\
&\forall t\, \forall s \subseteq t\, \forall x\, \forall y\ (x \leqslant_t y \to x \leqslant_s y)\\
&\forall t\, \forall x\, \forall y\ (x \leqslant_t y \to E(x; t) \land E(y; t))\\
&\forall t\, \forall_t x\ (x \leqslant_t x)\\
&\forall t\, \forall_t x\, \forall_t y\, \forall_t z\ (x \leqslant_t y \land y \leqslant_t z \to x \leqslant_t z)\\
&\forall t\ (t \subseteq t)\\
&\forall s\, \forall t\ (s \subseteq t \land t \subseteq s \to s = t)\\
&\forall s\, \forall t\, \forall r\ (r \subseteq s \land s \subseteq t \to r \subseteq t)\\
&\forall s\, \forall t\ (\neg(s \subseteq t) \to \exists r\ (r \subseteq s \land r \perp t))\\
&\forall \overline{u}\ (\exists t\, \phi(t, \overline{u}) \to \exists s\, \forall r\ (s \circ r \leftrightarrow \exists t\ (\phi(t, \overline{u}) \land t \circ r)))\,,
\end{aligned}$$

where $\overline{u}$ can be any sequence of variables not including t, s, or r. Here is a brief explanation of the axioms. The first two axioms state that everything must exist at some time, and no time can be completely unpopulated. The next two axioms ensures that existence and parthood is inherited by subintervals, i.e., that $E(x; t)$ means that x exists *throughout* t and that $(x \leqslant_t y)$ means that x is a part of y *throughout* t. The next axiom ensures that parts exist at the times at which they are parts. The next two axioms say that $\leqslant_t$ is reflexive and transitive. The rest of the axioms state that $\subseteq$ obey **GEM**. Note that $\leqslant$ is automatically a preorder, in that:

$$\forall x\ (x \leqslant x)$$
$$\forall x\, \forall y\, \forall z\ (x \leqslant y \land y \leqslant z \to x \leqslant z)\,.$$

A ***perdurance*** model is a persistence model satisfying the additional axiom:

$$\forall t\, \forall_t x\, \exists_t y\ (y \equiv_t x \land \forall s\ (E(y; s) \to s \subseteq t))\,.$$

This axiom states that for an object to exist throughout some interval, there must be an object that exactly overlaps with it throughout the interval and only exists at that interval.

Fact 5.2.1 *(Alternative Formulation of Perdurance).* A persistence model is a perdurance model iff it satisfies the axiom:

$$\forall t\, \forall_t x\, \exists_t y\ \ (y \equiv_t x \wedge \lambda(y) = t)\,.$$

Let **3D** be the language consisting of this syntax, the class of persistence model-assignment pairs, and the standard first-order satisfaction relation. Let **4D** be the restriction of **3D** to perdurance models. Clearly, we have:

Proposition 5.2.2 *(Schematic Translation from Perdurantism into Endurantism).* **4D** $\multimap$ **3D**.

Proof: Let θ be the formula:

$$\forall t\, \forall_t x\, \exists_t y\ \ (y \equiv_t x \wedge \forall s\ \ (E(y;s) \rightarrow s \subseteq t))\,.$$

Let $\mathtt{t}$ be defined as follows (where α is any atomic formula):

$$\begin{array}{rcl} \mathsf{s}(\alpha) & = & \theta \rightarrow \alpha \\ \mathsf{s}(\neg\,\phi) & = & \theta \rightarrow \neg\,\mathsf{s}(\phi) \\ \mathsf{s}(\phi \wedge \psi) & = & \theta \rightarrow (\mathsf{s}(\phi) \wedge \mathsf{s}(\psi)) \\ \mathsf{s}(\forall x\, \phi) & = & \theta \rightarrow \forall x\, \mathsf{s}(\phi) \\ \mathsf{s}(\forall t\, \phi) & = & \theta \rightarrow \forall t\, \mathsf{s}(\phi). \end{array}$$

It is straightforward to verify that $\mathtt{t}\colon$ **4D** $\multimap$ **3D**. ∎

The question now arises whether **4D** can interpret **3D**. It turns out the answer is affirmative if we are allowed to appeal to an additional predicate. Let R be some arbitrarily chosen unary predicate (for the "real" objects).

Proposition 5.2.3 *(Schematic Interpretation of Endurantism into Perdurantism).* **3D** $\multimap_\Gamma$ **4D**, where Γ is defined as follows:

$$\Gamma := \{R(x) \mid x \text{ is a variable}\} \cup \{\forall t\, \exists s \subseteq t\, \exists x\ \ (R(x) \wedge E(x;s))\}\,.$$

Proof: Define the translation $\mathtt{t}$ as follows:

$$\begin{array}{rcl} \mathtt{t}(\alpha) & = & \alpha \quad \text{for atomic } \alpha \\ \mathtt{t}(\neg\,\phi) & = & \neg\,\mathtt{t}(\phi) \\ \mathtt{t}(\phi \wedge \psi) & = & \mathtt{t}(\phi) \wedge \mathtt{t}(\psi) \\ \mathtt{t}(\forall x\, \phi) & = & \forall x\ (R(x) \to \mathtt{t}(\phi)) \\ \mathtt{t}(\forall t\, \phi) & = & \forall t\, \mathtt{t}(\phi). \end{array}$$

We will now show that $\mathtt{t}$ is model-preserving relative to Γ.

First, we show totality. Let $\mathcal{M} = \langle D, T, I \rangle$ be a persistence model. Define $\mathcal{M}^{\text{Perd}} = \langle D^{\text{Perd}}, T, I^{\text{Perd}} \rangle$ as follows:

$$\begin{array}{rcl} D^{\text{Perd}} & = & \{\langle a; t\rangle \mid a \in D_t\} \\ I^{\text{Perd}}(P^{n;m}) & = & \left\{\langle \overline{\langle a; \lambda(a)\rangle}; \bar{t}\rangle \;\middle|\; \langle \bar{a}; \bar{t}\rangle \in I(P^{n;m})\right\} \\ I^{\text{Perd}}(E) & = & \{\langle\langle a; s\rangle; t\rangle \mid a \in D_t \text{ and } t \subseteq s\} \\ I^{\text{Perd}}(\leqslant)_t & = & \{\langle\langle a; s\rangle, \langle b; r\rangle\rangle \mid \langle a, b\rangle \in I(\leqslant)_t \text{ and } t \subseteq s, r\} \\ I^{\text{Perd}}(\subseteq) & = & I(\subseteq) \\ I^{\text{Perd}}(R) & = & \{\langle a; \lambda(a)\rangle \mid a \in D\}\,. \end{array}$$

Where $g \in \mathsf{VA}(\mathcal{M})$, let $g^{\text{Perd}}(x) = \langle g(x), \lambda(g(x))\rangle$. It is tedious but straightforward to show by induction that $\mathcal{M}^{\text{Perd}}$ is a perdurance model satisfying Γ and for all ϕ, $\mathcal{M}, g \Vdash_{\mathbf{3D}} \phi$ iff $\mathcal{M}^{\text{Perd}}, g^{\text{Perd}} \Vdash_{\mathbf{4D}} \mathtt{t}(\phi)$.

Next, surjectivity. Let $\mathcal{N} = \langle D, T, I \rangle$ be a perdurance model and let $g \in \mathsf{VA}(\mathcal{N})$ be such that $\mathcal{N}, g \Vdash_{\mathbf{4D}} \Gamma$. Define $\mathcal{N}^{\text{End}} = \langle D^{\text{End}}, T, I^{\text{End}}\rangle$ as follows (where α is an arbitrary predicate):

$$\begin{array}{rcl} D^{\text{End}} & = & I(R) \\ I^{\text{End}}(\alpha) & = & I(\alpha) \restriction_{D^{\text{End}}}\,. \end{array}$$

Again, it is straightforward to check that $\mathcal{N}^{\text{End}}$ is a persistence model and that for all ϕ, $\mathcal{N}, g \Vdash_{\mathbf{4D}} \mathtt{t}(\phi)$ iff $\mathcal{N}^{\text{End}}, g \Vdash_{\mathbf{3D}} \phi$. ■

Corollary 5.2.4 *(Generalizing the Mutual Interpretation Between Perdurantism and Endurantism)*. Let Λ be a set of formulas. Let Λ^R be the set of formulas obtained from Λ by replacing all quantifiers with quantifiers bounded by R.

(a) Where $\mathtt{s}$ is defined as before, $\mathtt{s}\colon \mathbf{4D} + \Lambda \multimap \mathbf{3D} + \Lambda$.

(b) Where $\mathtt{t}$ and Γ are defined as before, $\mathtt{t}\colon \mathbf{3D} + \Lambda \multimap_{\Gamma \cup \Lambda^R} \mathbf{4D} + \Lambda$.

Schematic interpretability is not preserved through all modifications of these languages, however. So far, we have worked with fairly minimal articulations of these two theories of persistence that do not take a stand on any controversial principle of temporal mereology. But as it so happens, the two theories are typically associated with different principles of temporal mereology, and one might suspect that the disagreements arising from the dispute over persistence stem from a more basic disagreement over these principles. That is what we will now illustrate.

When one imports a principle of mereology into temporal mereology, there are two versions one can consider, which (very roughly) might be characterized as a version that applies primarily to three-dimensional objects and a version that applies to four-dimensional objects. More neutrally, we may say there are *synchronic* and *diachronic* versions of any given mereological principle. The synchronic version of the principle requires all the quantifiers to be restricted to objects that exist at a particular time, whereas the diachronic version concerns objects across time. For instance, consider ***Antisymmetry***:

$$\forall x\, \forall y\ (x \leqslant y \wedge y \leqslant x \to x = y)\,.$$

The synchronic version of this principle would be:

$$\forall t\, \forall_t x\, \forall_t y\ (x \leqslant_t y \wedge y \leqslant_t x \to x = y)\,.$$

That is, if some three-dimensional objects exactly overlap at any time, then they are identical. The diachronic version of this principle would leave the original formula untouched, but could be equivalently stated as follows:

$$\forall x\, \forall y\ (\forall t\ (E(x;t) \to x \leqslant_t y) \wedge \forall t\ (E(y;t) \to y \leqslant_t x) \to x = y)\,.$$

That is, if some four-dimensional objects exactly overlap at every time at which they exist, then they are identical.

The reflexivity and transitivity axioms (both in their synchronic and diachronic versions) are generally accepted by both endurantists and perdurantists. But they will diverge on most of the other mereological principles of interest. Thus, it makes sense to ask how endurantism, with its preferred mereological principles, compares with perdurantism, with its preferred mereological principles.

Below are some synchronic principles we may consider adding to endurantism:

Synchronic Antisymmetry. $\forall t\, \forall_t x\, \forall_t y\ (x \leqslant_t y \wedge y \leqslant_t x \to x = y)$.

Synchronic Supplementation. $\forall t\, \forall_t x\, \forall_t y\ (y <_t x \to \exists_t z\ (z \leqslant_t x \wedge z \perp_t y))$.

Synchronic Universalism. $\forall t\, \forall \overline{u}\ (\exists x\, \phi(x, t, \overline{u}) \to \exists x\ (x = \sigma_t y\, \phi(y, t, \overline{u})))$.

Synchronic Atomicity. $\forall t\, \forall_t x\, \exists_t y\ (y \leqslant_t x \wedge \mathsf{Atom}_t(y))$.

Synchronic Atomlessness. $\forall t\, \forall_t x\, \exists_t y\ (y <_t x)$.

Below are some diachronic principles we may consider adding to perdurantism:

Diachronic Antisymmetry. $\forall x\, \forall y\ (x \leqslant y \wedge y \leqslant x \to x = y)$.

Diachronic Supplementation. $\forall x\, \forall y\ (y < x \to \exists z\ (z \leqslant x \wedge z \perp y))$.

Diachronic Universalism. $\forall \overline{u}\ (\exists x\, \phi(x, \overline{u}) \to \exists x\ (x = \sigma y\, \phi(y, \overline{u})))$.

Diachronic Atomicity. $\forall x\, \exists y\ (y \leqslant x \wedge \mathsf{Atom}(y))$.

Diachronic Atomlessness. $\forall x\, \exists y\ (y < x)$.

It is an interesting question to what extent perdurantism and endurantism still remain schematically interpretable in one another when each side is equipped with its preferred set of mereological principles. As an illustration, when we add synchronic antisymmetry to endurantism and diachronic antisymmetry to perdurantism, we still have a schematic interpretation from the former to the latter.

Proposition 5.2.5 *(Interpreting an Antisymmetric Endurantism into an Antisymmetric Perdurantism)*. Define:

$$\theta = \forall x\ (R(x) \leftrightarrow \neg \exists y\ (\exists t\ (x \equiv_t y) \wedge \exists t\ (E(y;t) \wedge \neg(x \equiv_t y))))\,.$$

Then **3D**+***Synchronic Antisymmetry***$\multimap_{\Gamma \cup \{\theta\}}$**4D**+***Diachronic Antisymmetry***.

Proof: It suffices to show that the $\mathtt{t}$ defined earlier is still model-preserving relative to $\Gamma \cup \{\theta\}$. Surjectivity is trivial. For totality, let $\langle a, \lambda(a)\rangle, \langle b, \lambda(b)\rangle \in I^{\mathrm{Perd}}(R)$ and suppose $\langle a, \lambda(a)\rangle \equiv_t \langle b, \lambda(b)\rangle$ for some t. Then by definition, $a \equiv_t b$, so by ***Synchronic Antisymmetry***, $a = b$. Hence, $\langle a, \lambda(a)\rangle = \langle b, \lambda(b)\rangle$. Conversely, suppose $\langle a, t\rangle \notin I^{\mathrm{Perd}}(R)$. Thus, $t \subset \lambda(a)$, so $\langle a, t\rangle \equiv_t \langle a, \lambda(a)\rangle$. ■

Thus, **4D** with ***Diachronic Antisymmetry*** can interpret **3D** with ***Synchronic Antisymmetry***. But the converse does not seem to hold (though I have not yet found a proof that this is the case). Recall that in their more lax formulations, endurantism was a generalization of perdurantism, since it simply refused to impose the perdurantist temporal parts constraint. Thus, every perdurance model was automatically a persistence model. But it is not the case that every perdurance model satisfying ***Diachronic Antisymmetry*** is a persistence model satisfying ***Synchronic Antisymmetry***. In a sense, we would need to squeeze the persistence models into a tighter space to interpret **4D** with ***Diachronic Antisymmetry*** in **3D** with ***Synchronic Antisymmetry***. But this does not generally seem possible when some four-dimensional objects exactly overlap at one time but not another (as is possible according to perdurantism).[11] So while the two theories seem expressively on a par in the general case, it does not seem they are expressively equivalent when each is equipped with their preferred version of antisymmetry.

A similar kind of argument to the one above can also be used to show that for each of the principles above, when one adds the diachronic version to **4D** and the synchronic version to **3D**, the latter is interpretable in the former. This suggests an alternative characterization of the perdurantism-endurantism debate: the debate is not necessarily one about the different ways objects can persist through time, but is really a debate over the axioms governing temporal mereology (in particular, whether they should be construed synchronically or diachronically).

For example, perdurantists are often universalists about four-dimensional objects. That means that composition is completely unrestricted across both space and time: any collection four-dimensional objects, however far apart and disconnected in space and time, can be fused into a composite object. The perdurantist could, of course, hold that composition is only unrestricted across space (i.e., they could maintain ***Synchronic Universalism*** instead of ***Diachronic Universalism***). It is just that, given the picture of persistence they endorse where space and time are treated on a par, it would be unnatural for the perdurantist to restrict composition across time. By contrast, endurantists who also want to be universalists will often only hold that composition is unrestricted at any give moment. They will deny that, for example, Cleopatra and I compose a single object, for there is no single time at which all of the parts of this so-called object are present. Of course, there is no contradiction in adding ***Diachronic Universalism*** to endurantism. It is just not the most natural way for an endurantist to maintain unrestricted composition given the picture of persistence they endorse. Thus, whether or not the endurantist and perdurantist languages are expressively equivalent will depend on further commitments to principles that themselves might be disputed by each side.

Of course, the comparison does not stop there. To get a full picture of the relationship between perdurantism and endurantism, we would need to see how things change when we add further synchronic principles to endurantism and their diachronic counterparts to perdurantism. I leave such a comparative project to future work. But this does at least illustrate that the debate between perdurantism and endurantism is not *solely* over how objects persist. Rather, they are, at least in part, debates over whether temporal mereology should be done synchronically or diachronically. So even an anti-realist about the debate should concede that something still hangs on the debate, albeit indirectly. Even if we can choose which way to talk, these different ways of talking have very different structures that have an effect on our ability to adequately describe the world.

[11] We can generally restore interpretation if we replace the time-insensitive notion of identity in **3D** with a time-sensitive notion. For example, if we removed = from **3D**, we could interpret **4D** + ***Diachronic Antisymmetry*** in **3D** + ***Synchronic Antisymmetry*** by translating $x = y$ as $x \equiv y$.

Bibliography

Ackermann, Wilhelm. 1954. *Solvable Cases of the Decision Problem*. Amsterdam: North-Holland.

Barrett, Thomas William and Halvorson, Hans. 2016a. "Morita Equivalence." *The Review of Symbolic Logic* 9:556–582.

—. 2016b. "Quine's Conjecture on Many-Sorted Logic." *Synthese* 194:3563–3582.

Berto, Francesco, French, Rohan, Priest, Graham, and Ripley, David. 2017. "Williamson on Counterpossibles." *Journal of Philosophical Logic* 17:1–21.

Blackburn, Patrick, de Rijke, Maarten, and Venema, Yde. 2001. *Modal Logic*. Cambridge University Press.

Brogaard, Berit and Salerno, Joe. 2013. "Remarks on Counterpossibles." *Synthese* 190:639–660.

Caleiro, Carlos and Gonçalves, Ricardo. 2007. "Equipollent Logical Systems." In Jean-Yves Béziau (ed.), *Logical Universalis: Towards a General Theory of Logic*, 97–109. Basel: Birkhäuser-Verlag.

Carnielli, Walter A., Coniglio, Marcelo E., and D'Ottaviano, Itala M L. 2009. "New Dimensions on Translations Between Logics." *Logica Universalis* 3:1–18.

Carnielli, Walter A. and D'Ottaviano, Itala M. L. 1997. "Translations Between Logical Systems: A Manifesto." *Logique et Analyse* 157:67–81.

Chagrov, Alexander and Zakharyaschev, Michael. 1997. *Modal Logic*. Oxford: Clarendon Press.

Epstein, Richard L. 1990. *The Semantic Foundations of Logic. Volume I: Propositional Logics*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Etchemendy, John. 1990. *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press.

Feitosa, Hércules A. and D'Ottaviano, Itala M. L. 2001. "Conservative Translations." *Annals of Pure and Applied Logic* 108:205–227.

Font, Josep Maria. 2016. *Abstract Algebraic Logic: An Introductory Textbook*. London: College Publications.

French, Rohan. 2010. *Translational Embeddings in Modal Logic*. Ph.D. thesis, Monash University.

Gödel, Kurt. 1933. "Zur Intuitionistischen Arithmetik und Zahlentheorie." *Ergebnisse eines mathematischen Kolloquiums* 4:34–38. Reprinted in Gödel 1986, pp. 286–295.

—. 1986. *Collected Works. I: Publications 1929–1936*. Oxford University Press.

Grätzer, George. 2008. *Universal Algebra*. Springer Science & Business Media, 2nd edition.

Groenendijk, Jeroen A. G. and Stokhof, Martin J. B. 1991. "Dynamic Predicate Logic." *Linguistics and Philosophy* 14:39–100.

Groenendijk, Jeroen A. G., Stokhof, Martin J. B., and Veltman, Frank. 1996. "Coreference and Modality." In Shalom Lappin (ed.), *Handbook of Contemporary Semantic Theory*, 179–213. Oxford.

Heim, Irene. 1983. "File Change Semantics and the Familiarity Theory of Definiteness." In Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow (eds.), *Meaning, Use, and Interpretation of Language*, 164–189. Berlin: Walter de Gruyter.

Holliday, Wesley H. 2014. "Partiality and Adjointness in Modal Logic." In Rajeev Goré, Barteld Kooi, and Agi Kurucz (eds.), *Advances in Modal Logic*, 313–332. London: College Publications.

—. 2018. "Possibility Frames and Forcing for Modal Logic." *UC Berkeley Working Papers in Logic and the Methodology of Science.* Retrieved from https://escholarship.org/uc/item/0tm6b30q.

Hoogland, Eva. 2001. *Definability and Interpolation: Model-Theoretic Investigations*. Ph.D. thesis, Institute of Logic, Language and Information, University of Amsterdam.

Humberstone, Lloyd. 1981. "From Worlds to Possibilities." *Journal of Philosophical Logic* 10:313–339.

—. 2005. "Béziau's Translation Paradox." *Theoria* 71:138–181.

Jago, Mark. 2014. *The Impossible: An Essay on Hyperintensionality*. Oxford: Oxford University Press.

Jech, Thomas J. 2003. *Set Theory*. The Third Millennium Edition, Revised and Expanded. Berlin: Springer Monographs in Mathematics.

Jeřábek, Emil. 2012. "The Ubiquity of Conservative Translations." *The Review of Symbolic Logic* 5:666–678.

Kamp, Hans. 1981. "A Theory of Truth and Semantic Representation." In Jeroen A. G. Groenendijk, Theo M. V. Janssen, and Martin J. B. Stokhof (eds.), *Formal Methods in the Study of Language*, 277–322.

Kaplan, David. 1977. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals." In Joseph Almog, John Perry, and Howard Wettstein (eds.), *Themes from Kaplan*, 481–563. Oxford: Oxford University Press.

Karttunen, Lauri. 1976. "Discourse Referents." In *Syntax and Semantics*, 363–385. New York: Academic Press.

Kment, Boris. 2014. *Modality and Explanatory Reasoning*. Oxford: Oxford University Press.

Kocurek, Alexander W. 2017. "On the Concept of a Notational Variant." In Alexandru Baltag, Jeremy Seligman, and Tomoyuki Yamada (eds.), *Logic, Rationality, and Interaction: Proceedings from the 6th International Workshop*, 284–298. Berlin, Heidelberg: Springer.

Lewis, David K. 1986. *On the Plurality of Worlds*. Oxford: Blackwell Publishers.

Mares, Edwin D. 1997. "Who's Afraid of Impossible Worlds?" *Notre Dame Journal of Formal Logic* 38:516–526.

Mossakowski, Till, Diaconescu, Răzvan, and Tarlecki, Andrzej. 2009. "What Is a Logic Translation?" *Logica Universalis* 3:95–124.

Nolan, Daniel. 1997. "Impossible Worlds: A Modest Approach." *Notre Dame Journal of Formal Logic* 38:535–572.

Parsons, Josh. 2007. "Theories of Location." *Oxford Studies in Metaphysics* 3:201–232.

Pelletier, Francis Jeffry and Urquhart, Alasdair. 2003. "Synonymous Logics." *Journal of Philosophical Logic* 32:259–285.

Prawitz, Dag and Malmnäs, Per-Erik. 1968. "A Survey of Some Connections Between Classical, Intuitionistic and Minimal Logic." In H. Arnold Schmidt, K. Schütte, and H.-J. Thiele (eds.), *Contributions to Mathematical Logic*, 215–229. North-Holland, Amsterdam: North-Holland Publishing Company.

Priest, Graham. 2016. "Thinking the Impossible." *Philosophical Studies* 173:2649–2662.

Restall, Greg. 1997. "Ways Things Can't Be." *Notre Dame Journal of Formal Logic* 38:583–596.

Rothschild, Daniel and Yalcin, Seth. 2015. "On the Dynamics of Conversation." *Noûs* 1–25.

Schaffer, Jonathan. 2003. "Is There a Fundamental Level?" *Noûs* 37:498–517.

Schwarz, Wolfgang. 2018. "Semantic Possibility." In Derek Ball and Brian Rabern (eds.), *The Science of Meaning*. Oxford: Oxford University Press.

Sider, Theodore. 1993. "Van Inwagen and the Possibility of Gunk." *Analysis* 53:285–289.

Stalnaker, Robert C. 1976. "Propositions." In Alfred F. Mackay and Daniel D. Merrill (eds.), *Issues in the Philosophy of Language*, 79–91. New Haven: Yale University Press.

—. 1999. "Pragmatic Presuppositions." In *Context and Content*, 47–62. Oxford University Press.

—. 2012. *Mere Possibilities*. Princeton: Princeton University Press.

Straßburger, Lutz. 2007. "What is a Logic, and What is a Proof?" In Jean-Yves Béziau (ed.), *Logical Universalis: Towards a General Theory of Logic*, 135–152. Basel: Birkhäuser-Verlag.

Tarski, Alfred. 1944. "The Semantic Conception of Truth: and the Foundations of Semantics." *Philosophy and Phenomenological Research* 4:341–376.

—. 1983a. "Fundamental Concepts of the Methodology of the Deductive Sciences." In John Corcoran (ed.), *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, 60–109. Indianapolis: Hackett Publishing Company. Original work published in 1930.

—. 1983b. "On Some Fundamental Concepts of Metamathematics." In John Corcoran (ed.), *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, 30–37. Indianapolis: Hackett Publishing Company. Original work published in 1930.

—. 1983c. “The Concept of Truth in Formalized Languages.” In John Corcoran (ed.), *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, 152–278. Indianapolis: Hackett Publishing Company. Original work published in 1933.

Turner, Jason. 2011a. “Are Ontological Debates Defective?” In Robert Garcia (ed.), *Substance New Essays*. Substance Philosophia Verlag.

—. 2011b. “Logic and Ontological Pluralism.” *Journal of Philosophical Logic* 41:419–448.

Vander Laan, David A. 1997. “The Ontology of Impossible Worlds.” *Notre Dame Journal of Formal Logic* 38:597–620.

Varzi, Achille C. 2016. “Mereology.” In Edward N Zalta (ed.), *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/mereology/.

Warren, Jared. 2015. “Quantifier Variance and the Collapse Argument.” *The Philosophical Quarterly* 65:241–253.

Wigglesworth, John. 2017. “Logical anti-exceptionalism and theoretical equivalence.” *Analysis* 77:759–767.

Williams, John Robert Gareth. 2006. “Illusions of Gunk.” *Philosophical Perspectives* 20:493–513.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. London: Kegan Paul, Trench, Trubner & Co. Translated by Charles Kay Ogden.

Wójcicki, Ryszard. 1988. *Theory of Logical Calculi: Basic Theory of Consequence Operations*. Dordrecht: Kluwer Academic Publishers.

Woods, Jack. 2018. “Intertranslatability, Theoretical Equivalence, and Perversion.” *Thought: A Journal of Philosophy* 7:58–68.

Yalcin, Seth. 2007. “Epistemic Modals.” *Mind* 116:983–1026.

—. 2012. “Context Probabilism.” In Maria Aloni, Vadim Kimmelman, Floris Roelofsen, Galit W Sassoon, Katrin Schulz, and Matthijs Westera (eds.), *Logic, Language and Meaning: 18th Amsterdam Colloquium, December 2011, Revised Selected Papers*, 12–21. Berlin, Heidelberg: Springer.

List of Symbols

Chapter 1

List of Symbols

List of Symbols

Chapter 2

$\mathrel{-\!\!\triangleright}$	accessibility relation (modal logic), 56
$\mathbf{CPL}$	language of classical propositional logic, 55
CPL	classical propositional logic, 55
$\mathbf{FOL}$	language of first-order logic, 56
FOL	first-order logic, 56
$\mathtt{id}$	identity translation, 53
$\mathbf{IPL}$	language of intuitionistic propositional logic, 57
IPL	intuitionistic propositional logic, 57
$\mathbf{K3}$	Kleene's strong 3-valued logic, 60
$\Box$	necessity operator, 56
$\mathbf{LP}$	logic of paradox, 58
$\mathtt{Prop}$	proposition letters ($p_1, p_2, p_3, \ldots$), 55
$\mathcal{L}_{\mathtt{Prop}}$	propositional formulas, 55
$\mathbf{SOL}$	second-order logic, 59
$\overline{\alpha}$	sequence notation, 70
$\uparrow$	Sheffer stroke, 55
$\mathtt{ST}$	standard translation from modal logic into first-order logic, 56
$\mathtt{t}^{\mathrm{Th}}$	translation $\mathtt{t}$ lifted to theory spaces, 60
$\mathtt{t}, \mathtt{s}$	translations, 53
$\rightsquigarrow$	translation, translatable, 52
$\overset{\leftsquigarrow}{\rightsquigarrow}$	translation scheme, intertranslatable, 52
$\leftrightsquigarrow$	reversible translation scheme, translational equivalence, 63
$\cong$	translational isomorphism, 68
$\rightsquigarrow_{\Lambda}$	translation with auxiliary assumptions Λ, 70
V	propositional valuations, 55

Chapter 3

$:\subseteq$	definitional extension, 108
$-\!\!\bullet$	recursive translation, recursive translatability, 117

$\overset{\bullet\!-}{-\!\bullet}$	recursive translation scheme, recursive intertranslatability, 117
$-\!\bullet$	recursive translational equivalence, 117
$\overset{\bullet}{=}$	recursive isomorphism, 117
$\overset{\circ}{\subseteq}$	schematic fragment, 101
$-\!\circ$	schematic translation, schematic translatability, 98
$\overset{\circ\!-}{-\!\circ}$	schematic translation scheme, schematic intertranslatability, 99
$\circ\!-\!\circ$	schematic translational equivalence, 99
$\overset{\circ}{=}$	schematic isomorphism, 99
$\Theta^{\mathtt{t}}$	induced translation of schema Θ under $\mathtt{t}$ (only if $\mathtt{t}$ is schematic), 100
$\triangle^{\mathtt{t}}$	translation of $\triangle$ under $\mathtt{t}$ (only if $\mathtt{t}$ is schematic), 98

Chapter 4

$\propto_{\mathtt{t}}$	correlation relation for $\mathtt{t}$, 144
$\sqsubseteq$	discernible, 131
$\sqsupseteq\sqsubseteq$	discernibly equivalent, 131
$\sqsubset$	strictly discernible, 131
$\sqsubseteq_s$	strongly discernible, 138
$\sqsupseteq\sqsubseteq_s$	strongly discernibly equivalent, 138
$\sqsubset_s$	strictly strongly discernible, 138
$\leqslant$	expressible, 125
$\leqslant_{\mathsf{C}}$	expressible relative to C, 128
$<$	strictly expressible, 125
$\approx$	expressibly equivalent, 125
$\propto$	model-preserving, 143
∞	model-corroboration, 153
$\bigvee$	setwise disjunction, 135
$\neg$	setwise negation, 135
$\preceq$	weakly expressible, 129
$\sim$	weakly expressibly equivalent, 129
$\prec$	strictly weakly expressible, 129

Chapter 5

M	language of mereology, 168
AGEM	language of atomic general extensional mereology, **GEM** + ***Atomicity***, 169
$\tilde{\mathbf{A}}$**GEM**	language of atomless general extensional mereology, **GEM** + ***Atomlessness***, 169
EM	language of extensional mereology, **M** + ***Supplementation***, 169
GEM	language of general extensional mereology, **U** + ***Supplementation***, 169
$\mathbf{M}_{\leqslant}$	pure language of mereology, 168
N	language of mereological nihilism, 168
$\mathbf{N}_{\leqslant}$	pure language of mereological nihilism, 168
U	language of mereological universalism, 168
$\mathbf{U}_{\leqslant}$	pure language of mereological universalism, 168
M	class of mereological model-variable assignment pairs, class of mereological models, 168
$\mathcal{L}^1_{\mathtt{Pred}}(\leqslant)$	syntax for language of mereology, 167
$\leqslant$	part of, 167
Atom	mereological atom or simple, 167
$\perp$	disjoint from, 167
$\sigma x\, \phi(x)$	the fusion of all ϕs, 168
$\circ$	overlaps, 167
$<$	proper part of, 167
$\mathcal{L}^1_{\leqslant}$	syntax for pure language of mereology, 167
3D	language of endurantism, 188
4D	language of perdurantism, 188
D_t	objects that exist throughout t, 187
$I(\leqslant)_t$	interpretation of parthood restricted t, 187
$\equiv_t$	perfectly overlaps throughout t, 186
E	exists throughout, 186
λ	lifespan, 186

Index