Theory and Reality : Metaphysics as Second Science

Angere, Staffan

2010

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Theory and Reality

—

Metaphysics
As Second
Science

Staffan Angere

Department of Philosophy
Lund University

ISBN 978-91-628-8207-5
Printed in Lund, Sweden by
Media-Tryck in October 2010

For Lucius and Portia,
who were lost along
the way.

Preface

This is the day for doubting axioms.
With mathematicians, the question is settled; there is
no reason to believe that the geometrical axioms are
exactly true. Metaphysics is an imitation of
geometry, and with the geometrical axioms the
metaphysical axioms must go too.

—C. S. Peirce, "One, Two, Three:
Kantian Categories"

This book grew out of my curiosity about what the world is like in its most fundamental aspects. That curiosity got me interested in physics, and later in *meta*physics. At first, I was intoxicated by the contemporary metaphysics movement and its aims to free metaphysical reasoning from the shackles of epistemology and language. But, gradually, I became more and more disillusioned. It seemed to me that standpoints were generally accepted or rejected purely for psychological or social reasons, and the naturalist in me felt that such reasons simply were not relevant to questions of what the world is like.

In fact, as I discovered, much of contemporary philosophy is an *internal* affair: a debate is set up on certain premisses, and these are seldom questioned by the debating parties. As the debate proceeds, it takes on a life of its own, and defines its own norms for evaluating what is a good or a bad argument. Intuitions drive argument, and social groups form intuitions. In the end, the debate can move any distance from the—often quite concrete—questions that motivated it. Perhaps the most well-known philosophical school in which this is said to have occurred was Scholasticism; I suspect that much of what goes on in contemporary analytic and continental philosophy will be described in similar terms in the future.

The sciences—both deductive and empirical—are not similarly susceptible. They are generally constrained by fairly stable intersubjective methods of evaluation. These change slower than those of philosophy, so even if much science of the past has been given up, much of it also remains valid. Although purely social factors such as intellectual fashion do influence both the sciences and philosophy, the sciences are far less at their mercy. The greater subjectivity of traditional philosophy deprives it of its power to find out what the world is like, and the only way to regain that power, insofar as it is attainable at all, is to limit that subjectivity.

I have here tried to sketch an image of what an approach to metaphysics, as far as possible free of these defects, might be like. Ideas are gathered both from the sciences and the arts. On the one hand, this book is intended as a work of *scientific naturalism*, in that the proper methodology of philosophy is taken to be very similar to that of the sciences. On the other hand, the arts also have a large measure of objectivity by their role as image-providers, detached from questions of truth. An image is, in itself, not anything subjective, even if an interpretation of said image may be, and I believe the process of *imaging* to be crucial both to the sciences and to philosophy.

Within philosophy, I have mostly been inspired by the works of the giants of the 20th century: Carnap, Quine, Tarski and Wittgenstein among the dead ones, and Michael Dummett and Bas van Fraassen among those still living. Closer to me, I have received much inspiration and support from my supervisor Bengt Hansson, and also from professors Erik J. Olsson and Wlodek Rabinowicz of the Lund philosophy department. Furthermore, I would like to thank various attendants at seminars where parts of the book have been discussed, and my co-workers at the department, who were always ready to discuss my ideas, no matter how little sense they made: Robin Stenwall, Martin Jönsson, Stefan Schubert, Carlo Proietti, and many others.

An immense thanks goes to Sebastian Enqvist for meticulously reading through and commenting on the manuscript, for many enlightening discussions on the nature of philosophy and logic, and for support and friendship. But most of all, I owe infinite and unending gratitude to my wife Saga, and our cats Cassius, Ophelia, Hamlet and Othello, for keeping me sane enough to do philosophy, and for always being there for me. This book could not have been written without you.

Contents

Introduction

Metaphysics, despite being philosophy's most venerable strain, also remains one of its most questioned and criticised. For most part, this criticism is well motivated. Metaphysics was supposed to tell us about the fundamental constitution of reality, but since at least the 17th century, that has been the work of theoretical physics, and not philosophy. While physics has gone from success to success, metaphysics has seen very little actual progress since Plato: modern metaphysicians still concern themselves with problems of universals, instantiation, substance, essences, and the rift between appearance and reality. It is easy to draw the conclusion that metaphysics, as a research programme, has gone into regression, and that the parts of it that were once viable have been taken over by the sciences.

Why did this happen? The seeds of the collapse were sown already in the battle between the British empiricists and the continental rationalists during the 17th and 18th centuries. It is safe to say that the progress of science granted victory to the empiricists. Certainly, there was the Kantian programme of trying to show that empirical knowledge was confined to the world of appearances, and that a transcendental metaphysics was necessary to grasp reality as it really is. But the fact remains that the world of things–for–us is what we are immersed in, and it is this world that most directly piques our curiousity. That there may be another world behind the veil of appearance may be an intriguing thought, but perhaps more so for science fiction and theology than for science and philosophy.

Accepting that metaphysics studies the world of things as they are accessible to us should, however, not be confused with the quite different programme of analysing our "common sense" metaphysical concepts, mostly associated with Strawson's descriptive metaphysics (Strawson, 1959). Metaphysics, as it interests us in this book, is a subject purportedly dealing with what the world is like, and not primarily about our concepts. If there is a viable notion of descriptive metaphysics, apart from the psychological (and empirical) investigation into how we represent things mentally, we will not have much to say about it here. Our target is the real world, and what we think about it only serves as a stepping-stone, since these thoughts say something about the world only if what they say happens to be true.

The best methods for finding out what about this world is true or false are empirical, so it is easy to see why traditional metaphysics in the vein of the presocratics, Plato, Descartes and Leibniz must fail. "Armchair philosophy", as its detractors call it, is rationalistic, and though no metaphysician would categorise herself as an armchair philosopher, the fact remains that it is very rare for metaphysicians to do actual empirical experiments, or even to design or propose them, and so the armchair remains her weapon of choice.

We therefore ought to ask ourselves if we need metaphysics at all. What use is there for it, given that the sciences seem so much better at finding out about the world? This way of seeing the problem pits metaphysics *against* the sciences, as if they were two exclusive tools for finding out about the same thing. In a way this is true: both metaphysics and the sciences are about what the world is like. But it is also often held that there are important differences. Metaphysics is sometimes said to be concerned with the more "abstract", or the more general features of reality, while the sciences are held to be more specialised. Yet, physics certainly is as general as anyone could wish (it applies to *all* interactions, since if we find some interaction that it does not subsume, we will see that as an incentive to change our physics), so we still have no explanation why metaphysics does not conflict with physics.

Proceeding along Kantian lines, we may be drawn to the view that metaphysics should be transcendental, and investigate the *presuppositions* of the sciences. This is the road that leads to metaphysics as "first philosophy" — that which is required to justify the sciences. It is, of course, a descendent of Descartes' search for a secure foundation for all our knowledge. The sciences, however, seem to have proceeded quite well without such a foundation, and it is very doubtful that one will be found, or that even if one *is* found, it will be relevant to our scientific concerns. First philosophy, should it be possible at all, is of doubtful interest. The proper answer to the problem seems to be to reverse the priorities. Rather than first philosophy, metaphysics's proper place is as *second science.* It presupposes the sciences, and should work with their results, rather than attempt to justify them.

But how do we know that there is any meaningful work left to do, after the sciences have put forward their theories? We would have to go fairly deep into the philosophy of science to answer this question. It is worth noting, however, that instrumentalism did loom large in much of 20th century science. Theories are selected due to their predictive power, and we are regularly reminded not to read any kind of substantial claims into them. Philosophical versions of this view include the positivism of the logical empiricists, as well as van Fraassen's constructive empiricism (van Fraassen, 1980), in which commitment to a theory is taken to be commitment to its empirical adequacy, and empirical adequacy is explicated as truth of the observable parts of the theory. According to constructive empiricism, science does *not* commit itself to the whole of theories being descriptive of reality.

Science, in so far as its goals are instrumentalist, does leave room for metaphysics. Where the sciences claim that no more can be said because there are no empiciral tests that could settle the matter, metaphysics presumably could pick up the reins and investigate further. We can even envisage cases where its results may trickle back down into the sciences; models of natural phenomena created by metaphysicians, since they cannot conflict with the empirical data, are also models that are are available for use in the sciences. This means that, as far as they are described in scientifically useful terms, they can be used by scientists as well.

Metaphysics done *within* the sciences is often like this. As an example, we may take Minkowski's model of Einstein's special theory of relativity in terms of what we now call *Minkowski space* (Minkowski, 1908). Although such a model does not, by itself, supply any new testable consequences, and so is a "metaphysical" theory in the positivistic sense, its importance for understanding the theory of relativity cannot be overestimated. Almost all current textbooks on special relativity present it in terms of Minkowski space, and not in the more phenomenological terms that Einstein first gave it (Einstein, 1905). It is also safe to say that without the picture of Minkowski space, the question of other metrics—for instance those that are associated with curved space-time—would never have arisen, and so we would have had no general theory of relativity either.

Another example, also from physics,[1] is Bohm's "hidden variable" interpretation of quantum mechanics (Bohm, 1952). This interpretation is specifically designed not to give any new testable consequences, but only to provide a sort of framework, seen from which quantum mechanics makes sense in a classical manner. It *has* been criticised because of its lack of testable consequences, but this kind of criticism seems to me to miss the point. Its most important problems spring rather from the difficulty of adapting it in a natural way to newer theories, such as quantum field theory. Comparing Bohm's interpretation of quantum mechanics to Minkowski's space-time model of relativity, we may note what the second has, and what the first lacks, which makes Minkowski's metaphysical theory successful, and Bohm's unsuccessful so far. Minkowski spacetime, when used as a framework or a model, allows us to frame new theories which are impossible to frame without it, and which experiment have verified. Bohm's, on the other hand, makes the framing of an experimentally corroborated theory (quantum field theory) almost impossible, or at least very hard. The point is

[1] I am well aware of the tendency of philosophers of science to take almost all their examples from physics, to the neglect of all the other sciences, and I regret to say that I will be following suit here. Part of the reason for this is because physics is the science I am most familiar with, but it is also the case that physics, as being concerned with the most general and fundamental aspects of reality, holds special interest for metaphysics. Thus, while I in no way wish to promote the hegemony of physics among philosophers of science, I believe that it is somewhat more excusable when we are dealing with metaphysical questions.

pragmatic: Minkowski spacetime, as a model, has a *theoretical usefulness* that has so far not been found to be shared by the Bohmian interpretation of quantum mechanics.

A fundamental point of note here, for metaphysics, is that all *mathematics* may be done from the armchair (or at least from a desk with a computer), and few question the worth of that — even its higher reaches, whose applicability to empirical science may seem distant. Perhaps metaphysics could be more like this? Mathematics concerns itself with the design (or investigation, if you are a mathematical Platonist) of abstract structures. These are often applicable to empirical phenomena both in common-sense world views and the sciences. Can it be that metaphysics, as well, can be seen as such a process of structure-creation, with the actual fitting of structure to reality being left for the sciences?

This will indeed be the method primarily explored in this book. Metaphysics, as I see it, is a branch of *model theory*, in an extended sense of the word in which it stands for the discipline that studies the semantical correlates of theories and languages. Model theory, like classical metaphysics, is largely *a priori*, and does not purport to tell us, on its own, what reality is like. For this, it needs *semantics*, which is what connects it to theory, and an actual *theory*, which is what science supplies us with. All of these notions have their own problems, and all will concern us here. Our guiding methodology will however remain the *theory – semantics – model* connection, and our intention is to show how this may be put to use, in order to arrive at a conception of metaphysics that is both viable and scientifically respectable.

The first chapter contains an overview of various approaches to metaphysics. Starting with Quine's programmatic *On what there is*, the first chapter then discusses the perils involved in going from language to metaphysics. It criticises contemporary intuition-driven metaphysics, comments on naturalistic approaches, and then presents the main proposition put forward in the thesis: we should base metaphysics on *model theory.* But a model, logically speaking, is a mixture of interpretation and metaphysics. Therefore an important task is to separate these parts of it.

Chapter 2 introduces *theories*, which are defined as consequence operators on sets of truth-bearers. These can be used both for making claims, and for framing other theories. I avoid use of any analytic/synthetic or logic/material distinction. Some generalisations and specialisations of the concept are discussed, among which are algebraisation and probabilistic theories. Chapter 3 gives an abstract characterisation of *metaphysics* using category theory, and also contains examples of different kinds of metaphysics, and remarks on how these relate to one another. The central notion here is that of model, and a metaphysics, as a collection of ways something (e.g. the world) could be, is identified with a category of models.

In Chapter 4, we encounter a specific sort of model, based on a nondeterministic necessitation relation. These models (which I call *necessitarian* models) have roughly the same structure as a multiple-conclusion logic, and make up a very useful type of metaphysics, which will be used later in the book to derive theorems on the relation between theory and reality. Generalisations involving probabilistic necessitation are discussed, and questions of how to interpret these models in terms of more traditional metaphysical concepts are broached.

Chapter 5 is named "Semantics", and here we discuss various ways for theories to relate to models. One way, which fits well with necessitarian models, is based on the idea of *truthmaking.* Starting out from a simple satisfaction relation between models and truthbearers, we show that there are systematic ways to identify specific parts of models as truthmakers. These concepts are used in chapter 6, where we derive an isomorphism between the logical structure of a theory and the necessitation-structure of a metaphysics. This isomorphism allows us to go from theory to world, and thus gives us an answer to the question of what this relationship is.

The final chapter and the epilogue contains applications and a conclusion. We look at how the theory-world isomorphism can be used to answer questions about the philosophy of logic, mathematics, quantum mechanics, and philosophical problems of mind and metaethics. Questions dealt with include the relation of intuitionistic to classical logic, Platonism in mathematics, and the Bohr interpretation of quantum mechanics. We then take a step back, and consider some truly fundamental questions: in what way is *the world* a model? How should we do metaphysics? And, what considerations should we take into account, when we settle on a way to describe the world?

Two major influences on *Theory and Reality* are the conventionalism of Carnap, and the ontological relativism of Quine. These strains are combined with the Dummettian insight that *logic* and *metaphysics* are intimately related. Parts of the book are fairly heavily couched in the language of mathematics, although I will make no apology for this. Mathematics (and the part of it called *logic*) as I see it has as central a place in philosophy as it has in physics or economics. It supplies us with ways of thinking that can lead to much greater clarity and exactness than any other method. It provides us with common languages for communication, and it gives the often diverse opinions of various philosophers a common ground: there is usually very little variation in opinion over the validity of a mathematical proof, compared to a traditional philosophical argument.

However, this is not a thesis *about* mathematics. There are no really "deep" theorems in it, so I have avoided the practice of demarcating a ruling class of "theorems" from an underclass of "propositions" or "observations". The formal requirements (except where I discuss quantum mechanics) are only knowledge of first-order logic and Zermelo-Fraenkel set theory, but as always, fulfilling the formal requirements does not make everything easy. The reader is invited to skip parts she finds difficult on a first reading. Altogether, the book is an application of mathematics to philosophy. This, of course, invites the criticism that it *misses* something: that there are things that cannot be treated this way, and that applied mathematics is insufficient for metaphysics. This type of criticism is not new; Duhem quotes a "Cartesian" in 1740, commenting on Newton, as follows:

> Opposed to all restraint, and feeling that physics would constantly embarrass him, he banished it from his philosophy; and for fear of being compelled to solicit its aid sometimes, he took the trouble to construct the intimate causes of each particular phenomenon in primordial laws; whence every difficulty was reduced to one level. His work did not bear on any subjects except those that could be treated by means of the calculations he knew how to make; a geometrically analyzed subject became an explained phenomenon for him. Thus, this distinguished rival of Descartes soon experienced the singular satisfaction of being a

> great philosopher by sole virtue of his being a great mathematician. (Duhem, 1954, p. 49)

We do not see Newton as a "philosopher" at all any more, and nowadays we tend to see science and philosophy as crucially different. Still, I believe that the best kind of philosophy will always be the kind that lies close to science, and the best kind of science the one that touches on philosophy.

Finally, I would like to make a remark on various references to historical philosophers I that have used here and there. These are not to be taken as expositions of what the philosophers in question meant, or how they should be interpreted. Just as this book is not a book about mathematics, it is not one about the history of philosophy either. But just as mathematics, the history of philosophy furnishes us with a common conceptual framework. It can therefore be very useful for communication of ideas and for making comparisons and drawing analogies.

Chapter 1
What Metaphysics Can and Cannot Be

In this chapter, we give a brief overview of various approaches to metaphysics. We start with Quine's approach from *On what there is*, and try to gauge its strengths and weaknesses. The most important of these weaknesses will be found to be its close ties to first-order logic. The second section continues this thread, and deals with general problems inherent in inferring facts about the structure of the world from the structure of language. While language and world might not be completely separate, we have no reason to believe that they coincide completely either.

Section 3 discusses and criticises the currently common tendency to rely on intuition for metaphysical theorising. In contradistinction, I hold that intuition has no place at all in metaphysics, and ideally should play no role. This opens up the question of how to proceed, given that projecting language onto the world and employing intuition are both to be avoided. Section 4 treats possibilities for naturalism: the idea that philosophy should avail itself of roughly the same methods as the sciences. However, this turns out to be hard to do in practice.

Finally, I introduce the view of metaphysics that I prefer: metaphysics as model theory. For this purpose, we need a notion of "model" that lies somewhere between how it is used in logic, and how it is used in the sciences. I give some general remarks on what this kind of model theory might be, and then go into the question of how to connect theory to reality through model theory. This is to be done by the use of the concept of *truth*, and I therefore take up the question of what we are to mean by this word, and what role it plays for us.

1.1 The Last Great Metaphysician

Scientifically, the last progressive research programme in metaphysics was initiated by Quine in *On what there is* (Quine, 1948). Very freely summarised, the Quinean strategy for metaphysics (or ontology, which is the part of metaphysics he discusses) is as follows:

(*i*) Look to science for what theories of the world we have reason to believe are true.

(*ii*) Formalise these theories in classical first-order predicate logic with identity.

(*iii*) What we should believe exists is what the values of the bound variables in these formalisations have to range over in order for the theories to be true.

We have given the first step in terms of which theories are to be believed *true*, instead of the customary rendering "our *best* theories". Given Quine's pragmatism, the difference may be slight, but focusing on truth instead of "goodness" lets us avoid a problem noted by Melia: we have reason to believe many of our current best theories to be *false*, and thus these cannot be used for finding *actual* ontological commitments (Melia, 1995). It is better to let scientists (or possibly theorists of science) decide what theories are true as well, and treat this as given for the metaphysician. With this modification, it also becomes evident that the primary task of metaphysics (or ontology) is not to find out new truths, but rather to interpret (or in some cases *re*interpret) old ones.

The other side of the coin is that if most of our best theories are false, then it seems like we have very little to go by, if we are to apply Quine's methodology. This is not so, however. Many theories may be false, but they still contain subtheories (for instance, those dealing with the theory's observable consequences in given situations) that we have good reasons to believe to be true. This is why we have said that we should "look to" science for true theories: not every scientific theory is useful for finding ontological commitments, but almost all such theories contain theories that are.

The second step is where the metaphysician's ingenuity comes into play. Formalising a theory is somewhat like translating poetry. It is as much a creative as a deductive task, and different formalisations may be compared according to several criteria. Quine's first interest here was *parsimony.* If a formalisation F does not require quantification over some entities X and formalisation G does, but F and G are both adequate formalisations of the same theory, that theory itself is *not* committed to the entities in X. More specifically, if G is *reducible* to F, but F is not reducible to G, only the values of F's bound variables are among the theory's ontological commitments.[1]

On what there is thus in essence contains the basics of a research programme for metaphysics. It contains a methodology (briefly as described above) and principles for evaluation of theories, in terms of the sizes of their ontological commitments. Much good metaphysics was done in it, from Quine's own disentangling of Plato's beard in 1948, to Lewis's reduction of ZFC set theory to mereology and a primitive singleton operator in 1991 (Lewis, 1991). Lately it has become less and less prominent, although the principle that to quantify over something is to acknowledge its existence is often adhered to still, as we do not have any other criteria for ontological commitment that are as clear as

[1] The condition that F should not be reducible to G is necessary here. Two theories may be reducible to one another without being the same theory, or even logically equivalent. In such a case, it seems that neither the formalisation F's nor G's ontological commitments could be those of the theory.

Quine's.

There are probably as many reasons for this decline of Quinean metaphysics as there are metaphysicians. The most important, as I see it, is the primary place it grants to first-order predicate logic, with its standard referential semantics. This is quite arbitrary, as I shall argue by posing a few questions, in approximately increasing order of generality, about the choice of logic and semantics.

Why referential semantics? The standard Tarskian semantics of first-order logic is only one of the multitude that are conceivable. For Frege, for instance, semantics involved relations between signs and *functions* and *arguments*, rather than just objects and sets thereof. Using a Fregean semantics therefore would commit us to the existence of functions, no matter if we succeed in reducing them away or not.

We also have the various sorts of substitutional semantics, defended by Ruth Barcan Marcus (1961) and Peter Geach (1963). Interpreted this way, quantification commits us to nothing but the singular terms that may occupy the variable positions. Quine, of course, is critical to such attempts, since he takes the *fundamental* notion of variable to be the referential one:

> The variable *qua* variable, the variable *an und für sich* and *par excellence*, is the bindable, objectual variable. It is the essence of ontological idiom, the essence of the referential idiom. (Quine, 1972, p. 272).

However, he does not *disallow* use of substitutional quantification altogether. Rather, we have to translate a substitutionally-quantified theory into the "referential idiom" for us to be able to find the theory's true ontological commitments (Quine, 1969, p. 106). But, what if we simply *avoid* using the referential quantifiers in constructing our theory, and have no rules in mind for translating the theory into one that uses referential quantifiers either? Quine's method ceases to be applicable in such a case, and yet we may have good reason to hold substitutional theories to be true or false, and so to say something about reality.

Why first-order logic? Quine famously held second-order logic to be "set theory in sheep's clothing" (Quine, 1986, p. 66). Yet, to both

Frege and Russell, higher-order logics were not separate forms of logic at all, but just as logical as the first-order kind. More recent advocates of second-order logic such as Boolos (1975, 1984) and Shapiro (1991) have argued that limiting logic to the first-order kind is unnecessary and arbitrary, since, for instance, monadic second-order logic even is decidable (Skolem, 1919).

There are also other forms of quantification available, such as Henkin's branching quantifiers (Henkin, 1961) and Hintikka's independence-friendly logic (Hintikka, 1996). And while standard first-order logic, as Quine puts it, may possess "an extraordinary combination of depth and simplicity, beauty and utility" (Quine, 1969, p. 113), the question remains as to why these properties should make it the canonical vehicle for ontological commitment as well.

Why predicate logic? This may, at first, seem like a strange question. Standard sentential logic is not expressive enough for the needs of science, and so our interest in finding the ontological commitments of actual theories seems to force us into this choice. But it is still a problematic one, since predicate logic, especially with identity, is far from neutral when it comes to metaphysics. Vague objects, for instance, are ruled out, and also entities without identity conditions. Relations between infinitely many entities require set theory to be representable. More fundamentally, there is a kind of metaphysics inherent in predicate logic, in which self-subsistent *objects* have *properties* and stand in *relations*. While this very well may be a workable metaphysics, it is still a choice that should not be made in the logic, as it excludes alternatives without giving them a fair hearing. Ladyman and Ross (2007), for example, argue that contemporary physics is incompatible with the notion of a world of self-subsistent individuals. By tying ourselves to predicate logic with identity, we rule out such arguments beforehand.

Why classical logic? Despite Quine's insistence in *Two Dogmas* on the revisability of even the laws of logic, he remained a defender the sufficiency of its classical variant to his end. Yet, seeing the explosion of alternative systems from the 70's onward, with modal, many-valued, substructural, nonmonotonic, and constructive variants to mention a few, each with seeming applicability to their own areas, one cannot help but feel what a strait-jacket this is. The use of intuitionistic logic, for instance, does not necessarily have to make the idea of ontological commitment otiose, as we shall see in chapter 7. A methodology for metaphysics should ideally be neutral on the question of what logic, if any, is the "correct" one.

Why logic at all? Quine's idea is to let scientists determine what exists, but these do not, generally, express themselves in formal logics at all. Indeed, *any* thing that can be true or false (i.e., that purports to describe reality) seems to be possible to raise questions of ontological commitment over. Beliefs, diagrams, depictions, equations, speech acts, and natural-language discourse are all ways in which scientists represent their theories, and forcing this into the mold of a given logical system takes both creativity and skill. It also opens the question of whether the formalised version of the theory is *equivalent* to the pre-formalised one, since otherwise it will be of no use for determining the theory's ontological commitments. The more difference between the expressive strength of the theory's "natural" representation and the logical system we use, the harder this equivalence will be to establish.

As an example, we may take the difference between classical logic and English. Since Montague's papers on the semantics of natural language (Montague, 1970, 1973), it has been accepted that we *can* study the inferential properties of ordinary language discourse without prior translation into a formal language. But *non-formal* systems, such as those that admit of analytical consequence, generally lack the property of *structurality* (see section 2.4), which is commonly taken to be necessary for a notion of consequence to be *logical* (Wójcicki, 1988). In taking something else than logical form as grounds for consequence, we are therefore leaving the confines of logical systems. But since scientific theories in general at least depend on analytical consequence, we may want a methodology that accepts this habit as it is.

These questions all highlight the fact that Quine's reliance on first-order logic is a very real *limitation* on the applicability of his methodology.

But there are also other considerations: according to Quine, it is only the quantified variables that commit us to anything, so sentential-logical theories, for instance, have no ontological commitments at all. But what we take as quantifiable and what is not is to some extent up to us. Consider a language $\mathcal{L}$ for discussion about worlds in which where there are two objects, a and b, each of colour Red or Blue. The predicate-logical languages $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$ of table 1.1 are all versions of this language:

	Individuals	***Predicates***
$\mathcal{L}_1$	*the world*	Is Such That a Is Red & b Is Red (x), Is Such That a Is Red & b Is Blue(x), Is Such That a Is Blue & b Is Red(x), Is Such That a Is Blue & b Is Blue (x)
$\mathcal{L}_2$	a, b	Red(x), Blue(x)
$\mathcal{L}_3$	$a's$ *redness*, $a's$ *blueness*, $b's$ *redness*, $b's$ *blueness*	Is Instantiated(x)

Table 1.1: ***The languages $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$.***

Although $\mathcal{L}_2$ may seem the best choice among these, in terms of perspicuousness, Quine's preference for formalisations with minimal ontologies (his taste for "desert landscapes") recommends $\mathcal{L}_1$. The problem is that when we formalise, we generally have to make a trade off between ontological commitment and what Quine calls *ideology* – the predicates that our language must contain for the theories we are interested in to be expressible in them. The Quinean methodology's reliance on ontological commitment only captures one side of this trade off.[2]

[2] The opposite position—that only what predicates we use determine a theory's simplicity—is defended by Goodman in *The Structure of Appearance* (Goodman, 1951, pp. 59–63). David Lewis seems to place himself somewhere in the middle, since he argues that it is not commitment to entities that is to be avoided, but commitment to *kinds* of entities (Lewis, 1973, p. 87).

It is well known that Quine later came to distance himself from the metaphysical research programme that he incited. The main reason for this was his doctrine of *ontological relativity* (Quine, 1969), according to which a theory never *by itself* has an ontological commitment, but only in relation to some theory we may reduce it to. This is a corollary to his thesis of indeterminacy of translation from *Word and Object* (Quine, 1960b): in cases like that of the field linguist, not only the meaning of "gavagai" is indeterminate, but also its reference. This means that in order to secure reference for our terms, we need a system of *analytical hypotheses*—a kind of coordinate system that may be used to fix the references. The upshot is, as he puts it that "it makes no sense to say what the objects of a theory are, beyond saying how to interpret or reinterpret that theory in another." (Quine, 1969, p. 50)

The framework for metaphysics I am going to defend in this book will be compatible with the truth of ontological relativity, as I think it must be, if we are to remain naturalists when it comes to the philosophy of language.[3] But there is still work left for metaphysics to do, for metaphysics does not have to be *just* ontology, in Quine's sense. For one thing, we may have things that are common to *all* frameworks that a theory can be interpreted in. These would permit us to infer something about what the theory says exists, since just because theories do not have unique ontologies by themselves, that does not mean that *any* ontology would be acceptable for any theory. Instead of a single ontological commitment, we would have a *class* of ontological commitments compatible with the theory.

Secondly, it is also the case that not all systems of analytical hypotheses are equally interesting. In general, we are not interested in a theory's ontological commitments *per se*, but rather in its ontological commitments *as seen from our current theoretical framework.* The posing of a metaphysical question usually supplies us with a system of analytical hypotheses, which we can use for our answer.

The conclusion we arrive at is thus that Quine's methodology cannot

[3] It might occur to some current metaphysicians to take the problems of referential inscrutability to be soluble by use of the causal theory of reference. This merely pushes the problem back, however; instead of analytical hypotheses, we now need metaphysical hypotheses about the causal network of the world, in order to fix a term's reference.

be pursued, as it was laid out in *On what there is.* This does not mean that we cannot draw important lessons from it, and that some variant of it may be viable. The view of metaphysics I propose in section 1.5 may be seen as such a variant, since it shares many of Quine's fundamental standpoints, while trying to avoid some of its problems.

1.2 The Perilous Seas of Language

For Quine, as well as for Russell before him, studying the logical structure of language was a way to find out about the structure of the world. Yet, the supposed connection has also come under heavy fire recently. John Heil attacks what he calls the *Picture Theory*, and its use as a guiding principle:

> What exactly is the Picture Theory? As I conceive of it, the Picture Theory is not a single, unified doctrine, but a family of loosely related doctrines. The core idea is that the character of reality can be 'read off' our linguistic representations of reality—or our suitably regimented linguistic representations of reality. A corollary of the Picture Theory is the idea that to every meaningful predicate there corresponds a property. If, like me, you think that properties (if they exist) must be mind independent, if, that is, you are ontologically serious about properties, you will find unappealing the idea that we can discover the properties by scrutinizing features of our language. This is so, I shall argue, even for those predicates concerning which we are avowed 'realists'. (Heil, 2003, p. 6)

The picture theory is thus, at bottom, a theory about language. As such, it is of course not only criticised by metaphysicians, but also by philosophers of language. Ryle, to mention an influential example, calls it *the 'Fido'–Fido fallacy* (Ryle, 1957) — the idea that every part of a sentence corresponds to a part of reality. Austin (1950) explicitly distances himself from picture-type correspondence theories of truth, such as that of the *Tractatus*, and holds the correlation of sentences to the world to be purely conventional. And one of the view's strongest critics is Wittgenstein himself, who opens his *Philosophical Investigations* with a parody of it, as he finds it in Augustine's *Confessions* (Wittgenstein, 1953).

An instance of the Picture theory's influence is the tendency to base one's metaphysics on the subject-predicate distinction: many philosophers have held the contents of the world to be divided into *individuals* and *properties* such that the first of these instantiate the second. But, as Ramsey pointed out, it might very well be that this distinction is purely grammatical. Indeed, *all* singular terms could be like Quine's "sake", which *looks* like a name for an object, but is not reasonably taken to function as one (Quine, 1960b, p. 244). A more subtle influence of the picture theory can be seen in the idea that because "object" works as a count noun, the world has to contain a certain number of self-sufficient, well-individuated objects. I am not saying that any of these theories are *wrong*, but we should not infer their truth from the workings of language.

Yet, the picture theory has a very clear appeal. Contemporary formal semantics is very much based on the picture metaphor: words mean by referring to things (or functions, or sets, or other kinds of entities), and the meanings of sentences are determined functionally by the meanings of the words that they are made up from and their mode of composition. Through first Carnap and later Montague it has been extended to natural languages as well, and it seems to give some kind of understanding of how language works. If "Paris is north of Pisa" means that a certain thing (Paris) stands in a certain relation (being north of) to another thing (Pisa), then this should be true iff the original sentence is true. This in turn means that, since "Paris is north of Pisa" is true, "the thing *Paris* stands in the relation *being north of* to the thing *Pisa*" must be true as well. But this second sentence has a definite air of metaphysics.

Maybe we have moved too fast here. Does "the thing *Paris* stands in the relation *being north of* to the thing *Pisa*" really say *more* than "Paris is north of Pisa", so that it does not follow from this obvious truth? That would have to depend on how we interpret the two sentences: there is definitely a reading of them on which they are equivalent. But the whole point of expanding "Paris is north of Pisa" in terms of things and relations was to give meanings! How could there be a question of what the second sentence means then?

The truth is that no sentence ever interprets itself. "the thing *Paris* stands in the relation *being north of* to the thing *Pisa*" is as much in need of interpretation as "Paris is north of Pisa", and admits as many different types of metaphysics as it. The meaning, conceived as reference or as a condition on worlds, is inherent neither in the words themselves, nor in our usage of them.

Carnap, as one of the modern founding fathers of this kind of meaning theory, was well aware of this. In *Meaning Postulates*, he explicitly treats questions of how to assign intensions as one that is free for us to decide on (Carnap, 1956, pp. 222–229). His whole method of linguistic analysis in *Meaning and Necessity* is presented as motivated by usefulness, rather than any connections to what meanings "really" are. Referential as well as intentional semantics is a doubly conventional matter: not only is the usage of a word or a sentence decided by social conventions, but how this usage is to relate to the world is conventional as well.

Similar lessons can be extracted from Putnam's famous Twin Earth example, although Putnam himself certainly did not intend to draw them. The people on Twin Earth behave in exactly the same way as those on Earth, so *use* in the narrow sense of *behaviour* will not determine reference. But reference concerns what the world is like: "water is XYZ" is true iff "water" refers to a and "XYZ" refers to b, and a is identical with b. So linguistic behaviour does not determine what sentences say about the world.

It is common to suppose that what is missing between use and reference is causal or ostensive: what is in the mind does not determine reference, but what the world around the user is like does. But this is not a link that is permissible for us to use when we are to do metaphysics, since what the world is like is just what we want to find out. A causal theory of reference may possibly be useful if we already have a metaphysics and are trying to design a theory of language, but it can do no work when we are to go from language to metaphysics. Thus the relevance (or lack thereof) of causal reference to meaning is quite beside the point for us.

We do not have to say that meaning in general goes beyond use, however. As our focus here is on the theory–world connection, we can allow that this is underdetermined by use, without saying anything about whether "meaning" is so underdetermined or not. Accordingly, we will try to avoid using the word "meaning" altogether, instead employing "use" when it is this aspect that concerns us, and "semantics" for the connection between words or theories and the world.

Thus we will drive a wedge between linguistic usage, and language's possible connections with reality, in order to be able to study the second on its own terms.[4]. In this we follow Heil and other critics of "linguistic philosophy". But that the naive picture theory is false does not necessarily mean that mean that linguistic or logic analysis can tell us nothing at all about the world. Our linguistic usage does not float *entirely* free of what the world is like, even on more plausible accounts of language. That we should not impose one on the other does not mean that they are completely separate.

All we have to be careful about is to not confuse linguistic structure with metaphysical. A fundamental insight of the linguistic turn—that it is primarily with language that we connect with reality, and that the analysis of language therefore is *necessary* for understanding—remains untouched. That it is not sufficient is of course always worth pointing out. The structure of language is not the structure of reality, although there of course has to be *some* relation holding between the two, for language use to be possible at all. If nothing else, linguistic behaviour is as much a part of the world as any other kind of behaviour.

[4] This somewhat parallels Russell's important but neglected division between a word's *logical significance* and its *meaning in use* in *The Philosophy of Logical Atomism* (Russell, 1985, p. 142)

1.3 What's Wrong with Intuition?

Implicitly referring to Quine as "the last great metaphysician", as I did in the first section of this chapter, may seem almost perverse to some contemporary philosophers. Quine's metaphysical theorising is very limited in scope, concerning itself mainly with questions of ontology, and as we mentioned, he came to take exception to even that later on. Yet, his programme *did* supply inspiration for a generation of metaphysicians. Contemporary metaphysics, however, is generally much more indebted to the methods of Kripke. Above all, his insistence that we separate epistemology from metaphysics (for instance in his distinction between the *a priori* and the necessary (Kripke, 1981, pp. 34–39)) has inspired philosophers to proclaim the independence of metaphysical reasoning both from questions of knowledge and of language.

This would perhaps be fine if there clearly *was* such a thing as metaphysical reasoning. The problem is that when we sever the ties to language, logic and knowledge, it is hard to know what counts as a valid argument anymore. Do we *really* know that reality does not contain contradictions, for instance? A contradictory position may be epistemologically unacceptable, but how do we determine it to be *metaphysically* so?

Kripke, however, also supplies us with an *evaluative* principle: a theory is unacceptable insofar as it is counter-intuitive, or has counter-intuitive consequences, and acceptable insofar as it is intuitive. The following quote is from *Naming and Necessity*:

> [...] some philosophers think that something's having intuitive content is very inconclusive evidence in favor of it. I think it is very heavy evidence in favor of anything, myself. I really don't know, in a way, what more conclusive evidence one can have about anything, ultimately speaking. (Kripke, 1981, p. 42)

With a little imagination, we can see the beginnings of a new metaphysical research programme here. Metaphysical theorising is to be done on its own premisses, and theories are to be evaluated in terms of how far they save our "pretheoretical intuitions".[5] In this programme, the notion of *metaphysical necessity* often takes a central place; Lowe (1998), for instance, sees the entire role of metaphysics as explicitly dependent on the existence of metaphysical necessity. Ellis's "scientific essentialism" (Ellis, 2001) depends on metaphysical necessity to separate the essential properties of things from the contingent. And, for the most metaphysically influential application of them all, Putnam's once-held views on natural kinds (Putnam, 1975a) finds necessary *a posteriori* identities to be the glue that holds them together – that water is H_2O is to be something not only true in virtue of the meanings of our words, but a "logical necessity" in the primitive sense that it *couldn't* have been otherwise.

It is not my aim to argue against the notion of metaphysical necessity here, but neither do I intend to base any philosophy on it. The "intuitivistic" methodology is present even among metaphysicians who do not accord prime importance to questions of metaphysical necessity. Armstrong, for instance, advocates use of what he calls the *Euthyphro dilemma*, named after the dialogue of Plato in which Socrates asks whether that which is good is good because the Gods love it, or whether the Gods love it because it is good, as a metaphysically useful method. An example of its use is the following argument against "class nominalism", i.e. the theory that properties are classes, given by Armstrong in *Truth and Truthmakers*:

> It is useful to pose the *Eutyphro* dilemma here. It is in many ways the most useful dilemma in metaphysics, and the argument of this essay will rely on it at a number of points. Consider, first, the class of objects that are just four kilos in mass. Do the members of the class have the property of being just four kilos in mass in virtue of membership of this class? Or is it rather that they are members of this class in virtue of each having the mass-property? The latter view seems much more attractive. The class could have had different members, but the mass-property would be the same, it would seem. (Armstrong, 2004, pp. 40–41)

[5] This is of course not a principle exclusive to metaphysics; it is also very common in epistemology and ethics, and it furthermore rears its head in the philosophy of language now and then. The objections taken up against it below all apply to its use in these areas as well.

It is hard to imagine that Quine, despite being just such a "class nominalist", would take an argument like this seriously.[6] His evaluative standards are not Armstrong's, and arguments relying on counterfactual thinking, "in virtue of"-relations, and imagining the *same* class having *different* members, would simply cut no ice for him.

In order to assess contemporary intuitivistic metaphysics, we have to separate its two phases: (*i*) *rejection of linguistic analysis as a means for attaining metaphysical knowledge*, and (*ii*) *the use of intuitive content as an evaluative principle.* We have already accepted the first of these, at least partly: linguistic analysis is insufficient for metaphysics.

Thus we come to the second phase of the programme: evaluation of metaphysical theories with regard to intuitive content. *This principle must be rejected outright.* Metaphysical theories are theories about what the world is like, or may be like, and not only about what our beliefs about the world are like. They are true or false according to whether they describe reality as it is. The *ultimate* evaluative criterion of a metaphysical theory is therefore its truth—just as for a *physical* theory. Now, truth is of course very hard to determine, and when it comes to metaphysics, almost impossible. We therefore need to use indications of truth instead (again no difference with physics here). But it is precisely here that intuitivism fails, for, contrary to what Kripke claims in the above quote, *something's having intuitive content is no evidence at all for its truth*, at least when it comes to philosophy.

A statement such as this requires some motivation, and we may find an early defendant of it in it in Kant, as he criticises the use of "common sense" for metaphysics, in a lengthy passage in the *Prolegomena*:

> It is a common subterfuge of those false friends of common sense (who occasionally prize it highly, but usually despise it) to say, that there must surely be at all events some propositions which are immediately certain, and of which there is no occasion to give any proof, or even any account at all, because we

[6] Quine himself strenuously objects to being called a "class nominalist", since *nominalism*, for him (as for American philosophers in general, but not for Australians like Armstrong) is the view that there are no abstract objects, and Quine *does* believe in sets. He prefers to call himself a *class realist*, and an *extensionalist* about universals (Quine, 1981a).

otherwise could never stop inquiring into the grounds of our judgments. But if we except the principle of contradiction, which is not sufficient to show the truth of synthetical judgments, they can never adduce, in proof of this privilege, anything else indubitable, which they can immediately ascribe to common sense, except mathematical propositions, such as twice two make four, between two points there is but one straight line, etc. But these judgments are radically different from those of metaphysics. For in mathematics I myself can by thinking construct whatever I represent to myself as possible by a concept: I add to the first two the other two, one by one, and myself make the number four, or I draw in thought from one point to another all manner of lines, equal as well as unequal; yet I can draw one only, which is like itself in all its parts. But I cannot, by all my power of thinking, extract from the concept of a thing the concept of something else, whose existence is necessarily connected with the former, but I must call in experience. And though my understanding furnishes me a priori (yet only in reference to possible experience) with the concept of such a connection (i.e., causation), I cannot exhibit it, like the concepts of mathematics, by visualizing them, a priori, and so show its possibility a priori. This concept, together with the principles of its application, always requires, if it shall hold a priori as is requisite in metaphysics —a justification and deduction of its possibility, because we cannot otherwise know how far it holds good, and whether it can be used in experience only or beyond it also.

Therefore in metaphysics, as a speculative science of pure reason, we can never appeal to common sense, but may do so only when we are forced to surrender it, and to renounce all purely speculative cognition, which must always be knowledge, and consequently when we forego metaphysics itself and its instruction, for the sake of adopting a rational faith which alone may be possible for us, and sufficient to our wants, perhaps even more salutary than knowledge itself. For in this case the attitude of the question is quite altered. Metaphysics must be science, not only as a whole, but in all its parts, otherwise it is nothing; because, as a speculation of pure reason, it finds a hold only on general opinions. (Kant, 1783, pp. 109–110)

These paragraphs could just as well have been written in reply to Kripke, although "common sense" is a strictly narrower concept than intuitiveness; something may be intuitive, but not be common sense, but anything that is common sense must therefore also be intuitive. If common sense is unreliable, intuition must be so as well. Kant's point is simply that intuition is not enough for us to draw any conclusions except the most trivial ones, such as those that follow from the principle of contradiction.

Contemporary critics of intuition-driven philosophy include Hintikka (1999), Sosa (2007), Weinberg et al. (2001); Machery et al. (2004), Cummins (1998) and Ladyman and Ross (2007). Largely from empiricist positions, they object to the rationalist methodology inherent in intuitivism. Indeed, the motivating force behind intuition-driven philosophy *is* Cartesian: "intuitions" are what Descartes's "clear and distinct ideas" have become, in contemporary parlance. But we know much more about the human psyche now than we did in the 17th century. In particular, the theory of natural selection tells us that those traits of our psychology that have been propagated primarily are those that enhance likelihood of survival, or at least of producing fertile offspring.

This means that "common sense" about those things relevant to our survival is likely to be fairly reliable. It is quite easy to show, decision-theoretically, that the greatest chance of survival generally belongs to those who have most of their beliefs about things which affect our survival ability true. Philosophy, however, is totally irrelevant to survival from an evolutionary point of view. Natural selection has no way of weeding out veridical intuitions about the basic constitution of matter, for instance, from false ones, because humans have not generally been killed before they can procreate due to having erroneous metaphysical intuitions. Or bluntly put: having a true metaphysical theory does not help you getting laid.

Contemporary physics bears this out clearly: we have reason to believe that the world is an extremely counter-intuitive place, and our intuitions have been shown to be wrong at least as many times as they have been shown right. Not even our *logical* intuitions can be trusted—ask a logician (or a logically trained person in general) from before 1920 if we from something's having the both the property F and one of the properties G and H always can draw the conclusion that it must have either both F and G, or both F and H, for instance. Before Birkhoff's and von Neumann's work in quantum mechanics, it is unlikely that anyone would have answered no to this, and yet we know that there *are* counterexamples to the "law" of distributivity.[7] But if not even these intuitions are reliable, why would intuitions about things like counterfactual cases, property instantiation, or the dispositions of electro-finks be?

For Kripke, phase (i) and phase (ii) were interdependent. Intuition proves that the proposed linguistic analyses are wrong, and if we cannot rely on linguistic analysis to produce truth, some other means has to be employed, and what could that be besides intuition? It should, however, be clear by now that I hold intuition to be of no use at all here. Even if we grant (i), which I do, we will have to find some other ground for our metaphysical theorising. If this *should* turn out to be impossible, the honest reaction will not be to say "well, then we have to settle for intuitions after all", but rather "so much the worse for metaphysics".

1.4 Naturalistic Metaphysics

If you are a metaphysician, chances are you have not included yourself among the targets of the last section's critique of intuitivism. Many metaphysicians nowadays like to think of themselves as *naturalists*,

[7] The classical philosophical defense of this position is Putnam's *Is Logic Empirical?* (Putnam, 1968). The common way to "reinstate" classical logic would be to say that quantum mechanics does not give us a particle's *properties*, but only the results of *measurements*. Apart from being unpalatable to a realistically-minded metaphysician, this has the further problem that we can regain the failure of distributivity fairly easily. Consider a tunneling experiment, where we fire an electron e at a known speed v towards a thin membrane. We can then take $F(x)$ to be "when measured, x is found to be moving in a line from the electron gun towards the membrane with speed v", $G(x)$ to be "when measured, x is found to be in front of the membrane" and $H(x)$ to be "when measured, x is found to be behind, or inside, the membrane". Then $F(e)$ is true, and $(G(e) \vee H(e))$ must be true as well. But neither $(F(e) \wedge G(e))$ nor $(F(e) \wedge H(e))$ can hold, for both would violate the uncertainty principle.

though it may sometimes be hard to find out exactly what this means. For Armstrong, it is the ontological thesis that space-time and its contents are all that exist (Armstrong, 1997, pp. 5–6), and as such a substantial metaphysical hypothesis. More commonly it is taken to be more of a guiding principle, loosely inspired by Quine's slogan that "philosophy of science is philosophy enough" and the idea that philosophy is to be continuous with science, rather than an attempt to furnish a foundation for it. Philosophy, to be relevant, must on this conception be scientifically informed.

There are at least two types of metaphysical naturalism. The first, which we will refer to as *weak* naturalism, merely dictates that philosophy should not contradict the sciences, but rather be inspired by them and work together with them. According to weak naturalism, we cannot produce a valid philosophical argument that time but not space is unreal, for instance, for time is just as real as space in relativity theory. But there is also a stronger reading, which focuses on scientific method as the sole means for finding things out about the world. *Strong* naturalism, as we will call it, seems to be in direct contradiction with intuitivism.

To evaluate strong naturalism, we need to get a grip on what parts of scientific method are applicable to metaphysics. A principle popular among current metaphysicians is *Inference to the Best Explanation* (IBE): from a set of data, taken as given, we infer the truth of the best theory that explains this data. This principle is seemingly in use in the sciences, so why should not metaphysicians avail themselves of it as well?

We should be careful here. There are several principles in the vicinity of IBE, and not all of them are equally valid. Two processes that *are* in use in the sciences are those I will refer to as *abduction* and *Inference to the Most Probable Explanation* (IMPE). By abduction, we will mean the framing of hypotheses, without deciding whether to believe them or not. It is a crucial part of science. Such hypotheses may have varying degrees of "goodness" due to fit with other theories, likelihood conferred to data, testability, simplicity, and other properties. In some cases, there may be only one known hypothesis worthy of investigation.

IBE goes far beyond this however, and says that we may infer the *truth* of such a hypothesis. But this, I hold, is *not* something that is commonly done in the sciences. That a theory gives the best explanation for a phenomenon is not a reason to believe in it, but to *test* it. It is not until positive results of such a test are in that we should invest our credence in it. A scientist *qua* scientist has no business placing trust in a theory designed to account for phenomena. It is only when the theory has been matched against *new* data that we may infer anything about its truth.

It is here that IMPE plays a role: we can, for example, use Bayesian updating, and then pick the theory with the largest posterior probability. But such a probability may be fairly low, and thus it is not clear that even IMPE is a valid principle. Perhaps we should talk about inference to a *sufficiently* probable explanation instead.

There are also important disanalogies between purported use of IBE in the sciences, and its use in philosophy. First of all, what is it we explain? In the sciences, it is empirical data. In philosophy, however, we often take the given to include intuitions, and their unreliability has already been pointed out. What we should try to explain is not why our intuitions are *true*, but only why we have them, and that may be a job better suited for evolutionary biology, developmental psychology, and sociology, than for philosophy.

Even if we limit ourselves to IBE of purely empirical data, the important fact remains that IBE, for the sciences, primarily appears as abduction. It is a stepping stone, and not an endpoint. The primary tests remain empirical, and a theory with no chance of ever being empirically confirmed or disconfirmed is simply not taken seriously, no matter how well it explains the data. In philosophy, on the other hand, we have no way of testing the results of IBE, independently of IBE itself. This means that using IBE as the sole test for validity of a theory involves a gross overestimate of what the principle is able to do: it can be used to direct our attention to theories that are worth testing, but it cannot, on its own, give any validity to metaphysical theorising.

A strongly naturalistic metaphysics that does not depend on IBE, as well as a general programme to naturalise metaphysics, is defended in Ladyman's and Ross's *Every Thing Must Go* (Ladyman and Ross, 2007). Their guiding principle is what they refer to as the *Principle of*

Naturalistic Closure (PNC):

> Any new metaphysical claim that is to be taken seriously should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses jointly explain more than the sum of what is explained by the two hypotheses taken separately, where a 'scientific hypothesis' is understood as an hypothesis that is taken seriously by institutionally *bona fide* current science. (Ladyman and Ross, 2007, p. 30)

Ladyman and Ross see the task of metaphysics as one primarily of unification of scientific theories. They cite Philip Kitcher's work on scientific explanation (Kitcher, 1981, 1989) as an inspiration, and one may indeed say that so long as we accept Kitcher's view , the goals of metaphysics — to give scientific explanations of theories — are *the same* as the goals of theoretical science. This is why I have classed their methodology as strongly naturalistic.

In order to substantiate the notion of "explaining more" that Ladyman and Ross use, let us introduce the notion of *explanatory power* $e(h)$ of an hypothesis h. For simplicity, assume that explanatory power is ordered by a relation $\succ$, and that there furthermore is an operation of *addition* $(+)$ defined on this structure. A metaphysical hypothesis h_m must then perform a service in showing that $e(h_1 \& h_2) \succ e(h_1) + e(h_2)$, where h_1 and h_2 are scientific hypotheses, for it to pass the PNC.

For this, we cannot of course in general have that $e(h_1 \;\&\; h_2) = e(h_1) + e(h_2)$, so it must genuinely be the case that some hypotheses together explain more than the sum of what they explain individually. One interpretation that satisfies this is to take $e(h)$ to be the set of *phenomena* that can be explained by hypothesis h, take $\succ$ to be the superset relation, and the sum operation to be set union. On this reading, h_m must be necessary as a premiss for us to show that $e(h_1 \& h_2) \supset e(h_1) \cup e(h_2)$.

This is, however, probably not what Ladyman and Ross have in mind. As followers of Kitcher, they hold explanatory power to be *unifying* power. For Kitcher, this is a property of a *generating set* $G(D)$ for a set D of derivations of the hypotheses we are interested in, where a generating set is a set of *argument-patterns* that the elements of D

instantiate. Unifying power is taken to increase with the number of conclusions of the derivations in D (i.e. the number of hypotheses we can derive), increase with increasing stringency among the patterns in $G(D)$, and decrease with the number of patterns in $G(D)$. That $e(h_1 \,\&\, h_2) \succ e(h_1) + e(h_2)$ can then be interpreted as the claim that $h_1 \,\&\, h_2 \,\&\, h_m$ is to have a smaller generating set than $h_1 \,\&\, h_2$ have on their own. A metaphysical hypothesis must unify actual scientific hypotheses in order to be acceptable.

This reading of the PNC does however make it hard to determine what makes the hypothesis h_m *metaphysical*, except that we have chosen to call it so. Any theoretical hypothesis in the sciences should be assumed only in so far as it explains phenomena, and on a unificationist understanding, this means that it should unify them. PNC does not only dictate that the goals of metaphysics are the same as those of science, but also that the methods of metaphysics are the same as those of theoretical science. We are therefore justified in asking in what way PNC is a principle for metaphysics at all.

Ladyman and Ross primarily see the difference between metaphysics and the sciences as one of scope: the individual sciences are specialised in a way in which metaphysics is not, and so metaphysics has the task of unifying hypotheses when these belong to *different* sciences, while we may assume that within their areas, unification may be achieved by the sciences themselves. The divisions between sciences thus determine which claims are metaphysical, and which are not. The problem with this is that the sciences are not really this discrete, except when we identify them with departments at specific universities, and even then we often have crossover subjects like physical chemistry, chemical biology, and neuropsychology as well.

The distinguishing marks of metaphysics thus do not necessarily show themselves as far as we only try to unify hypotheses pairwise. But what if we consider a large number of hypotheses, all belonging to what traditionally is seen as different sciences, or even one from each and every science? Now Ladyman and Ross are very critical of the notion of reduction of one science to another, but if it would be the case that all sciences *were* reducible to physics, then it still would be physics that ought to effect the unification. A unified theory would in that case be a *physical*, and not a metaphysical theory.

If, on the other hand, reductionism is false, and neither physics nor any other of the sciences could ever unify such a set of hypotheses, we should ask ourselves how metaphysics could do better. I will not attempt to answer this question, as I think the only thing that can be said for it is that it has not been shown that metaphysics has succeeded in this yet, nor, for that matter, that it *is* impossible. To some degree, the idea looks promising. Metaphysics, as dealing with conceptual systems, could plausibly attempt the task to unify such systems from different sciences, while leaving their laws and particular statements out.[8] For the PNC to be fulfilled, we do however need something stronger. An example of such metaphysical unification would have to be given in the form in which Kitcher gives theories of genetics and evolution, so that one can see clearly whether actual unification has been done. Certainly, no such unification is given in *Every Thing Must Go*, so we will have to wait and see.

The *prima facie* problems of strong naturalism that we have encountered mainly seem to center on the question of whether there still is any meaningful work for metaphysics to do after the sciences have staked out their areas of interest. Strongly naturalistic metaphysics is on the verge of sliding into the sciences and being swallowed—eliminated—by them. This does not have to be wrong, and we should of course not presuppose that there *is* anything for metaphysics to do. Yet, it is also worth exploring other directions in which metaphysics could be useful, as we wait for examples of metaphysical unification to show up.

Although Ladyman and Ross are to be commended on the strength of their naturalism, the most common form of it is the weak one. Often, it is taken to mean simply that metaphysics should not concern itself with *particular* statements, such as that there is no elephant in the room I am in now, but only with the *general* ones, such as what it is that makes claims of non-existence true. An instance of this is Armstrong's *a posteriori realism* about universals: while his arguments for the general structure of the space of universals are *a priori*, he leaves

[8] A discipline where this is attempted is ontology, in the sense a computer scientist uses the word. In computer science, an ontology *is* a formal representation of a set of concepts for reasoning about things in a given domain.

the identification of which specific universals exist to empirical science (Armstrong, 1989, p. 87). Likewise, metaphysicians often take themselves to be concerned with the general nature of necessity, but hold that which truths really are necessary or contingent is a scientific matter—it is, supposedly, science that tells us that water is necessarily H_2O, but only contingently the main ingredient in our lakes and oceans.

Interpreted this way, naturalism is compatible with fairly large doses of Cartesian rationalism and reliance on intuition, as we have seen from our discussion of Armstrong's "Eutyphro dilemma" above. The problem with this is that intuition is no more a guide here than it is when it comes to particular questions. Indeed, one could even argue that it is *less* of a guide: when we have intuitions, they usually regard specific things. We (or at least I) do not really have intuitions about universal generalities.

This illustrates the dangers with weakly naturalistic metaphysics: since science is *silent* on so many questions of interest to metaphysicians, it is all too easy to slide back into intuitivism. And although this may be a problem more with the metaphysicians than with naturalism itself, it also shows that weak naturalism does not give the metaphysician what she needs. This leaves us searching for some kind of middle road between weak and strong naturalism—a metaphysics whose methodology is *inspired* by that of science, but not necessarily *identical* to it. The vagueness of this notion, however, places it in constant danger of collapsing into weak naturalism, and from there into intuitivism. Our verdict on naturalistic metaphysics must therefore be that it so far just affords the barest sketches of a research programme, and that while its general motivation may be sound, its details remain to be worked out.

1.5 Metaphysics as Model Theory

Not all scientific progress consists in unification. We also have the very important process of *model building*: designing mathematical, mechanical, mental, computational, or even physical models that fulfil the postulates of a theory. Such models are important not only for concretising abstract theory, but also as perspectives from which to suggest new theories, or revisions of old ones. A model of a theory T is an answer to the question "what could the world be like, given that T is true?"

Of the various types of model available, the mathematical ones are particularly useful. Using a mathematical model lets us *prove* things about it. This, in turn, gives us far greater clarity than any other known method. The importance is not that mathematical proof is more *certain* than other forms of argument, but that it gives a much deeper understanding. Therefore proof is essential to scientific thinking, and if we are to approach metaphysics scientifically, we should be able to prove things in metaphysics no less than in physics.

On this view, metaphysics consists in the construction of world-representations. It is thus not quite an empirical science, but it can still be well connected to science. Its closest kin is mathematics, rather than physics. This section will contain some broad outlines of what I take this "model-theoretic" conception of metaphysics, as I will refer to it, to consist in, and how it is related to the others, as well as to science and other parts of philosophy.

Let us start with model theory itself.[9] As a subject, it is usually said to have started in the 1950's, although certain results that were later seen as model-theoretical had appeared before, such the Löwenheim-Skolem theorem from 1920, Gödel's completeness theorem from 1930, and Tarski's definition of truth from 1936. Its inception, as Chang and Keisler explain in their classic book on model theory (Chang and Keisler, 1973, p. 3), was the realisation that a theory could have more than one model, due to the development by Bolyai and Lobachevsky of non-Euclidean geometry, and Riemann's construction of a model of

[9] I am aware that the word "model" often is used in very different ways in science and in logic. I will not try to capture any of these uses perfectly however, but instead introduce a model concept that can do work both logically and scientifically.

geometry in which the parallel postulate was false but all the other axioms were true.

A *model* as we will understand it is something that, given an *interpretation* of a theory, can be used to determine the semantic values of statements in that theory. Such an interpretation is a function h from entities in the theory (for instance its sentences, terms and predicates) to entities in the model. By a *semantics* $\mathcal{S}$ for a theory we will mean an assignment of *semantic values* to pairs of statements and interpretations of these. For our purposes, the most important semantic values are *truth* and *falsity.* These concepts are illustrated in fig. 1.1 below.

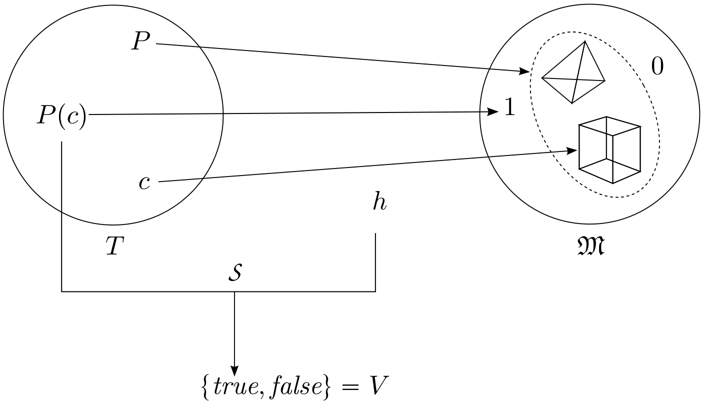


Figure 1.1: ***Theory, model, interpretation and semantics.***

In this example, T is a theory whose language contains the predicate P, the individual constant c, and the sentence $P(c)$. $\mathfrak{M}$ is a model containing a cube and a tetrahedron, together with two objects 0 and 1. The interpretation h interprets c as referring to the cube, P as referring to the set containing the cube and the tetrahedron, and $P(c)$ as referring to the object 1. For example, $P(x)$ might be "x is a shape". The semantics $\mathcal{S}$ assigns a semantic value from the set V to the sentence $P(c)$, given the interpretation h. In the case depicted, we interpret $P(c)$'s taking the value 1 under h as $P(c)$ being *true* under this interpretation.

We will refer to a set of models with a common type of structure as a *model space*. *Model theory*, as we interpret the term, investigates the properties of model spaces. This usage deviates from what is the regular one; Chang and Keisler (Chang and Keisler, 1973), for instance, hold model theory to be the sum of universal algebra and logic. This would make their "model theory" more like what we have called *semantics*, and it corresponds better with what mainstream, first-order model theory has been working on. Mainstream model theory deals with models made using set-theoretical algebraic constructions, for use in interpreting theories formulated in first-order logic. I mean something much wider with the word "model" here; we will talk about models for all kinds of theories, and not only those formulated in first-order logic. We will not even assume that they have to be formulated in a language at all, but accept that they can be beliefs, diagrams, or for that matter matehematical structures themselves, as it is common to see them in the structural (Sneed, 1971; Stegmüller, 1979) and semantic (Giere, 1979; van Fraassen, 1980) conceptions of theories. We will also not take the structures that can serve as models to necessarily be sets, relational systems or algebras; mathematics, even though much of it can be formulated in terms of set theory, is a broader subject than that and studies any kind of abstract structure.

What connects a theory to the actual world, rather than to an arbitrary model, is the notion of *truth*. This is of course a very controversial concept, philosophically. We will try to avoid most of the controversies by adopting what I will refer to as a *thin conception of truth*. First off, let us start with the notion of a *truthbearer*. I will refer to anything to which we may ascribe truth or falsity as a *possible claim*, or more often just a *claim*. Examples may include beliefs, sentence tokens, propositions, speech acts, diagrams, depictions, etc. I do not assert that any of these actually exist, nor that they *are* truthbearers, but only that if you believe in them, and believe that they can be true or false, they are to be included in what I have called claims. Given this notion, we may gloss the thin conception of truth as follows:

TT: A claim p is true iff p says that the world satisfies φ, and the world does satisfy φ, for some condition φ on worlds.

This is of course not proposed as a definition; for one thing, it employs the notions of *satisfaction* and *condition*, which are unlikely to be less complex than truth itself, and it also talks about such things as "worlds" and "saying that". We may instead view it as a sort of *condition* on truth-definitions: *given* explications of truth, worlds, "saying that", conditions, and satisfaction, this is something that must hold between them. As such, it gives a partial definition of truth in the sense that it rules out *some* theories. The thin theory of truth itself can then be taken as the meta–theory that says that one of the theories not thus ruled out is the correct one.

As weak as TT is, it still contradicts some positions. For instance, it is incompatible with coherentism about truth (that what is true is what is entailed by our most coherent body of beliefs) coupled to the belief that the world is independent of our beliefs of it. But it is not incompatible with coherentism *per se* — if we hold *both* that truth is that which follows from our most coherent body of beliefs, and that that worlds *are* bodies of beliefs, for instance, that can satisfy TT (see section 3.4.3).

The motivation behind the thin theory of truth is the same as that behind correspondence theories: that what the world is like is what determines what is true or false. This much is arguably a part of the very meaning of the word "true", so that denying it would be something like denying that bachelors are unmarried. The appropriate entry in the *Compact Oxford English Dictionary* for "true", for example, is "in accordance with fact or reality". The interpretation is in agreement with deflationists such as Horwich:

> It is indeed undeniable that whenever a proposition or an utterance is true, it is true *because* something in the world is a certain way—something typically external to the proposition or utterance. (Horwich, 1998, p. 104)

and even anti-realists, such as Dummett:

> If a statement is true, there must be something in virtue of which it is true. This principle underlies the philosophical attempts to explain truth as a correspondence between a statement and some component of reality, and I shall accordingly refer to it as principle C. The principle C is certainly in part constitutive of our notion of truth [...] (Dummett, 1976, p. 52).

I will therefore take TT to be *trivially* true, and denials of it to be due to conceptual confusion rather than substantial disagreement. Someone who proposes a theory of truth in which truth is independent of what the world is like, is better interpreted as proposing a replacement for our regular concept — perhaps because that concept is held to be useless in practice, or incoherent. This debate will not concern us at the moment. Truth, for us, is a *starting point*, from which to set out on our metaphysical odyssey. We simply assume it to be obtainable, in so far as it is obtainable at all, by scientific method. As metaphysicians, we *use* the truths given to us; it is not our primary task to find out new ones, or to question those we are given by the sciences.[10] Dummett continues the above quote by claiming just such a task for his principle C:

> [...] but it is not one that can be directly applied. It is, rather, regulative in character: that is to say, it is not so much that we first determine what there is in the world, and then decide, on the basis of that, what is required to make each given statement true, as that, having first settled on the appropriate notion of truth for various types of statement, we conclude from that to the constitution of reality. (Dummett, 1976, p. 52).

It is for this task that TT (our version of Dummett's principle C) is enough. We only need *some* link between our true theories, and what the world is like, for us to be able to reel in reality (or at least parts of it) by pulling on it. One question of importance, however, is why it has to be *truth*. It is just one of the circle of semantic concepts which includes truth, reference and satisfaction. Tarski, for instance, took satisfaction as fundamental, disregarded reference entirely (because none

[10] By this i do not, of course, mean that a philosopher could never challenge claims of the sciences, but only that when we do so, we do it as theorists of science rather than as metaphysicians.

of the languages he wrote about had individual constants), and defined truth from there. Modern logicians are likely to take both reference and satisfaction to be the basic concepts, and truth to be derivative.

It is easy to see the advantages of this approach: knowing what things satisfy $P(x)$ and what the constant c refers to, it is a trivial matter to find out whether "$P(c)$" is true. On the other hand, from just knowing that "$P(c)$" is true, we can tell next to nothing about what things satisfy $P(x)$, or what c refers to. We need far more information than that, such as in what circumstances (or "worlds") $P(c)$ *would have* been true, what *other* individual constants can be replaced for c *salva veritate*, or even knowledge of the inference relations in the entire language.[11] But even given these, we can never be *certain* that we will be able to regain determinate referents for our singular terms; it was just problems like these that drove Quine to his position in *Ontological Relativity* (Quine, 1969), and his later appreciation of ontology as meaningless. So it would seem that we basically *have to* start with taking reference and satisfaction for granted, if we are to do any metaphysics at all.

Doing so would, however, be to succumb to wishful thinking. The sentence is the basic unit of meaning—it is what is asserted in a speech act—and the meanings of words are derivative. Frege put the point best by claiming that "Only in the context of a sentence do the words mean anything" (Frege, 1884, §62).[12] We do not have any way of referring to objects, or of predicating anything of them, that does not presuppose the referring words' roles in making assertions. Since meaning must be determined by use if we are to see it as a social phenomenon at all, the meaning of such referrals or predications must be determined, as far as it is determined at all, by their use in assertions. [13]

More explicitly spelled out, the argument is the following: sentences'

[11] Brandom's theory of language is an example of one that takes inference relations to be fundamental, and derives reference for terms, and satisfaction of predicates, from this network (Brandom, 1994).

[12] Frege uses the word "bedeuten", but since this was before his distinction between *Sinn* and *Bedeutung*, I have interpreted him as using it in its customary sense, and translated it as "mean" instead of the philosophically more common "stand for".

[13] The classic work here is Dummett's, on Frege's philosophy of language, where these points are explained far better than I ever could (Dummett, 1981).

meanings are determined by their uses, i.e. what the effects of uttering them are. Words (such as names and predicates) receive their meaning derivatively, from the stable contributions they make to the meanings of sentences in which they appear. The way in which we assign these word-meanings does however depend on our underlying intuitive preconceptions of metaphysics: we must assume individual objects to exist and have certain properties for them to be the meanings (or referents) of individual constants, for instance. In order to make the whole process of inferring a metaphysics from a theory explicit, it is therefore safer to focus on the semantical properties of whole claims, such as truth.

Similar considerations apply to other kinds of claims, such as beliefs. The view of ideas as depictions of objects, which refer to that which they are depictions of by being similar to them, which was attributed by Berkeley to Locke, has long since been given up, if it was ever held by anyone at all. Beliefs, just as statements, must be determined by their relations to manifest behaviour. Just as we do not have any *fundamental* linguistic act of referring, we do not have such a fundamental mental act either.

It may seem that questions such as these would be of interest primarily for the philosophy of language and the philosophy of mind. Why cannot we, as metaphysicians, simply leave the question of priority to these subjects, assume that it can be sorted out there, and that *some* notions of reference and predication will be available for us to use? The reason is that reference and predication are not metaphysically neutral: in employing them, we have already taken reality to consist of individual, self-subsistent objects, and things that can be said *about* these. On some semantics, such as Frege's (which may be seen as the *intended* interpretation of predicate logic—after all, he invented it with that interpretation in mind, and certainly not as a purely formal calculus), predicates stand for entities as well. Granting ourselves reference and satisfaction there would therefore commit ourselves to a full Platonic heaven. As metaphysicians, we want assumptions like these to be the *result* of our theorising, and not silent presuppositions.

This is why we focus on truth as the central semantic concept for use in metaphysics. Reference and satisfaction, if we find that we need them, will have to come in at a later stage. This means that the semantics we work with primarily will be the semantics of sentential languages and similar structures. We *will* discuss languages of predicate logic as well, but these will be treated as a special case of our general theory.

We argued, when discussing Quine's methodology, that metaphysics should not concern itself with formulating its own theories about what the world is like, but rather interpret theories formulated by the sciences. Such an interpretation, when expressed mathematically, *is* a type of model. This is why I say that metaphysics is model theory: its subject-matter is the construction of models for theories we have reason to believe to be true.

What differentiates model-theoretic metaphysics from the more traditional kind, except for its greater reliance on mathematics, is the extra level of abstraction involved in treating model spaces, rather than single models. The metaphysician's task is limited to the design of a *type* of model, and she has no say in what model in a given model space is the one corresponding to the actual world. That is entirely up to science, through the mediation of semantics, and if science does not suffice for making a unique choice, then nothing else will.

Chapter 2
Theories

Here we give an explication of the concept *theory* which is broad enough to cover empirical theories as well as logics and natural languages. A theory A is defined as a consequence operator C_A on a language L_A, and we can identify the truths of the theory with its theorems. We introduce several variations, among which are algebraic, many-valued and probabilistic theories.

We also discuss the use of theories as frameworks for working with other theories. Certain sets of claims (the *closed* sets) in a theory form theories of their own, called *strengthenings*. The strengthenings of a theory form a kind of logic, with lattice-theoretical properties. More distant connections between theories can be captured using theory transformations. A *theory homomorphism* is a consequence-preserving function between theories' languages, and a *translation* is a kind of homomorphism that not only preserves consequence, but reflects it as well.

Finally, we discuss the matter of *necessity*. This concept will be of importance for us later on, in a metaphysical setting. Here, however, we treat it as a modality of claims, and investigate its relations to the consequence structure of a theory. This type of necessity is thus inherently theory relative.

2.1 Logic and Theory

Traditionally, the notions of *logic* and *theory* have been taken to be exclusive of one another. Aristotle's subject matter in the *Physics* seems quite different from that in the *Organon*, to take an early example. Yet, we are unwilling to deny that statements true in virtue of their logical properties are *true*, just as those which are true in virtue of how the world is. And just as both logic and empirical investigations aim at truth, both material and formal implications can be used for drawing inferences.

The first modern philosopher to take these similarities seriously was Carnap. In *The Logical Syntax of Language*, he lets his languages contain two types of formal rules: the *logical* rules, or *L-rules*, which are stable under replacement of non-logical symbols, and the *physical* rules, or *P-rules*, which are not (Carnap, 1937, §51). The difference can be interpreted as one concerning the basis of the inferences allowed: L-rules can give only what follows from a sentence's *logical form*, which is a hypothesised property shared by those sentences that may be obtained from one another by replacement of non-logical symbols.

This difference naturally depends on our being able to give a classification of which symbols are *logical*, and which ones are *descriptive*. Carnap never gave one, and the debate is still lively — one of the currently most popular accounts seems to be Tarski's, of permutation invariance of the domain (Tarski, 1986). Symbols commonly accepted as logical include those for conjunction, disjunction, negation, quantification, necessity and identity. Symbols that are not commonly included are set membership, part-whole relations, and predicates and individual constants that denote physical properties or objects. Nevertheless, we are far from any kind of consensus on what a logical symbol is, or even which symbols are logical.[1] For Peano, set membership was a paradigmatically logical relation, while it is not for us. To a large extent, the difference seems to be purely conventional.

This uncertainty over the line between logical and non-logical constants translates into a vagueness in the notion of logical form, and

[1] Cf. MacFarlane's careful discussion in the *Stanford Encyclopedia of Philosophy* (MacFarlane, 2005).

thus also into blurriness in the boundary between logical and physical (or *material*) consequence. Quine is the philosopher best known for exploring this blurriness — most famously as regards analytical consequence in *Two Dogmas* (Quine, 1951), but also more directly in *Carnap and Logical Truth* (Quine, 1960a).

One does not have to go all the way to Quinean pragmatism, and deny the very distinction between logical and material consequence altogether, however. Wilfrid Sellars, another prominent philosopher as deeply influenced by Carnap as Quine was, accepts the distinction, but holds that material rules of inference are necessary for us to be able to capture the notion of validity. In *Inference and Meaning* (Sellars, 1953), he argues that the existence of subjunctive conditionals requires our language to contain material rules of inference, since the coarse-grainedness of strict implication (which is the object-language equivalent of logical consequence) makes it incapable of distinguishing between different counterfactual conditionals.[2] In *Some Reflections on Language Games* (Sellars, 1954), he shows how material rules of inference are needed to ground even the logical ones, since we cannot learn to play a language game (Sellars's Wittgensteinian word for what seems to be a "language"' in the Carnapian sense), unless some material rules of consequence are in place.

No matter where you stand in these questions, it should be obvious that there is a reading of "consequence" that is wider than formal or logical consequence. In keeping with the (very) broadly Quinean methodology I have been inspired by in this book, we will not presuppose any kind of absolute difference between this wider ("material") notion of consequence, and the narrower "logical" kind. This does not, of course, mean that no such distinction could be introduced, but only

[2] In a way, later theorising may be seen as having borne these speculations out: Lewis, in his seminal work on counterfactuals, refers to them as *variably strict conditionals* (Lewis, 1973, 13–19). His analysis furthermore contains crucial elements, such as his ternary similarity relation $\leqslant_i$, which we have no reason to believe to be expressible as a set of "extra premisses" of first-order logic. The reason why he is able to see himself as dealing with the *logic* of counterfactuals is of course that he only discusses very general structural conditions on $\leqslant_i$. Nevertheless, in order to ascertain whether an actual inference *is* valid or not, we need to have access to the full similarity relation, and not only a few tidbits of information about it such as whether it is reflexive or not.

that it is nothing we will take for granted.

For us, another important advantage of not tying ourselves to a specific logical consequence relation is that this allows us to avoid the points of criticism we raised against Quine's programme in section 1.1. Much of this centered around Quine's dogmatic reliance on first-order logic with identity (henceforward "FOL"), and the problems of arbitrariness and limit in scope that this brings. *Prima facie*, one way of tackling these problems would be to select some other logic as our foundation, which does not have the limitations of first-order logic. But which one? If there is one thing we should have learnt from the classicism–intuitionism debate in the philosophy of mathematics, it is that showing that some logic is the "right" one is incredibly difficult. But it is also the case that some of the questions of sect. 1.1 pull in different directions: some are about why FOL is too *weak*, and some are about why it is too *strong*.

The only way out of this, if we are to approach metaphysical methodology in general, without bias, is to assume no specific logic at all. A few properties *will* follow from our theory of theories, such as that entailment is transitive. If this is unpalatable, it is possible to modify the framework presented here by basing it on non-monotonic consequence operators instead of monotonic ones, for instance. We will indicate how to generalise the theory concept in section 2.4.

In 1.1.5, we introduced the notion of a *claim*: anything that might be true or false. Claims are typically connected in different kinds of systems, and it is these that we will refer to as *theories*. The "glue" that holds the claims together in such a system is *consequence*, which we will represent using the Tarskian notion of a *consequence operator*: a function C on the subsets of a set L, such that the following hold, for any $X, Y \subseteq A$:

$$\begin{array}{rl} (\textit{Reflexivity}) & X \subseteq C(X) \\ (\textit{Idempotence}) & C(X) = C(C(X)) \\ (\textit{Monotonicity}) & \text{if } X \subseteq Y \text{ then } C(X) \subseteq C(Y) \end{array}$$

A *theory* A is a consequence operator C_A on a set L_A of claims, called the theory's *language*, together with a set S_A called A's *subject matter.* Expressed set-theoretically, it is an ordered triple $\langle L_A, C_A, S_A \rangle$. The role that we have focused on here, for the theory, is as a *vehicle of inference.* It justifies the inferences we make between claims in the manner that inferring claim q from claims $p_1, \ldots, p_n$ is justified by the theory A iff $q, p_1, \ldots, p_n \in L_A$, and $q \in C_A(p_1, \ldots, p_n)$, which we also will write as

$$\{p_1, \ldots, p_n\} \vdash_A q$$

In the limit, where the theory allows us to infer a claim from no premisses at all, and thus $q \in C_A(\varnothing)$, we say that q is an A-*truth.* We denote the set of all A-truths by $\top_A$.

We have called L_A the theory A's language, even though not all claims need to be linguistic entities. Taking A's claims as *thoughts*, we may for instance speak about a "language of thought", though not necessarily in the substantive sense that Fodor and others use the term. All we require of a language is that it is a set of claims, of any kind whatsoever. We may even have heterogenous languages, in which some claims are thoughts, others are sentences, and yet others are depictions of states of affairs. Such languages can be useful for studying logical relations holding between claims of different domains.

The third part of a language is the subject matter. This plays the same role as the set of "intended applications" in the Sneed-Stegmüller tradition of structuralist theory of science (Sneed, 1971; Stegmüller, 1979). This is necessary since many actual theories contain indexical elements. For instance, theories in physics often mention "the system", and which system is intended may differ from application to application. In many cases, the applications do not even exist: physics has to be applicable to thought-experiments as well as actual systems, or much of the reasoning done by physicists would be invalid.

This means that we should not interpret S_A extensionally. Some kind of set of descriptions of what things the theory is about, or may be applied to, is sufficient. This allows us to have theories about things that do not exist, or things that we do not know whether they exist or not. It does not rule out theories whose subject matter just is "the world", of course, even if such theories probably are more uncommon in practice than what one may get the impression of when reading contemporary philosophy of science. Nevertheless, whenever we leave out specification of the subject matter in the description of a theory, we will assume it to be applicable to the world, no matter how it is. Thus a theory defined as $\langle L_A, C_A \rangle$ will be assumed to have an implicit subject matter, and be an abbreviation for a theory $\langle L_A, C_A$, "the world"$\rangle$

Consequence, as a property of a theory A, is a purely theory-relative concept. $X \vdash_A p$ is to be interpreted as "A allows inferring p from X", and does not in itself involve anything external to said theory. That something is an A-truth thus does not mean that it is *true*, but only that it is true according to A.

Often all kinds of models or semantics are attached to theories to motivate the inferences allowed. The guiding principle here is that if p is inferable from X, then whenever the claims in X are true, p is true as well. A semantics is then used to flesh out this "whenever" in terms of models, situations, possible worlds, interpretations, etc. But such a semantics remains secondary to the theory and its consequence operator itself. We generally decide on truth or falsity of claims through different kinds of testing, such as experiment, observation, proof, or counterexample. A consequence operator can be motivated through the "it has always worked so far" methodology, and all motivation has to include this as a part.

We do not want to downplay the importance of semantics, on the other hand. This book is to a large extent about the relation between claims and the things they are about. But to be able to approach such questions in an unbiased way, it is very useful to "bracket" the semantical presuppositions of a theory. This is possible because the theory as consequence operator is self-sufficient: two users of it can communicate, so long as they treat consequence the same way, even if one of them motivates the relation through one kind of semantics, and the other through another.

Bracketing allows us to avoid questions about "intended" interpretations. It also allows us to consider theory first, and the question of what the world is like given a theory second. It is thus very useful if we are to do metaphysics as secondary to scientific theory, rather than as first philosophy.

A few examples of theories, some of which we will return to in the last chapter, are the following:

Sentential logics. We will mainly discuss two varieties of sentential logics: the classical and the intuitionistic kind. Both, moreover, constitute classes of theories, rather than single ones. Assume that S is a set of sentences. Let the *sentential language* $\hat{S}$ be the smallest subset containing S which is closed under the following conditions:

(*i*) $\top \in \hat{S}$ and $\bot \in \hat{S}$.

(*ii*) If $p \in \hat{S}$ then $\neg p \in \hat{S}$.

(*iii*) If $p \in \hat{S}$ and $q \in \hat{S}$ then $(p \wedge q) \in \hat{S}$, $(p \vee q) \in \hat{S}$, and $(p \rightarrow q) \in \hat{S}$.

Given any set of sentences S, we define the *intuitionistic logic over* S as the theory $I(S)$ with language $L_{I(S)} = \hat{S}$ and consequence operator

$$C_{I(S)}(X) = \left\{p \in L_{I(S)} \;\middle|\; p \text{ is an intuitionistic consequence of } X\right\}$$

The *classical logic on* S is defined as the theory $C(S)$, with the same language as $I(S)$ and the same definition of consequence operator, except for the replacement of "classical" for "intuitionistic". The sets $\top_{I(S)}$ and $\top_{C(S)}$ are the sets of classical and intuitionistic tautologies, respectively, in the language $\hat{S}$.

Classical predicate logics. Again, this comprises a class of theories. First of all, for every ordinal number n, we have a different class of logics: the nth order ones. Then, for every order, we have different logics depending on what predicates, variables, and function letters we have. Assuming that $\langle S, C\rangle$ is an nth order logic, whose set of sentences is S, we can define C as

$C(X) = \{\varphi \in L \mid \varphi$ is true in all models of nth order logic where all sentences in X are true$\}$.

This definition defers the problem to defining the notion of *truth in a model of nth order logic*. The most common of these are those of Tarski, for first-order logic (although they can fairly easily be extended to higher-order logics), and the proof-based ones, where a model can be taken to be a set of sentences, and truth in a model equated with derivability from that set. For first-order logics these coincide, but depending on whether we allow proofs to be infinite, and on how Tarskian models are defined for higher-order logics, they may come apart for higher orders.

***ZFC* set theory.** Unlike many logics, ZFC set theory is a specific theory rather than a class of them. Its language is a variant of predicate logic: one with no function symbols and the single binary predicate "$\in$". Let Ax be an axiomatization of ZFC set theory in this language. We then define the consequence operator as

$$C_{ZFC}(X) = \{\varphi \in L_{ZFC} \mid \varphi \text{ is true in all models of FOL where all sentences in } Ax \cup X \text{ are true}\}.$$

Here, we have the same choice for our interpretation of "truth in a model" as we had in our last example. For second-order ZFC, in contrast to the first-order theory, differing choices give rise to different theories. Moreover, for proof-theoretic consequence, we have a class of different systems — all finite subsystems of the model-theoretic version, which contain among others the basic second-order logic of Frege's *Begriffsschrift*.

The standard model-theoretic version of second-order logic is the one that holds the greatest interest for most philosophers: it permits us to give categorical axiomatisations of Peano arithmetic, and almost-categorical (that is, categorical up to cardinality) axiomatisations of ZFC. For this theory (call it ZFC^2), we do not have any finite set of axioms. This does not, in any way, prevent it from being a *theory* in our sense: as soon as we have a clearly determined set of claims (the sentences of ZFC^2), and a fact of the matter of which inferences are *valid* or *invalid* (which is given by the model-theoretic consequence notion for ZFC^2), we have a theory.

Classical mechanics. This is a typical example of an empirical theory, and also a good example since it is so well-known. In its Hamiltonian formulation, classical mechanics is used to derive properties about a *physical system*. The state of such a system can in general be described through a set of generalised coordinates. In the simplest case of n free particles, the system is determined by $6n$ coordinates as a function of time – 3 for each particle i's position $\mathbf{q}_i(t)$, and 3 for each particle's momentum $\mathbf{p}_i(t)$.

A theory describing such a system can be defined over a language L_{CM} generated by of formulae of the form

$$p = \text{the value of observable } \mathbf{A} \text{ is } x \text{ at time } t$$

where $\mathbf{A}$ is a specified real function of the coordinates of the system, x is a real number, and t is a time. The system itself can be described as a point in $6n$-dimensional space, and its evolution in time as a trajectory in this space. The observable $\mathbf{A}$ takes such a point as argument, and gives a real value.

The consequence operator C_{CM} can be defined as one of a mathematical framework (e.g. ZFC, together with an appropriate collection of definitions) combined with Hamilton's equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_i(t) = -\frac{\partial}{\partial \mathbf{q}_i}\mathcal{H}(\mathbf{p}(t), \mathbf{q}(t), t)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{q}_i(t) = \frac{\partial}{\partial \mathbf{p}_i}\mathcal{H}(\mathbf{p}(t), \mathbf{q}(t), t)$$

where $\mathcal{H}$ is a function called the *Hamiltonian*, which gives the total energy of the system in each of its configurations. This function is characteristic of the system, and thus of a theory of classical mechanics for a specific system.

To obtain classical mechanics in full generality, we need to get rid of the hard-wiring of $\mathcal{H}$ to the theory. This means that we also will have to include sentences for specifying the Hamiltonian in L_{CM}. Since we will not dwell much on classical mechanics in this book, we will not go into detail of how to do so here.

2.2 Truths and Theories as Claims

We have this far focused on theories' roles as vehicles of inference. It should however be obvious that this is not *all* they are. If A justifies the inference from p to q, then, according to A, if p is true, q is true as well. This is, at the very least, a necessary condition for an inference being justified. In the case where $p \in \top_A$, A justifies drawing the conclusion p from no premisses at all (or equivalently, given monotonicity, from *any* premisses). According to A, all claims in $\top_A$ must therefore be true.

This allows us to extend the notion of truth from claims to theories. Let $true_A$ be the set of all claims in A that *actually* are true under some interpretation we have settled on (not to be confused with the set $\top_A$, which contains the claims in A that are true *according to* A). We define:

> A theory A is *true* iff for any set $X \subseteq true_A$, $C_A(X) \subseteq true_A$.

The case where $X = \varnothing$ is not meant to be excluded here, since that is what makes A's truth entail the truth of all A-truths. Intuitively, the definition says that a theory is true when all the inferences it allows are truth-preserving. This definition is dependent on the notion of truth for *claims*. Since we already have said as much as we will about what this is in section 1.1.5, we will take it for granted here.

Truth, as we have defined it for a theory, is similar but not identical to *soundness*. A logic is *sound* iff it is *impossible* for any set of premisses in the logic's language to be true, without those things the logic says follows from these premisses also being true. Soundness is thus a modal concept. Truth, as we have interpreted it here, is its non-modal cousin: a theory is true iff, for any set of claims in the theory's subject matter, if these are *in fact* true, then everything that is a consequence of them, according to the theory, is also in fact true. We can therefore say that a theory is *sound* iff it is necessarily true.

The possibility for a theory to be true makes it a kind of claim (remember, we have taken claims to be *any* entities to which it is meaningful to ascribe truth or falsity). Theories can thus be elements in the language other theories. Can they be elements in the language of themselves? If our underlying set theory is Cantorian (which we will assume it to be), the answer is no; the axiom of foundation prohibits infinite descending chains or cycles in the membership relation. Since the language of a theory is an element of that theory, a cycle would ensue if the theory itself were an element of its language. Thus, a theory cannot talk about itself, or, for that matter, about theories that include itself. This also helps our theory of theories to avoid liar-type paradoxes.

In a similar way, we can also see that there must be claims that are not theories. Starting with an arbitrary theory A, and following the elementhood relation downwards, we must always come in a finite number of steps to some theory B whose language does not contain any sets at all, and thus not any theories. We have two ways this may happen: either B is the *empty theory* whose language is empty (the "theory of nothing" in the strictest sense), or its language consists of claims that are not theories. But it is easy to see that the empty theory is true, just from the definition of truth of theories. If all elementhood chains of all theories ended in it, all theories would therefore be true. Since this is not the case, there must be claims that are not theories as well.

Due to the well-foundedness of theories, we can always *consolidate* them by including the language of the theories in them in their own subject matter. Call a theory A *consolidated* iff, for any theory $B \in L_A$

(*i*) $L_B \subseteq L_A$.

(*ii*) For any $X \subseteq L_B$, $C_A(\{B\} \cup X) \cap L_B = C_B(X)$.

Condition (*i*) is simply that the language of B is to be included in that of A. (*ii*) requires A's consequence operator to coincide with B's over B's language, so long as B is held to be true as well. This may be held to follow directly from our definition of what the truth of a theory is. We can also see that $C_A(B) \cap L_B = \top_B$, so B implies that the B-truths are true, according to A. We do not necessarily have the reverse implication: the B-truths may all be true without B being true, since B, as a claim, says more than $\top_B$.

We say that a theory A *contains* a theory B, or that B is a *subtheory* of A, iff $L_B \subseteq L_A$ and A's and B's consequence operators coincide on

L_B. Given any non-consolidated theory A, there are generally many consolidated theories that contain it. Nevertheless, the existence of consolidated containing versions of any non-consolidated theory lets us confine our attention to these hereafter.

Apart from the containment relation, there is one other important relationship that holds between theories. We say that B is a theory *in* A, or that B is a *strengthening* of A, iff

(*i*) $L_B = L_A$.

(*ii*) $C_B(X) = C_A(\top_B \cup X)$.

The meaning of (*ii*) is that B is obtainable from A by *fixing* a set of A's claims, and regarding them as true. This set then becomes the set of B-truths. Strengthenings are important because they do not really add to the expressive power of A: everything we can claim in B, we could just as well have claimed in A, by citing the elements of $\top_B$ as further premisses (this is proved in a more formal manner in lemma 2.2 below). Thus, claiming X in B is *the same thing* as claiming $X \cup \top_B$ in A.

What is the importance of strengthening? Why require that

$$C_B(X) = C_A(\top_B \cup X)$$

rather than the more general $C_A(X) \subseteq C_B(X)$, for instance? We must here consider the role of theories not only as subjective entities, but as tools for communication as well. Suppose that you and I are conversing using a theory A, and I want you to accept the move from p to p', which is not allowed in A. There is no way for me to communicate this intention except to say that something or other holds, and this done by making one or more claims in A.

In theories which are compact and have well-behaved implication and conjunction connectives, the difference disappears. Using the deduction theorem, we can express

$$\{p_1, \ldots, p_n\} \vdash q$$

as

$$\vdash (p_1 \wedge \ldots \wedge p_n) \rightarrow q$$

which allows us to take any strengthening of a consequence operator to be a strengthening in our sense. But not all strengthenings, as Lewis Carroll famously pointed out, can be handled like this. We cannot introduce the properties of "→" this way, for instance.

The most common examples of our kind of strengthening in the literature occur when A is a logic. Then, any axiomatic extension of A is a theory in A, and any set X of sentences in A determines a theory in A, which we will call the theory *generated* by X. Here, our usage of the word "theory" touches that of the logician. In logicians' parlance, "theory" means "logically closed set of sentences", and, as the theorem below shows, when A is a logic, such theories correspond one-to-one with the theories in A.

Theorem 2.1 : If B is a theory in A, then $\top_B$ is closed in A, and for any closed set X in A, there is a unique theory B in A such that $X = \top_B$.

Proof. Let B be an arbitrary theory in A, so that $X \subseteq L_B$, $C_A(\top_B \cup X) = C_B(X)$, for all $X \subseteq A$. Then, in particular, $C_A(\top_B) = C_B(\varnothing) = \top_B$, so $\top_B$ is closed in A. Now assume that X is an arbitrary closed set in A. We can then define a theory $B = \langle L_A, C_B \rangle$, where $C_B(Z) = (X \cup Z)$, for every $Z \subseteq L_A$. To show that different closed sets X and Y are the truths of *different* theories, all we have to do is to note that no two distinct theories in A can have the same set of truths. □

Theorem 2.1 proves that the set of theories in A has the same structure as the set of subsets of L_A that are closed under C_A. We call such a set of closed sets the *closure system* $\mathcal{CS}(A)$. Such a system, as can be found in any book on lattice theory (see for example Davey and Priestley, 2002, p. 46), is a *complete lattice*: a structure $\mathfrak{T} = \langle S, \preccurlyeq, \bigwedge, \bigvee \rangle$, where S is a set, $\preccurlyeq$ is an order on S, and $\bigwedge : \wp(S) \mapsto S$ and $\bigvee : \wp(S) \mapsto S$ are functions that give the *greatest lower bound*, or *meet*, and a *least upper bound*, or *join*, of arbitrary subsets of S, in the order $\preccurlyeq$.

When X is a set of claims in the subject matter of the theory A, we refer to the weakest theory in A that includes X among its truths as $Th_A(X)$ (the theory generated by the set X). We have shown that every theory A gives rise to a complete lattice $\mathfrak{T}_A = \langle \mathcal{T}_A, \preccurlyeq, \bigwedge, \bigvee \rangle$—the *theory space* $\mathfrak{T}_A$—where

$$\begin{array}{rl}
\mathcal{T}_A & \text{is the set of all theories in } A, \\
X \preccurlyeq Y & \text{iff } \top_Y \subseteq \top_X \text{, in which case we say that } X\ A\text{-} \\
 & \textit{entails } Y, \\
\bigwedge \mathcal{X} & = Th_A(\bigcup_{X \in \mathcal{X}} \top_X), \text{ and} \\
\bigvee \mathcal{X} & = Th_A(\bigcap_{X \in \mathcal{X}} \top_X).
\end{array}$$

For *pairs* of theories $\{X, Y\} \subseteq \mathcal{T}_A$, we use the notation $X \wedge Y$ and $X \vee Y$ for meets and joins. Strictly, most of these symbols should be subscripted with what theory space they belong to, but we will rely on the context to determine this.

We call the set of theories $\mathcal{T}_A$ the *theory space* of A. The meanings of $\bigwedge$ and $\bigvee$ are clarified by the following theorems, and their accompanying lemma:

Lemma 2.2 : If A is true then, for any theory B in A, B is true iff all claims in $\top_B$ are true.

Proof. First, assume that B is true. Then $\top_B \subseteq true_A$, since the truths of a true theory are all true, from the definition of truth for theories above. Conversely, assume that $\top_B \subseteq true_A$, and that X is an arbitrary subset of $true_A$. For B to be false, there must be some $p \notin true_A$, such that $X \vdash_B p$. But this would require that $X \cup \top_B \vdash_A p$, and we have already assumed X and $\top_B$ to be all true, and C_A to be truth-preserving, so such a situation cannot arise. Thus, B is true as well. □

Theorem 2.3 : If A is true, then $\bigwedge \mathcal{X}$ is true iff all theories in $\mathcal{X}$ are true.

Proof. By definition,

$$\bigwedge \mathcal{X} = Th_A(\bigcup_{X \in \mathcal{X}} \top_X)$$

is true. From lemma 2.2, $Th_A(\bigcup_{X \in \mathcal{X}} \top_X)$ is true iff $\bigcup_{X \in \mathcal{X}} \top_X \subseteq true_A$. But it follows straightforwardly, by the use of elementary set theory, that this holds iff $(\forall X \in \mathcal{X})(X \subseteq true_A)$. □

Theorem 2.4 : If A is true, then $\bigvee \mathcal{X}$ is true if some theory in $\mathcal{X}$ is true, and $\bigvee \mathcal{X}$ is the strongest theory that follows from some theory in $\mathcal{X}$ being true.

Proof. Again, we use lemma 2.2 to work with the theories' truth-sets instead of their consequence operators. Assume that there is a theory $B \in \mathcal{X}$ such that $\top_B \subseteq true_A$. Then, since

$$\top_{\bigvee \mathcal{X}} = \bigcap_{X \in \mathcal{X}} \top_X$$

it follows trivially that if all claims in $\top_B$ are true, all claims in $\top_{\bigvee \mathcal{X}}$ must be true as well.

For the second part of the theorem, let Y be some theory such that $X \preccurlyeq Y$ for all $X \in \mathcal{X}$. We then obviously have that $\top_Y \subseteq \top_X$ for all $X \in \mathcal{X}$, and thus that $\top_Y \subseteq \bigcap_{X \in \mathcal{X}} \top_X$. But this is equivalent to $\bigvee \mathcal{X} \preccurlyeq Y$. □

Unfortunately, we cannot strengthen the implication from some theory in $\mathcal{X}$ being true to $\bigvee \mathcal{X}$ being true in theorem 2.4 to an equivalence: it may be that $\bigvee \mathcal{X}$ is true, although none of the theories in $\mathcal{X}$ are true themselves. This happens, for instance, in quantum-mechanical cases under certain interpretations: here, we can have it true that the spin of a certain particle in a given direction is either up or down (this follows from the interpretation), without it being the case that it is up, or that it is down. For the theories

$$U = Th_{QM}(\textit{the particle's spin in direction } d \textit{ is up})$$
$$D = Th_{QM}(\textit{the particle's spin in direction } d \textit{ is down})$$

we then always have that $U \vee D$ is true, even though each of U and D can fail to be so.

2.3 Theory Transformations

For any kind of mathematical structures, the question of transformations between instances of this structure is one of central importance. Call $h : A \to B$ a *theory homomorphism* if h is a function from L_A to L_B such that

$$h[C_A(X)] \subseteq C_B(h[X])$$

for all $X \subseteq L_A$. A theory homomorphism is a *consequence-preserving* mapping in the sense that if $X \vdash_A p$ holds, then $h[X] \vdash_B h(p)$ must hold as well. There is also a different way to look at it: let as before the *closure system* $\mathcal{CS}(A)$ be the set of subsets of L_A that are closed under C_A. A closure system, as we noted in the last section, is a complete lattice. But it is also almost the set of closed sets of a *topology*: if C_A fulfils the conditions that $C_A(\varnothing) = \varnothing$ and $C_A(X \cup Y) = C_A(X) \cup C_A(Y)$ as well, it fulfils all the Kuratowski closure axioms. Importing the concept of a *continuous function*—one for which the preimage of an open set always is open—from topologies to closure systems, we can, through use of the following lemma, prove that homomorphisms are exactly the continuous functions in this sense (cf. Lewitzka, 2007 for a similar approach).

Lemma 2.5 : $p \in C_A(X)$ iff p is in all sets in $\mathcal{CS}(A)$ that contain X.

Proof. Assume for contradiction that $p \in C_A(X)$, and that there is a closed set $S \in \mathcal{CS}(X)$ such that $X \subseteq S$ but $p \notin S$. Then by monotonicity $C_A(X) \subseteq C_A(S)$, but by idempotence $C_A(S) = S$, so $C_A(X) \subseteq S$. But then we must have that $p \in C_A(X)$, contrary to our assumption. For the other direction, assume that p is in all closed sets that contain X. Since $C_A(X)$ is a closed set, it must contain p. □

Theorem 2.6 : A function $h : L_A \to L_B$ is a homomorphism iff it is continuous.

Proof. Let $\mathcal{CS}(A)$ and $\mathcal{CS}(B)$ be the closure systems of A and B. Let $h : L_A \to L_B$ be a theory homomorphism. We show that if $Y \in L_B$ is closed, then $h^{-1}[Y]$ is closed as well, since this is equivalent to the same condition on *open* sets, and thus expresses continuity. Let $X = h^{-1}[Y]$, and suppose that $Y = C_B(Y)$. Then since h is a homomorphism, we have that $h[C_A(X)] \subseteq C_B(Y)$, from which it follows that $C_A(X) \subseteq h^{-1}[C_B(Y)] = h^{-1}[Y] = X$. Thus $X = C_A(X)$.

In the other direction, let $h : L_A \to L_B$ be a continuous function, and let $p \notin C_B(h[X])$. Then, by the preceding lemma, there is a closed set $S \subseteq L_B$ such that $h[X] \subseteq S$ but $p \notin S$. Since h is continuous, we have that $h^{-1}[S]$ must be closed as well. Let $q \in h^-1[\{p\}]$. Then, again by the last lemma, we must have that $q \notin C_A(X)$. Applying h on the left gives that $p \notin h[C_A(X)]$. □

Among the theory homomorphisms, some are especially useful. Let a *theory isomorphism* be a bijective homomorphism h such that h^{-1} is a homomorphism as well. Let a *theory embedding* be an injective homomorphism $h : A \to B$ such that

$$h[C_A(X)] = C_B(h[X]) \cap h[L_A]$$

A theory embedding requires the consequence operator of B to correspond exactly to that of A on the image of B in A. It is easy to see that if there is a theory embedding from A to B, then A is isomorphic with a subtheory of B. Theory embedding is, however, in general a somewhat too strong criterion to be really interesting. We call $h : A \to B$ a *theory translation* when h is a homomorphism, which may or may not be injective, for which the embedding condition above holds. A translation still reflects the consequence structure of its domain, but may identify claims in L_A that have the same place in this structure. Every embedding is thus a translation, but not every translation is an embedding.

An example of a translation is the transformation from a propositional language to the Lindenbaum algebra of that language. This takes every sentence p to the set of all sentences equivalent to it, and thus it is not injective. Nevertheless, the Lindenbaum algebra has, in a very clear sense, the *same* consequence structure as the language we started with, even if its cardinality in can be different.

To be a translation is a purely structural property. But consider the theories $A = \langle L_A, C_A \rangle$ and $B = \langle L_B, C_B \rangle$ such that

$$\begin{aligned} L_A &= \{\text{snow is white},\ \text{something is white}\} \\ L_B &= \{\text{grass is green},\ \text{something is green}\} \end{aligned}$$

whose consequence operators allow $p \vdash p$ for any p (as all consequence operators do), and for which

$$\begin{aligned} \text{snow is white} &\vdash_A \text{something is white} \\ \text{grass is green} &\vdash_B \text{something is green} \end{aligned}$$

There is a unique translation from A to B, namely the one that takes “snow is white” to “grass is green”, and “something is white” to “something is green”. But this is surely not a valid translation! “Snow is white” and “grass is green” do not *mean* the same thing at all.

The reason why we can see this is that we are currently using a larger theory (most likely some form of English) that contains both A and B. This theory does *not* allow inferring either “snow is white” from “grass is green” or its converse. We can make these ideas precise by defining an *F-translation* from A to B, where A and B are subtheories of F, as a translation $h : L_A \to L_B$ such that

$$C_F(X \cup \{p\}) = C_F(X \cup \{h(p)\})$$

for all $p \in L_A$ and $X \subseteq L_A$. The rightness of a translation thus depends on which theory we evaluate it in, and F licenses a certain translation h iff that translation only takes claims to claims that are equivalent to them, according to F.

It is worth mentioning that we have not made any reference to *meaning* here. If there is such a thing as absolutely analytic consequence, we can require that F's consequence operator be analytic. Then F will allow only those translations that preserve meaning. But if analyticity, as Carnap held, always is relative to a formal language, all we can say is that F's consequence operator is F'-analytic, for some theory F' of which F is a subtheory. That, in turn, can only mean that F's consequence operator conforms to that of F'.

2.4 Variations on the Theory Theme

The notion of theory that we use is almost absurdly broad. In many cases, we have more structure available, although in others, we actually have even less. This chapter indicates some specialisations and generalisations of the concept used, all of which will be useful further on.

2.4.1 *Formal Theories*

We have approached theories as consequence operators defined on unstructured sets of claims, and this is their most general form. In many cases, however, we have access to further information. One of these is where the language L is a *formal language*, i.e. one which can be generated recursively. But it is not absolutely necessary that L be a language for this kind of structure to be applicable; we may also hold certain thoughts or beliefs to be obtainable from others, by use of preestablished mental transformations, for instance. Hume's *ideas* were of this kind, since according to him the complex ideas were constructed from simple ideas, which are copies of simple expressions (Hume, 1739, Book I, ch. I, sct. I). Leibniz's *terms* also have this structure, since they (at least in one of his interpretations) correspond to natural numbers, and the number of a complex term is the product of the numbers of the terms it is composed of. *Simple* terms are those whose numbers are prime (Leibniz, 1679).

Such a structure will be represented as an *algebra*, which is a mathematical structure of a kind we now briefly will describe. An algebra $\mathfrak{A}$ is a set A (the *carrier*) together with a finite or infinite sequence of functions $\{f_i\}$, $i \in \mathbb{N}$. Each of these (the *operations* of the algebra) is a function from n_i-tuples of elements of A, to elements of A, where n_i, for any i, is a natural number (zero included). We call the sequence $\{n_i\}$ the *signature* of $\mathfrak{A}$.

A slight generalisation of this concept is that of an algebra whose operations admit countable sequences of arguments, rather than merely finite sequences of them. The most important of these for us are the σ-algebras, which are algebras $\mathfrak{S} = \langle \mathcal{S}, \bigcup, ^C \rangle$, such that $\mathcal{S}$ is a set of subsets of some set S, $\bigcup_{i=1}^{\infty} X_i$ is the union of the X_i's, and X^C is the complement of X in S. These algebras are crucial for probability theory, and we will encounter them frequently in this context. We will also consider some slightly more general σ-algebras, where the elements of $\mathcal{S}$ do not need to be sets, and $\bigcup$, ^{C} can be other operations than union and set complement.

We say that an algebra $\mathfrak{B} = \langle B, g_1, g_2, \ldots \rangle$ is a *subalgebra* of another algebra $\mathfrak{A} = \langle A, f_1, f_2, \ldots \rangle$ iff $\mathfrak{A}$ and $\mathfrak{B}$ have the same signature, $B \subseteq A$, and $g_i(x_1, \ldots, x_{n_i}) = f_i(x_1, \ldots, x_{n_i})$ for all i and all $x_1, \ldots, x_{n_i} \in B$. This entails that the carrier of a subalgebra is closed under the operations of that algebra.

If $\mathfrak{A} = \langle A, f_1, f_2, \ldots \rangle$ and $\mathfrak{B} = \langle B, g_1, g_2, \ldots \rangle$ are algebras of the same signature, a *homomorphism* from $\mathfrak{A}$ to $\mathfrak{B}$ is a function φ from A to B, such that $g_i(\varphi(x_1), \ldots, \varphi(x_{n_i})) = \varphi(f_i(x_1, \ldots, x_{n_i}))$ for all i and all $x_1, \ldots, x_{n_i} \in A$. A homomorphism from $\mathfrak{A}$ to $\mathfrak{A}$ is called an *endomorphism* on $\mathfrak{A}$. There is a theorem of universal algebra (see

Grätzer, 1979, p. 36) that says that the image of any endomorphism is a subalgebra of the algebra on which the endomorphism is defined.

Finally, we need the notion of a *free* algebra. For a subset $X \subseteq A$, we say that X *generates* $\mathfrak{A}$ iff every element in A can be obtained by applying the operators of $\mathfrak{A}$ on elements of X some finite number of times. Let an *extension* of a function $\varphi : X \to Y$ be a function $\varphi^+ : X^+ \to Y^+$ where $X \subseteq X^+$, $Y \subseteq Y^+$, and $\varphi^+(x) = \varphi(x)$ for every $x \in X$. $\mathfrak{A}$ is a *free algebra* with the *generators* X iff every function φ from X to the carrier of some algebra $\mathfrak{B}$ with the same signature as $\mathfrak{A}$ can be extended uniquely to a homomorphism from $\mathfrak{A}$ to $\mathfrak{B}$. In a free algebra, every endomorphism is uniquely determined by how it transforms the elements of that algebra's generators. We can thus view the generators as *atomic elements*, and the elements of A as those obtainable by applying the operators of $\mathfrak{A}$ (the "connectives") to these generators. An endomorphism is then a *substitution* of some of the atomic elements elements of A, with arbitrary elements thereof.

We are now ready to define the central concepts of this section. Say that an algebra $\mathfrak{A} = \langle L_A, f_1, f_2, \ldots \rangle$ is a *formalisation* (or an *algebraisation*) of the theory A iff the following condition holds:

(*Structurality*) $\varepsilon[C_A(X)] \subseteq C_A(\varepsilon[X])$, for any $X \subseteq L_A$ and any endomorphism ε on $\mathfrak{A}$.

The structurality condition (which is also called *logicality*, see Wójcicki, 1988, p. 22) essentially says that when we are to determine if something follows or not, we can disregard the specifics of atomic elements, and only look at the structure imposed by the operators. It can equivalently be written as the condition that $X \vdash p$ entails that $\varepsilon[X] \vdash \varepsilon(p)$, so that consequence is preserved under substitutions. This holds in sentential logics, for instance: the atomic sentences are *sentence variables*, which may take on the meaning of any other sentence in the language. Whatever follows from a set of sentences in such a logic, follows from the structure that the connectives (i.e. the operators) have imposed on it. The bearer of consequence for a sentential language is *logical form* – the pattern of connectives in our sentences. This is why we have called the imposition of an algebra on a theory so that structurality holds a *formalisation* of that theory. Further reasons for interest in the condition come from the very general type of semantics it allows, based on so-called *matrices*, which we will encounter in ch. 5.

Most formal theories are propositional languages. In fact, it is very difficult to even formalise predicate logic, since complex predicate-logical sentences are not built using sentences, but using terms, predicates and quantifiers. And even if we limit ourselves to just full sentences, structurality does not hold, since their internal, non-sentential structure influences whether they can be derived from one another. To satisfactorily handle predicate logic algebraically, more complex structures would have to be used.

Another important property that we would like to have in a formalisation is *self-extensionality.* We say that the formalisation $\mathfrak{A}$ is self-extensional iff

$$p_k \dashv\vdash_A q_k$$

for $k = 1 \ldots n$ entails that

$$f(p_1, \ldots, p_n) \dashv\vdash_A f(q_1, \ldots, q_n)$$

for all operations f of $\mathfrak{A}$. A self-extensional formalisation allows us to disregard the specifics of individual claims, and instead concentrate on equivalence-classes of them, even algebraically. If self-extensionality does not hold, logically equivalent claims cannot be substituted for one another. This is the case in certain strongly intensional logics, such as logics of belief that do not allow inference of "a believes that p" from "a believes that q", where p and q are logically equivalent.

The following is an example of a non-linguistic formal theory, which has the structure of classical logic.

Levi's conceptual frameworks. Isaac Levi, in *The Fixation of Belief and Its Undoing* (Levi, 1991), adopts a system of *beliefs* as a basis for his theory of belief revision, as opposed to the more common approach that involves working with sets of sentences (Gärdenfors, 1988). This is interesting as an example of a purportedly non-linguistic theory, which still has a *logical structure.*

A *Levian conceptual framework* is a set B of *potential states of full belief*, partially ordered by a relation $\preccurlyeq$ called *strength*, such that if state a is stronger than state b, then anyone who is in state a believes *more* than someone who is in state b. Alternatively (or equivalently, on Levi's theory), if $a \preccurlyeq b$, then a *entails* b. This ordering is furthermore assumed to have the structure of a *bounded complemented lattice*, i.e. to be such that for every pair of belief states a and b, there is a strongest belief state entailed by them both (their *join* $a \vee b$), a weakest belief state that entails both of them (their *meet* $a \wedge b$), and for any belief state c, there is a belief state c' such that $c \vee c' = \top$ and $c \wedge c' = \bot$, where $\top$ is the unique weakest belief state in the conceptual framework, and $\bot$ is its strongest belief state. It is furthermore required to be *distributive*, which means that we must have, for any belief states a, b and c, that $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$ and $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$.

It is well known that a complemented distributive lattice is equivalent to a *Boolean algebra*, which is the algebra of classic propositional logic. We will therefore use a Boolean algebra for the algebraisation. Let $\langle B, \preccurlyeq \rangle$ be a Levian conceptual framework. A *filter* in such an framework is a subset of B that is closed under entailment and under meet of any two of its elements. We define the *theory* T for this framework to be $\langle B, C \rangle$, where $C(X)$, for any $X \subseteq Y$, is the intersection of all filters in $\langle B, C \rangle$ that contain X. An algebraisation of T is then a Boolean algebra $\mathfrak{T} = \langle B, \wedge, \vee, \neg, \top, \bot \rangle$, such that, for any endomorphism ε on $\mathfrak{T}$, $\varepsilon[C(X)] \subseteq C(\varepsilon[X])$ for all $X \subseteq B$.

2.4.2 *Many-valued theories*

Consequence, as it is usually conceptualised, is very much concerned with the preservation of *truth* and does not say anything about falsity, or any other semantic property. But we ideally would like to use consequence to find out not only about what is true, but also what is false. If we have that $X \vdash p$, and know that p is false, we want to be able to infer that some claim in X has to be false as well.

It may be thought that this information is already encapsulated in a consequence relation. Should not falsity simply be definable as the absence of truth? But this is not proper for all kinds of theories or logics. For example, in a theory in which we allow vague concepts, we may want to admit cases where p is neither true nor false. Defining "p is false" as "$\neg p$ is true" is somewhat better, but is possible only if the right kind of negation is available. Finally, attempting the definition "p is false" $\underset{def}{=}$ "$p \vdash \bot$", where $\bot$ is a known falsity, invites the question of how such a falsity is to be identified.

The proper way to handle these problems seems to me to be to define a consequence operator not on bare claims, but on assignments of semantic values to these claims instead. Writing

$$v : p$$

for the assignment of value v to the claim p, we can then define inference rules like

$$\{t : p \rightarrow q, f : q\} \vdash f : p$$

which captures a version of *modus tollens.*

Consequence for a many-valued theory A is defined as a function on sets of assignments on the theory's language L_A instead of directly on sets of claims. We can still assume such a consequence operator to satisfy the same axioms as before, i.e. reflexivity, idempotence and monotonicity. Using bold-face italics for sets of assignments, we thus write

$$\boldsymbol{Y} \subseteq C_A(\boldsymbol{X})$$

when the assignments in the set $\boldsymbol{Y}$ are inferable from those in the set $\boldsymbol{X}$.

Defining consequence in this way gives us a significant increase in expressiveness. As Carnap discovered, traditional consequence is particularly inept at constraining semantics for propositional languages: any set of inference rules for classical propositional logic permits semantics with more than two truth values, and furthermore semantics where the negation of a false sentence does not have to be true (Carnap, 1943).[3]

But in another sense, defining consequence on assignments rather than claims might *not* seem to incur any essential generalisation. In the simplest case, saying that p is true is equivalent to saying that p. That p is false may not be expressible in all theories, but it is definitely expressible in some. In any case, "p has semantic value v" is often as much a claim as anything else, since it generally can be true or false.

The difference, of course, lies in interpretation. In the many-valued case, we regard the assignment as part of the metatheory, but in terms of traditional consequence, it is part of the object theory. This is similar to the difference between Hilbertian and Tarskian consequence: we can very well see consequence as holding between single claims rather than between sets of claims and single claims, as long as we allow sets of claims to be claims themselves, and keep in mind to interpret a set of claims as true iff all the claims in the set are.

The most important type of many-valued theories for us will be the ones whose assigned set of semantic values is $\{t, f\}$, where t stands for *true* and f for *false*. Such a theory will be called *bivalent*, while one whose set of semantic values is $\{t\}$ will be called *single-valued*. Traditional logic is single-valued, since it is concerned about nothing but preservation of truth.

We can give rules for bivalent consequence, just as for single-valued. The most important one (apart from reflexivity, idempotence and monotonicity), which connects truth and falsity with consequence, is

$$\boldsymbol{X} \cup \{t : p\} \vdash t : q \text{ iff } \boldsymbol{X} \cup \{f : q\} \vdash f : p$$

This rule expresses the principle of contraposition for bivalent consequence relations.

[3] The explanation for this fact is given in Church's review: no amount of axioms can distinguish between Boolean algebras of different cardinality. Since truth corresponds to the top of a Boolean algebra, and negation to complement, any element which is neither top or bottom will be false, and also have a false negation (Church, 1944).

2.4.3 *Probabilistic theories*

Probabilistic consequence gives a generalisation of the standard, deterministic kind. The fundamental idea here is that we want to capture the probability a certain set of premisses give to a claim, rather than merely whether it follows logically or not. Thus, we want to have a collection of consequence operators C^{π}, where $\pi \in [0,1]$, such that

$$p \in C^{\pi}(X) \text{ iff } P(p \mid X) = \pi$$

where $P(p \mid X)$ is conditional probability measure, defined on pairs of single claims and sets of claims. Thus $X \vdash^{\pi} p$ can be read as "the probability of p given the truth of all claims in X is π". We assume that $C^{\pi_1}(X) \cap C^{\pi_2}(X) \neq \varnothing$ implies that $\pi_1 = \pi_2$, so that no claim ever is assigned more than one probability.

How does C^{π} work, for a specific value of π? For $\pi = 1$, we should expect it to be a consequence operator in the regular sense. For other values, we should not. Even if q is true 50% of the time when p is, there is no reason to believe that the same holds when both p and another claim p' are true. This is easiest to see when we take $p' = \neg q$, in which case we should get that $\{p, \neg q\} \vdash^{0} q$ rather than $\{p, \neg q\} \vdash^{0.5} q$. In short, probailistic relations are not monotonic.

One way to proceed is to widen the theory concept to admit non-monotonic consequence operators, and give general axioms for these. Since this will take us too far afield, we will not do so here, but instead concentrate on the intended interpretation. Let A be a theory, and let A_{triv} be the maximal strengthening of A, for which $C_{triv}(\varnothing) = L_A$. Let a *probabilistic theory on* A be a pair $\langle \mathfrak{S}_A, Ev \rangle$, where $\mathfrak{S}_A$ is a σ-algebra $\langle \mathcal{T}'_A, \bigcup, ^C \rangle$, such that $\mathcal{T}'_A \subseteq \mathcal{T}_A$, and $Ev : \mathcal{T}'_A \times \mathcal{T}'_A \to [0,1]$ is a function from pairs of theories included in $\mathcal{T}'_A$ to real numbers in the interval $[0,1]$.

$\mathcal{T}'_A$ gives us the set of subtheories of A for which probabilistic inference is defined. We assume that

- if $B \in \mathcal{T}'_A$, then there is a theory $B^C \in \mathcal{T}'_A$ such that $B \wedge B^C = A_{triv}$ and $B \vee B^C = A$. Furthermore, $(B^C)^C = B$ and $(B_1 \wedge B_2)^C = (B_1^C \vee B_2^C)$.

- if $B_1, B_2, \ldots$ is a sequence of theories in $\mathcal{T}'_A$, then $B_1 \vee B_2 \vee \ldots$ is in $\mathcal{T}'_A$.

It follows, as usual, that since $\mathcal{T}'_A$ is closed under joins and complements, and fulfills the criteria of an *orthocomplemented lattice* (Birkhoff, 1967, p. 52), it is closed under meets as well.[4] The function Ev is to be interpreted so that $Ev(B_1, B_2) = \pi$ holds iff the truth of theory B_1 gives evidence of strength π as to the truth of theory B_2, where this strength is taken to be a conditional probability. We therefore assume Ev to fulfil the conditions

(*i*) $Ev(B_1, B_2) = 1$ iff $B_1 \preccurlyeq B_2$

(*ii*) if $B_1, B_2, \ldots$ is a sequence of theories in A such that $B_i \wedge B_j = A_{triv}$ for all $i \neq j$, then

$$Ev(B_1 \vee B_2 \vee \ldots) = Ev(B_1) + Ev(B_2) + \ldots$$

(*iii*) $Ev(B_1, B_2 \wedge B_3) = Ev(B_1, B_2)\; Ev(B_1 \wedge B_2, B_3)$

The first of these affirms Ev as an essentially *logical* form of conditional probability (cf. Carnap, 1950). A subtheory B_1 gives evidence of strength 1 to a subtheory B_2 iff B_1 A-entails B_2. The second guarantees that evidence is additive over theories that cannot be true together, and the third that conditionalisation works as usual for probabilities.

Using Ev, we can easily obtain a set of probabilistic consequence operators with the desired properties. For each probabilistic theory $\langle \mathfrak{S}_A, Ev \rangle$ on A, define the *probabilistic consequence operator* to be a set C^π_A of functions on $\wp(L_A)$, indexed by real values $\pi \in [0, 1]$, such that

$$p \in C^\pi_A(X) \text{ iff } Ev\,(Th_A(X),\; Th_A(\{p\})) = \pi$$

[4] Requiring complements to exist rules out theories that are built on intuitionistic logic. This is unfortunate, but since we will apply probabilistic theories primarily to quantum mechanics, we will not go into how to generalise the notion of probabilistic theory to theories without complements. For a start, see Roeper and Leblanc, 1999, pp. 182–185.

Unless $\mathcal{T}'_A$ contains all subtheories of A, this will not define $C^\pi_A(X)$ for all values of X. It is, however, the best we can do, since it is impossible to define a measure on *all* subtheories of a given theory in the general case.[5] It is obvious that C^1_A is a consequence operator by definition, but in general probabilistic consequence is very different from regular consequence. For most values of π, it does not fulfil the closure axioms. Not only the monotonicity condition has to be replaced, but also reflexivity: we should not expect $p \in C^0_A(\{p\})$ to hold, for example. Nevertheless, many of our current best theories of the world are probabilistic. The following is an example.

Quantum mechanics. For quantum mechanics, we need to be more careful than for classical mechanics in assigning properties to systems. Let L_{QM} be a set of sentences of the forms

Preparation:	the system is prepared in state ϱ at t.
Measurement:	observable $\mathbf{A}$ is measured at t.
Observation:	the value of observable $\mathbf{A}$ at t is in the set V.

where ϱ is a density operator, $\mathbf{A}$ is an observable, t is a time, and V is a Borel set of real numbers.[6] We use $p, p_1, p_2, \ldots$ for preparation sentences, $m, m_1, m_2, \ldots$ for measurement sentences, $o, o_1, o_2, \ldots$ for observation sentences, and $s, s_1, s_2, \ldots$ to refer to sentences of any one of these classes. Let $t(s)$ be the time mentioned in such a sentence, and where s is a measurement or observation sentence, let $\mathbf{O}(s)$ be the observable involved in it.

[5] Consider, for example, a theory for describing where in a real interval $[0, 1]$ a certain point is, such that each subset X of the interval corresponds to a claim "the point is in X". There is a one-to-one correspondence between claims in this theory and its strengthenings, but as is well-known, it is impossible to define a suitable measure on all the subsets of $[0, 1]$ (Fremlin, 2000, §134B).

[6] A density operator is a positive self-adjoint linear operator with trace 1 on a Hilbert space. An observable is a self-adjoint linear operator. A Borel set is a set constructible from intervals of real numbers by using complement and countable unions and intersections.

A density operator ϱ expresses a probability measure over all possible states of a physical system. Preparation of the system consists in subjecting it to some process such that we can assign probabilities to its states after that process is complete. Measurement consists in the performance of an experiment on the system, and an observation is the result observed through such a measurement.

In quantum mechanics, observables are linear operators on a Hilbert space, and the possible values of a measurement are the eigenvalues of these observables. Let $\mathbf{Q}_V^{\mathbf{O}}$, where $\mathbf{O}$ is an observable and V a Borel set, be a projection operator defined to take every point of the Hilbert space to a point with eigenvalue 1 iff $\mathbf{O}$ takes the same point to a point with an eigenvalue inside V. $\mathbf{Q}_V^{\mathbf{O}}$ can be read as "measuring $\mathbf{O}$ gives a value in V", and is itself an observable called a *question*. As shown by von Neumann (1955, pp. 252–254) and Mackey (1963, ch. 2.2), all observables can be defined in terms of such questions.

The evidence function for QM, and thus also the set of probabilistic consequence functions C^{π}_{QM} is determined by the quantum theory. One of the most central properties of these can be formulated as

$$o \in C^{\pi}_{QM}(\{p, m\})$$

where $t(p) < t(m) = t(o)$, $\mathbf{O}(m) = \mathbf{O}(o)$, and

$$\pi = Tr\left(\mathbf{U^{-1}}(\Delta t)\ \varrho\ \mathbf{U}(\Delta t)\ \mathbf{Q}_V^{\mathbf{O}(m)}\right)$$

Here, Tr is the trace function, V is the value set of the observation o, $\Delta t = t(m) - t(p)$, and $\mathbf{U}(t)$ is a linear operator indexed by real numbers called the *time evolution operator*, which governs how the physical system changes over time when left undisturbed. If the system is isolated, we have

$$\mathbf{U}(t) = e^{i\mathbf{H}t/\hbar}$$

where $\mathbf{H}$ is an observable called the *Hamiltonian*, whose eigenvalues are the total energies of different states of the system. It plays the same role as the Hamiltonian in classical mechanics, but is quite different mathematically, since the quantum-mechanical Hamiltonian is a linear operator, and the classical one a real function. These formulae together give the probability of making a certain observation, given that the system was prepared in a state ϱ, that an observable $\mathbf{O}(m)$ is measured, and that the time between the occurrence of these is $t(m) - t(p)$.

Since the occurrence of an observation also is a kind of preparation, we furthermore need principles for deriving what kind of preparation it is. The quantum mechanical rule for inferring preparations from observations is

$$p' \in C^{1}_{QM}(\{p, m, o\})$$

where $t(p') = t(m) = t(o) > t(p)$, p is a preparation statement with density operator ϱ, and p' is a preparation statement with density operator

$$\varrho' = \frac{\mathbf{Q}_V^{\mathbf{O}(m)} \ \varrho \ \mathbf{Q}_V^{\mathbf{O}(m)}}{Tr(\mathbf{Q}_V^{\mathbf{O}(m)} \ \varrho)}$$

For more complex sets of premisses, we can define consequence recursively. This is easiest if the set of premisses is finite, so we assume this to hold. Time-order the premisses X using a function $ord_X : \mathbb{N} \to \wp(X)$ such that $s \in ord_X(0)$ if $t(s)$ is the earliest time among the premisses, and $s \in ord_X(k+1)$ if $t(s)$ is the next larger time-value in X after that of the sentences in $ord_X(k)$ (such a value exists because we have assumed X to be finite, although the assumption that it is well-ordered by t would suffice as well).

Let the consequence operators $C^{\pi}_{QM}[k]$, where k is a natural number, be defined as

$$s \in C^{\pi}_{QM}[k](X) \text{ iff } s \in C^{\pi}_{QM}(ord_X(k))$$

Using a time-ordering such as ord_X, we can always calculate the probabilities of observations by gradually stepping through the sentences of X. A “collected” consequence operator can be defined as the union of the $C^{\pi}_{QM}[k]$, for all k. This consequence operator can then be extended by adding logical connectives, and it can also be made algebraic, although we do not have space to do so here.

2.5 Necessity and Possibility

The theory space of a theory A embodies the role of A as a *framework*, since it determines what theories about its subject matter are available. Some of these, such as the A-*trivial* theory $Th_A(L_A)$, are generally false, given that A is true, but as we have not ruled out theories containing only true claims in their language, we cannot hold $Th_A(L_A)$ to be *always* false no matter what A is. In order to find the theories *ruled out* by A, we would have to specify not only how it transmits truth, which is what C_A tells us, but also how it transmits falsity. This could be done by using a bivalent consequence operator, as described in section 2.4.2. However, we will avoid this complication for now and take a short cut.

Where X is a set of claims, write $v : X$ for the set of assignments $\{v : p \mid p \in X\}$. Let $R_A(X)$, for a bivalent theory A, be the set of claims assigned the value *false* by C_A, when X is a set of claims assigned the value *true*, i.e.

$$R_A(X) = \{p \in L_A \mid f : p \in C_A(t : X)\}$$

R_A is what Carnap called a *rule of refutation*, which tells us what it takes to prove claims false (Carnap, 1942, p. 157). It is a function on sets of claims, rather than on sets of assignments. It is generally not a consequence operator, since we for any consistent claim p should have $p \notin R_A(\{p\})$.

Rules of refutation, when added to a single-valued theory, extend its power somewhat. They do not give the full power of a bivalent theory, however, since they do not specify what we may infer from the falsity of claims, or from combinations of truth and falsity. Nevertheless, they give a useful intermediary, and they are also easily specifiable from most common logics. We can often define a refutation operator for a single-valued theory as

$$R_A(X) = \{p \in L_A \mid C_A(X \cup \{p\}) = L_A\}$$

and a *set of A-falsehoods* $\perp_A$ as

$$\perp_A = R_A(\varnothing)$$

This does not work for all theories, but only for those that satisfy the principle known as *Ex Falso Quodlibet* or *explosivity.* Thus we rule out some theories whose inference machinery is built on minimal logic (Johansson, 1936), positive logic (Hilbert and Bernays, 1934) or various forms of relevant logics (Belnap and Anderson, 1975), for instance. For true generality, we need a many-valued theory. However, due to the greater familiarity of single-valued logics, we will primarily concentrate on these.

The theory, when used as a framework, is the theory used as *logic*. Conversely, by viewing a theory B as a theory in A, we focus on B as variable and regard A as the fixed theoretical framework: that which is *necessary* from our point of view. This notion of necessity is of course relative to what framework we have used, and since frameworks are theories, we have here a notion of *relative theoretical* necessity: a theory is necessary relative to the theory F iff F entails it, impossible iff F refutes it, and possible relative to F iff it is a theory in F which is not refuted. But, since the only theory in F entailed by F is F itself, and the only theory in F that entails the F-absurd theory is the F-absurd theory itself, modality, when seen as a relation between theories, is a fairly simple matter.

The situation changes somewhat when we go from theories to their claims. Every claim p in a theory F's language corresponds to a theory $Th_F(\{p\})$ called the *principal theory* generated by p. The extension of modality to claims can then proceed by defining a claim to be F-necessary iff its principal theory is F itself, F-impossible iff its principal theory is the F-absurd theory, and F-possible otherwise. It is a trivial matter to check that when $p \in L_F$, p is F-necessary iff $p \in \top_A$, F-impossible iff $p \in \bot_A$, and F-possible otherwise.

How do these concepts tie in with more usual notions of modality? The literature, generally, mentions several types of necessity. Fine (2002), for instance, distinguishes the *metaphysical*, *natural* and *normative* necessities, and takes logical and mathematical necessities to be subspecies of the metaphysical. Kripke famously held that the only *real* necessity is the metaphysical, and that even much of what we take to be “true by definition”, such as that the standard metre is one metre in length, is not really necessary at all.

At a first glance, it might seem that we only can be dealing with *de dicto* necessity here, since consequence concerns claims rather than objects. But this is not clearly so: claims can be indexical, and can thus be *about* things. For instance, we can have a theory in which from “that is red” we can derive “that is coloured”. If a certain thing fulfils “that is red”, then we can draw the inference that it is coloured as well. This inference is then necessary in the theory, but it concerns things outside it as well, and thus shares properties with *de re* necessity. Still, it is *de dicto* necessity that is primary for us, and *de re* necessity will have to be determined in terms of it somehow.

When it comes to *de dicto* necessity, we have all the resources required to describe it completely. Such necessity is fully determined by what sentences in a language L are treated as necessary, possible or impossible. Given any such partition of L, we can define a theory M whose language is L, and whose consequence operator is such that $\top_M$ coincides with the necessary sentences in L and $\bot_M$ coincides with its impossible sentences. Theories are thus able to represent systems of modality.

This, again, makes it clear that the theory *itself* really is nothing but a *structure*. It can be *used* in several ways, some of which are:

(i) To justify an *inference* by showing that the theory’s consequence operator allows that inference.

(ii) To make a *truth claim*, which is to hold that the theory’s consequence operator is *truth-preserving* in its *actual* subject matter.

(iii) To *frame* other theories in, by expressing their consequence operators in terms of the theory’s, or equivalently by showing them to be strengthenings of it.

(iv) To make a *necessity claim*, which is to hold that the theory’s consequence operator is *necessarily truth-preserving*, i.e. that it preserves truth in *all* situations it is applicable to.

It is (iv) that we have encountered here. To say that A is necessary is to claim A in a certain *mode*. Using a many-valued theory, we can make this more precise. Let an *alethic-modal* theory be a theory whose set of semantic values consists of non-empty strings of N's, t's and f's. We may read an assignment such as

$$fNt : p$$

as *it is false that it is necessarily true that* p, or more succinctly, p *can be false.* Of course, these kinds of modalities do not by themselves make up a modal logic as such, since they cannot be embedded inside claims in the same theory. But since many assignments, as we noted, are claims themselves, we can always define larger theories with such assignments as claims, and in these, we are free to introduce sentential connectives. Whether this is reasonable or not depends on whether the *necessity* of p is something that can be true or false. We will not take a stand on this question here.

Given any kind of modality, there is some theory that we can use to represent that modality. For an example, let $L = \langle L_L, C_L \rangle$ be the theory of a language of first-order logic, and let Nec be the set of metaphysically necessary sentences in L_L. We can then define the theory of metaphysical necessity $M = \langle L_M, C_M \rangle$, where $C_M(X) = C_L(X \cup Nec)$, for all $X \subseteq Y$. M, by itself, says nothing about necessity, however. It is only when we use it to make a claim of metaphysical necessity that this notion enters.

Many kinds of modality may be held to flow from the subject matter of the theories themselves. Thus we may sometimes speak about the *canonical* modality of a theory A. A theory of physics, for instance, is most naturally seen as concerned with physical (or nomological) necessity. A theory in mathematics, insofar as it consists of claims derivable from the axioms of, for instance, ZFC set theory, deals with the mathematically necessary. And many theories of metaphysics, as Lowe claims, concern what is metaphysically necessary (Lowe, 1998). So although there may be no *law* that metaphysicians can make only metaphysical necessity claims, we often have reason to interpret them that way, in the absence of contrary evidence.

However, many metaphysicians hold there to be something *special* with metaphysical necessity that makes it more *real* or more *fundamental* than other kinds. We will not have anything to say about this supposed difference, since it will not affect our usage of the concept. But there is *something* to the idea that some forms of necessity are more fundamental than others. If B is a theory in A, then B's inferences can be expressed in terms of A's, and taking these to be necessary means that B-necessity can be reduced to a form of A-necessity. We should therefore ask ourselves whether there is some *most* fundamental theory, which we can use to base any other on. Such a theory would give a *minimal* logic in the true sense of the word.

Section 2.2 introduced two types of relationship between theories: strengthenings and containments. But we can also combine these. Write $A \sqsubseteq B$, and say that A *frames* B, if B is a subtheory of a strengthening of A. Equivalently, we can say that that

$$B \sqsubseteq A \text{ iff } C_B(X) = C_A(X \cup \top_B^A) \cap L_B \text{ for some set } \top_B^A \subseteq L_A$$

A framing theory may be larger than the theories it frames, but its consequence operator can still capture those of its framed theories. The set $\top_B^A$ gives the claims in A that we must hold true to arrive at B's consequence operator from A's. It is easy to see that $\top_B = \top_B^A \cap L_B$ whenever $B \sqsubseteq A$.

The following theorem characterises the framing relation.

Theorem 2.7 : $\sqsubseteq$ is a partial order.

Proof. The only condition that is not trivial is antisymmetry. Assume that $C_B(X) = C_A(X \cup \top_B^A) \cap L_B$ and $C_A(X) = C_B(X \cup \top_A^B) \cap L_A$ for all X. This can hold only if $L_A = L_B$, so $\top_B^A = \top_B$ and $\top_A^B = \top_A$, and we have that $C_A(X \cup \top_B) = C_B(X \cup \top_A)$. But $\top_B \subseteq C_B(X)$ for all X, and $\top_A \subseteq C_A(X)$, so this means that $C_A(X) = C_B(X)$. □

The question we have asked—whether there is a most fundamental framework—can then be posed as: does $\sqsubseteq$ have a top? I.e. is there some theory F such that $A \sqsubseteq F$, for any possible theory A?

There is a simple reason why such a theory cannot exist: it has to contain all theories as subtheories, and since the class of theories is as numerous as the class of sets, it cannot be a set itself. But we may rephrase the question again. *Given* a set of theories, can we always create a theory that frames them all? The answer to this question turns out to be *yes*, as long as we allow the introduction of new claims. Let $\mathcal{X}$ be a set of theories. Define F as a theory whose language is the union of the languages of the theories in $\mathcal{X}$, and whose consequence operator is minimal (i.e. is such that $C(X) = C(Y) \Rightarrow X = Y$). Extend F to a theory F' by adding to F's language, for every theory $A \in \mathcal{X}$, a claim t_A, such that

$$C_{F'}(X \cup \{t_A\}) \cap L_A = C_A(X)$$

It is clear that such an extension is possible, since each instance of consequence $X \cup \{t_A\} \vdash_{F'} p$ is not an instance of $X \cup \{t_B\} \vdash_{F'} p$ unless $A = B$. The claim t_A can be read as "theory A is true", and allows us to import the consequence operator of A into F'. Clearly, $A \sqsubseteq F'$ for all $A \in \mathcal{X}$, so F' frames every theory in $\mathcal{X}$.

F' is not, however, a *minimal* framing theory, so it is not a meet of the $\mathcal{X}$'s, in the lattice-theoretic sense. In fact, generally no such meet exists. It is therefore always possible to adopt a theory that is neutral among a given set of theories, but we have no reason to believe such a theory to be neutral with regard to other theories *not* in the set. The structure of the class of all theories is thus that of a *directed class*, i.e. a partially ordered class in which every set has an upper bound.

This characterisation tells us something about how theories work as frameworks. There is no universal logical framework, even though any selection of theories can be placed in a common one. The theory concept is *indefinitely extendible*, to borrow Dummett's term (Dummett, 1991a, 316–319). Whenever we have some theories, we have a method of making a new theory is not among those we had before. In this it is similar to the concepts of *set* or *ordinal number*.

In fact, we can show this indefinite extendibility in a more direct way. Suppose that we use a theory F as framework. Any theory in F will presuppose the consequence operator of F, and thus none of these will be able to contradict F, without falling into self-contradiction. But it is obvious that any theory *can* be contradicted, as they are really nothing more than inferential systems. So there must be some weaker framework F' in which F can be false, i.e. such that $F \sqsubseteq F'$, but $F' \not\sqsubseteq F$. Any claim can be meaningfully denied.[7]

How is this related to *necessity*? The indefinite extendibility of the theory concept translates to an indefinite extendibility of the concept *possible world.* Suppose that we have a class Ω of all worlds that are possible. Some (but possibly not all) subclasses of these correspond to *possible claims*, namely claims that the actual world is an element of a certain class. Let F_Ω be the theory that has these as language, and which has consequence defined so that $X \vdash p$ iff the intersection of the classes that X correspond to is contained in the class p corresponds to.

Assume that Ω itself corresponds to a claim p_Ω in this theory. This will be the case, for instance, if F_Ω has a classical or intuitionistic negation, an orthonegation, or any other way to form claims true in all worlds. We can, of course, still question whether p_Ω holds. We can say "Ω is a class of ways the world could have been, but it isn't". Thus there must be some world in which p_Ω is false (since non-contradictory claims correspond to non-empty sets of worlds), but this cannot be a world in Ω, so Ω could not have contained all possible worlds to start with.

One could hold, of course, that the worlds we take recourse to in such an extension are not possible, but *impossible.* But this seems to be a mere splitting of hairs. They are certainly impossible from the point of view of F_Ω, but not from the point of view of a weaker theory. They can do the same work, semantically, as possible worlds can, which is to act as elements in sets that correspond to claims. The only difference lies in which inferences they can ground.

[7] Note that I do not say that any theory can be meaningfully *doubted* here; that is a psychological question which philosophers probably are poorly equipped to deal with.

Chapter 3
General Metaphysics

In this chapter, we will try to say something about what the world, or things in general, can be like *in themselves*, i.e. apart from any pregiven connection to a language or theory. However, much of our normal thinking about the world is influenced by classical logic, and thus we begin with trying to find what this logic presupposes about what the world might be like. This investigation is then used as an example of the things we want to be able to say about things: that they are like one another in certain respects, that they are parts of one another, etc.

A framework based on category theory is sketched, which will allow us to approach questions like these without presupposing reality to have a certain structure. Thus, just as the previous chapter treated theories as *sui generis* entities without a specific, given structure, this attempts to do so for metaphysics. We give examples of different types of metaphysics (or model theories), in order to indicate the wealth of options available to us.

Finally, we say a few words on the relation between *model* and *world*. Rather than taking this to involve some kind of structural similarity, we adopt an interpretation according to which the world *is* a model. This will allow us to treat semantics as dealing not only with theory–model relations, but with relations directly to reality as well.

3.1 Classical Models

As we have noted, theories are used to make claims about either the world or some parts or aspects of it (i.e. the theory's subject matter), and these claims are true iff the subject matter is as the claim describes it. This constitutes an intensional way of looking at truth: given that the theory is true, it is about *something* (or possibly *some things*), and which of the claims in the theory's language are true is then determined by what this subject is like. We might say that we hold the subject fixed and ask for its properties.

The notion of *model* allows us to turn this picture around, and approach the matter extensionally. A model, as we will use the term, is anything usable as a representation of the subject of some theory. Since the world can be used to represent itself, and the same holds in general for every existing thing (for example in a so-called "Lagadonian" language, where everything stands for itself), everything is a model. But useful models are in general *epistemically accessible*, in that we can gain knowledge about them either empirically or deductively. The attractiveness of the second method is probably the reason why mathematical objects are so popular as models: these have exactly the properties that follow from those we explicitly attribute to them, and no others.

The important point is that when we treat something as a model, we see it as having its properties essentially, and it is this that allows us to turn the intensional characterisation of truth into an extensional one. Informally, we say that the claim $p \in L_A$ is *true in the model* $\mathfrak{M}$ iff the supposition that the theory A's subject-matter is as $\mathfrak{M}$ represents it entails that p is true. Two models are *A-equivalent* iff the same claims in L_A are true in them.

Since anything can be a model, it is permissible to take A's subject matter to be a model $\mathfrak{A}$ (for *actual*) as well. As $\mathfrak{A}$ does not in general share the nice epistemic properties of mathematical models, we often deal with these instead. Any model $\mathfrak{M}$ will be said to be *appropriate* for the theory A iff $\mathfrak{M}$ is A-equivalent to $\mathfrak{A}$, i.e. iff the claims in L_A that are true in $\mathfrak{M}$ are those and only those that are actually true, no matter which these are.

Any true theory's subject matter $\mathfrak{A}$ is naturally appropriate for that theory, but there may be other models that are, as well. An A-appropriate model $\mathfrak{M}$ is one that, *as far as A is concerned*, is impossible to tell apart from A's actual subject. While the next chapter will deal with how to make talk about truth in models exact, this one will center on the models themselves. It is titled *metaphysics* because, as the models include everything that exists, the study of models includes the study of everything that exists. This is one sense in which, as I have argued, metaphysics *is* model theory.

Things are however not quite as simple as this. Mainstream model theory (henceforward "MMT"—the term is from Hodges's entry in the *Stanford Encyclopedia of Philosophy* (Hodges, 2005)) is quite a different thing from the kind of model theory that we have envisaged. The most important difference is that *being a model*, in MMT, is a relative property: a model is always a model *for a language.* While one may sometimes want to see models as models *of* a subject, we do not want to tie them as strongly to a specific language as MMT does. For us, models are free-floating citizens "in their own right" as well, and it is the job of semantics to connect models to languages, or more generally, to theories.

The discipline of model theory is generally taken to fall under the field of universal algebra, which is a part of mathematics that we already have encountered: any formalisation of a theory is an algebra, and it is a trivial matter to show that any algebra is isomorphic to a formalisation of some theory. MMT adds to the operations in the algebra an ordered set of *relations* defined on the algebra's carrier set, where a relation simply is a set of n-tuples of elements of the carrier. For the rest of this section, we will refer to such a structure as a *Tarskian* model—its usual name in MMT is simply "structure".

The formal definition proceeds as follows: assume that $\mathcal{L}$ is a first-order language with n function symbols and m predicates, where both n and m are at most countable. Then the pair $\langle k_i \rangle_1^n$ and $\langle l_i \rangle_1^m$ of sequences of length n and m such that k_i is the arity of $\mathcal{L}$'s i:th function symbol and l_i is the arity of $\mathcal{L}$'s i:th predicate is $\mathcal{L}$'s *signature.*[1] A Tarskian model for $\mathcal{L}$ is a sequence $\mathfrak{M} = \langle D, f_1, \ldots, f_n, R_1, \ldots, R_m \rangle$, where

[1] We do not intend to exclude any of the cases where $n = 0$, $n = \infty$, $m = 0$, or $m = \infty$ here. When both n and m are 0, the model is essentially just a set.

D is a non-empty set,

f_i is a k_i-ary function on D for all i from 1 to n, and

R_i is an l_i-ary relation on D for all i from 1 to m.

We call D the *domain* of $\mathfrak{M}$, the sequence f_i the *functions* of $\mathfrak{M}$, and the sequence R_i the *relations* of $\mathfrak{M}$. The *signature* of the model is the same as the signature for the language it is a model of, and this is one of the things that makes a Tarskian model so tied to its language. In the language, however, the signature determines the arities of predicates and function symbols, while in the model it stands for arities of functions and relations. These functions and relations are, in turn, subsets of Cartesian powers of D.

There are two things worth noting here, if we are to take a Tarskian model to be the subject of a theory, and in the limit, a representation of the world. First of all, a Tarskian model is *Platonistic* in that it employs non-concrete entities (more specifically, functions and relations created from sets). But it is also in a certain sense *non-extensional*: the models $\langle D, R_1, R_2\rangle$ and $\langle D, R_2, R_1\rangle$ are different (unless $R_1 = R_2$), so the identity of a model is not determined by its domain, which relations hold in it, and how the functions act on the elements the domain. We also need to know which predicates correspond to which relations, and which function symbols to which functions, and this is given by the relations' and functions' positions in the number series. This position, in turn, is a property of the model as a whole (since it is ordered), but not of the relations and functions themselves.

It is essentially a trick of Tarski's to rely on matching index numbers to find out which predicates correspond to which relations.[2] A more explicit approach is to bring in the language $\mathcal{L}$ itself, and see the model as a function from $\mathcal{L}$'s symbols to relations and functions on D. But this makes the model–language tie even tighter, and we are trying to separate the two here. If a model is *too* dependent on its language it is

[2] I do not mean to imply that Tarski *invented* this trick. It is used, among other places, in defining homomorphisms between algebras. For example, there are in general *no* nontrivial homomorphisms between a ring defined as $\langle R, +, \cdot\rangle$ and one defined as $\langle R, \cdot, +\rangle$, even if they have the same signature algebraically.

a fallacy of the same type involved in the "picture theory" to take it to correspond to a way the world can be.

We can however also take the model concept in the other direction. What determines what *exists* in a model $\mathfrak{M}$ is usually taken to be the domain (this is another interpretation of Quine's criterion of ontological commitment). The rest of the model has to do with the *interpretation* of $\mathcal{L}$ in $\mathfrak{M}$. What if we separate these?

Let us call any set M a *thin* model. This is almost as simple and structure-less as models can get, but not quite. A set still has some structure: its cardinality. This, in turn, is the only thing that standard predicate logic preserves unless we fix the interpretation of some non-logical constants.[3] Thus the thin notion of model is also quite natural for predicate logic, but a predicate logic from which we have stripped away the interpretative aspects.

So there are at least two notions of model in MMT floating around—thin and Tarskian. Which one is correct? We do not have to decide, but can take them to be alternative ways of explicating what the world can be like according to classical predicate logic. When we discuss semantics, we will see what differences the choice gives rise to.

As Tarskian models are extensions of universal algebras (for languages with only functional symbols, models *are* universal algebras), structural relationships such as homomorphisms, isomorphisms and embeddings hold between them. An important part of MMT concerns how the existence of such relationships between models corresponds to relationships between the sets of sentences that are true in those models. The next section will develop the theory for structural relationships in general. In this section, we will confine ourselves to those that hold between Tarskian models, and between the sets that make up thin models.

Since models, for us, are representations of parts or aspects of the world, this question is equivalent to the one of how such parts or aspects are related. We will primarily be interested in three types of relationship, which informally can be explained as follows:

- $\mathfrak{M}_1$ is *embeddable in* $\mathfrak{M}_2$ when $\mathfrak{M}_2$'s structure contains $\mathfrak{M}_1$'s.

[3] This is what drives the so-called "Newman problem": since the only officially logical predicate is identity, the only things we can really say about models in standard predicate logic is how many things there are in them.

- $\mathfrak{M}_1$ is *reducible to* $\mathfrak{M}_2$ when $\mathfrak{M}_2$'s structure is obtained by identifying structurally indistinguishable parts of $\mathfrak{M}_1$'s.[4]

- $\mathfrak{M}_1$ is *isomorphic to* $\mathfrak{M}_2$ when they have the same structure.

These all concern the models themselves, rather than their relationships with any language. There are also a couple of interesting relationships that we need to bring in theories and semantics for, such as theoretical (and logical) equivalence, but these will be the subject of chapter 5.

The concepts outlined above are simplest for thin models. Since the only structure a set has is its cardinality, one set X is embeddable in another set Y iff there is an injection from the first set to the second. X is a reduction of Y iff there is a surjection from Y to X (i.e. if every element in X is the image of some, and generally more than one, element in Y). X and Y are isomorphic iff there is a bijection between them. By the Schröder-Bernstein theorem, X and Y are also isomorphic if they are mutually embeddable. Furthermore, if X is embeddable in Y and Y is a reduction of X, then there is, by the axiom of choice, a one-to-one function $g : Y \to X$ as well, so we again have mutual embeddability, and thus isomorphism.

Tarskian models admit more interesting structural relationships. The carriers of these, as in the case of thin models, are still functions between the models' domains, but because Tarskian models have functions and relations defined on them, structure-preserving transformations need to respect these. The fundamental entity here is the homomorphism, which is an extension of the algebraic concept. Formally, a homomorphism $h : \mathfrak{M}_1 \to \mathfrak{M}_2$, where $\mathfrak{M}_1 = \langle D_1, f_1, \ldots, f_n, P_1, \ldots, P_m \rangle$ and $\mathfrak{M}_2 = \langle D_2, g_1, \ldots, g_n, Q_1, \ldots, Q_m \rangle$, is a function h from $\mathfrak{M}_1$'s domain to $\mathfrak{M}_2$'s that fulfils the following requirements:

(i) $\mathfrak{M}_1$ and $\mathfrak{M}_2$ have the same signature.

[4] There is another use of the word "reduction" in MMT, which concerns functions between models of with different signatures. Unfortunately, there seems to be no commonly accepted name for the relationship we use here, so since we will have no use for the other notion of "reduction" in this text, I have appropriated the word. Our use also complies with how the word is used in constructing a "reduced product" of models.

(*ii*) $h(f_i(x_1, \ldots, x_k)) = g_i(h(x_1), \ldots, h(x_k))$ for all i from 1 to n.

(*iii*) If $\langle x_1, \ldots, x_k \rangle \in P_i$ then $\langle h(x_1), \ldots, h(x_k) \rangle \in Q_i$, for all i from 1 to m.

A homomorphism does not "lose" any structure, but the image may have more structure than the preimage. In terms of first-order logic, a homomorphism is a map between models that preserves the truth of *positive existential* sentences, where a positive existential sentence is one equivalent to some sentence that contains no occurrence of any of the symbols $\forall$, $\neg$, $\rightarrow$ or $\leftrightarrow$ (Hodges, 1993, pp.47–49).

The three types of relationship above correspond to different types of homomorphism. Call a homomorphism *strong* if it satisfies not only the left-to-right direction of (*iii*) above, but also the reverse direction, i.e. that $\langle x_1, \ldots, x_k \rangle \in P_i$ *iff* $\langle h(x_1), \ldots, h(x_k) \rangle \in Q_i$. An embedding is then an injective strong homomorphism, and a reduction is a surjective strong homomorphism.

For another viewpoint, we can use the standard semantics of first-order logic to characterise these relationships. An embedding is a homomorphism that preserves the truth of *existential* sentences: those built up from quantifier-free formulas using only $\exists$, $\wedge$ and $\vee$ (Hodges, 1993, pp.47–49). Such a sentence can only assert the existence of things, and not deny any thing's existence or say that something holds for *everything* in a class. This concurs with the intuitive notion of what an embedding is supposed to be, since it means that under the standard semantics, if $\mathfrak{M}_1$ is embeddable in $\mathfrak{M}_2$, then everything that exists in $\mathfrak{M}_1$ exists in $\mathfrak{M}_2$ as well.[5]

Reductions may at first seem somewhat less natural, but they have significant uses as well, since they are generalisations of the algebraically important technique of taking the quotient of an algebra under a congruence relation on it. Semantically, a reduction is a homomorphism that preserves the truth of sentences equivalent to some sentence that contains no occurrence of "=" in a negated context:

[5] We are using a certain structural interpretation of what it means for something to "exist" here. An embedding does not guarantee that the elements of $\mathfrak{M}_1$'s domain themselves are elements of $\mathfrak{M}_2$, but only that the same existential sentences are true. Thus, the existence used is relativeised to a first-order language.

Theorem 3.1 : Let $h : \mathfrak{M}_1 \to \mathfrak{M}_2$ be a surjective homomorphism between Tarskian models. Then h is a reduction iff h preserves the truth of sentences that contain no essential occurrence of "=" in a negated context.

Proof. Due to a theorem of Lyndon (1959, p. 148), a set of sentences is preserved under so-called *Q-maps* iff it is equivalent to a set of *Q-positive* sentences. A Q-map, for some set Q of relations, is a homomorphism h in which $R(h(x_1), \ldots, h(x_n)) \Rightarrow R(x_1, \ldots, x_n)$ for any relation R that is *not* in Q. A Q-positive set of sentences is one in which no essential occurrence of any relation in Q occurs in a negated context. As Lyndon treats identity as a relation among others, it suffices to apply this theorem to $Q = \{\text{"="}\}$. □

Reductions thus mirror the predicate and functional structure of a model, but may identify elements of the domain that have the same place in this structure. For any Tarskian model $\mathfrak{M}$, and any a, b in $\mathfrak{M}$'s domain, let $a \sim b$ iff a and b stand in exactly the same relations in $\mathfrak{M}$ with everything, and the results of all functions in $\mathfrak{M}$ are invariant under the exchange of a with b. Then, a reduction is a function that may identify only those elements a, b for which $a \sim b$. It only disregards "differences without a difference", so to say.

The sufficiency of mutual embeddability for isomorphism does not hold for Tarskian models. The model $\mathfrak{M}_1 = \langle\{-\infty\} \cup \mathbb{R}, \leqslant\rangle$, where $\leqslant$ is the regular ordering of the extended reals, is embeddable in the model $\mathfrak{M}_2 = \langle\mathbb{R} \cup \{+\infty\}, \leqslant\rangle$, and $\mathfrak{M}_2$ is furthermore embeddable in $\mathfrak{M}_1$. But $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are not isomorphic, since $\mathfrak{M}_1$ contains a least element while $\mathfrak{M}_2$ does not, and $\mathfrak{M}_2$ has a largest element, which $\mathfrak{M}_1$ lacks. However, we still have that if $\mathfrak{M}_1$ both is embeddable in and *reducible to* $\mathfrak{M}_2$, then $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are isomorphic, although it is more convenient to wait until the next section for the proof.

This concludes our brief overview of first-order models. How do they stand as representations of reality? There is really no way to tell yet, since we have to know *how* they represent first, and that is given by semantics. The picture they provide of the world (a set with set-theoretically defined relations and functions on) might not be a familiar one, since we are used to thinking of the world as "concrete", and sets as "abstract". That does not mean that the world could not be a Tarskian model after all, just as the fact that quantum fields may be quite unfamiliar or even impossible to imagine except as mathematical objects does not exclude the possibility that the world is made up by them. The world, in its more fundamental aspects, cares little for our intuitions.

This also means that we cannot *assume* that the world is a Tarskian model, however. After all, it might be a quantum field instead, or something else entirely. Thus, what we really need is a notion of *model* that is broad enough to cover these cases, and just about any other as well. The next section contains an attempt at achieving this.

3.2 Abstract Nonsense

At first, the idea of a general theory of models might seem impossible. The appropriate *OED* entry on "model" reads "A representation of structure, and related senses", but how are we to interpret this unless we settle on what to mean by "structure"? Indeed, the difference between different kinds of models may be taken as differences in how "structure" is interpreted. Seen this way, Tarskian models furnish us with an explication of what structure is, although like any explication, others may be better for other purposes.

But there *are* ways to employ structures without settling on specifically what they are. One important tool here is *group theory*, which can be used for describing symmetries (i.e. invariants), even if we do not know the structure of the thing they are symmetries *of*.[6] This cannot

[6] Philosophers who argue that group theory should be used for describing models in this manner include van Fraassen (1989) and French and Ladyman (2003). Furthermore, this approach is very much a paradigm of the theory of measurement, where we usually say that a relation between two measured values corresponds to something in reality iff it is invariant under automorphisms of the scale type (see Suppes, 1959), and the automorphisms of any algebra form a group. But there are

be the whole story when it comes to structure, however: in a group, every transformation has an *inverse*, i.e. a transformation that cancels the effect of the first one, and it is this that limits them to describing *exact identity* of structure. But there are many structural relationships apart from isomorphism, such as the existence of embeddings and more general homomorphisms

Category theory is a part of mathematics well suited for treatment of all kinds of transformations, and not only those having inverses. Analogously to a group, where the fundamental entities are the isomorphisms, the fundamental entities of category theory are structure-preserving transformations called *morphisms* (or sometimes *arrows*, due to the fact that they often are drawn as arrows in diagrams). Almost all mathematical structures form categories, i.e. classes of objects with such morphisms defined on or between them, and much of mathematics can be reformulated in terms only of the properties of these. In short, categories are ideal for representing structure without having to prejudge the question of *what* structure is.

Formally, a category $\mathcal{C}$ is a collection of the following:

(*i*) A class obj, called the *objects* of $\mathcal{C}$. When no possibility of confusion seems likely, we will also use the category's name to refer to the class obj of its objects, and rely on the context to disambiguate whether we mean the entire category or only its object class.

(*ii*) A class hom, called the *morphisms* of $\mathcal{C}$. These are the structure-preserving transformations.

(*iii*) Two mappings $\mathsf{dom} : \mathsf{hom} \to \mathsf{obj}$ and $\mathsf{cod} : \mathsf{hom} \to \mathsf{obj}$. Given $f \in \mathsf{hom}$ such that $\mathsf{dom}(f) = a$ and $\mathsf{cod}(f) = b$, we write this

calls for allowing other structure-preserving mappings here as well: Luce, Krantz, Suppes and Tversky argue in the third volume of their classic treatise *Foundations of Measurement* that we should allow as meaningful relations that are invariant under non-automorphic endomorphisms as well, if there are any (Luce et al., 1990, ch. 22). The endomorphisms of an algebra do not, however, in general form a group, but only a *monoid*: an associative algebraic structure with identity, but without inverses. A different proposal is given by Guay and Hepburn (2009), according to which the appropriate mathematical structure to use for symmetry is the *groupoid*. A groupoid is a group where the binary operation is only partially defined.

as $f : a \to b$, and say that a is f's *domain* (or *source*) and b is f's *codomain* (or *target*). We use the notation $\mathsf{hom}(a, b)$ for the class of morphisms of $\mathcal{C}$ that have the object a as domain and the object b as codomain.

(*iv*) A partial mapping $\circ : \mathsf{hom} \times \mathsf{hom} \to \mathsf{hom}$ called *composition*, such that for any $f, g, h \in \mathsf{hom}$, if $f : a \to b$, $g : b \to c$, and $h : c \to d$, then $f \circ (g \circ h) = (f \circ g) \circ h$. This is usually expressed as the condition that *composition is associative.* We assume $f \circ g$ to be defined iff $\mathsf{cod}(g) = \mathsf{dom}(f)$.

(*v*) A mapping $\mathsf{id} : \mathsf{obj} \to \mathsf{hom}$ such that for any morphism $f : a \to b$ we have that $\mathsf{id}(b) \circ f = f \circ \mathsf{id}(a) = f$. The morphism $\mathsf{id}(a)$ is called a's *identity morphism*, and is also written as 1_a.

The most well-known example of a category is $\mathcal{V}$, whose object class is the class V of all sets, and for which morphisms are set-theoretic functions, dom and cod give these functions' domains and codomains, 1_a is the identity function on the set a, and $\circ$ is function composition. This is, incidentally, also the category that describes the structural relationships of thin models, since these just are sets.

The category we will focus on in this section is that of Tarskian models. The class of all Tarskian models of a given signature Σ forms the object-class of a category $\mathcal{T}_\Sigma$ whose morphisms are the homomorphisms between the models of signature Σ.

Let $\mathcal{T}$ be the category that is the union of all $\mathcal{T}_\Sigma$, for any signature Σ. Categories of models such as $\mathcal{T}$ or $\mathcal{V}$ will be referred to by us as *model spaces.* Since a model, as we have used the term, is the representation of how the subject of a theory can be, a model space is intended to represent all possible ways a potential subject for a theory can be. The semi-formal definition is as follows.

Definition 3.1 : A *model space* is a category $\mathcal{M}$ where

- obj is a class of *models* — for our purposes, some kind of entities assumed to have some kind of structure.

- hom is the class of all structure-preserving mappings between elements of obj.

- $\mathsf{dom}, \mathsf{cod}, \circ, \mathsf{id}$ are the domain, codomain, composition and identity mappings for obj and hom.

As advertised, we take the notion of structure as unanalysed: there is simply no formalism general enough to describe the interior of every possible kind of structure. The usual representations, such as a set with a set of relations on it, all depend on prior ontological assumptions, such as that structures have atomic parts, and are constituted by relations-in-extension over these parts. We will of course use several representations of structures in this book, but none of these are to be taken as explications of the notion of structure *in general.* Category theory, on the other hand, allows us to work with structures "from the outside"—in terms of what they *do* rather than what they *are*—and so the rest of this section will be devoted to the purely category theoretical aspects of model spaces.

The morphisms themselves may sometimes be quite hard to interpret: one standard textbook on category theory (McLarty, 1992, p.5) describes them, abstractly, as a "kind of picture" of the domain in the codomain, but hastens to add that this does not really tell us much so long as we do not know what the domain is like. We have said that they are "structure-preserving", but even this is fairly vague. What we do *not* mean is that the codomain has to contain the *same* structure as the domain. It must contain at least the structure of the domain, but may be more structured as well.

In $\mathcal{T}$, where morphisms are the homomorphisms of the preceding section, we noted that these are the maps between models that preserve the truth of existential-positive sentences. The "preservation of structure" involved here is the preservation of fundamental (atomic) relations, in the sense that if $h : \mathfrak{M}_1 \to \mathfrak{M}_2$ is a homomorphism, then all the fundamental relations that hold in $\mathfrak{M}_1$ also hold in the image of $\mathfrak{M}_1$ under h. But there can also be other fundamental relations that hold in h's image, and non-fundamental relations (i.e. those that hold because of the recursive specification of the satisfaction relation) may

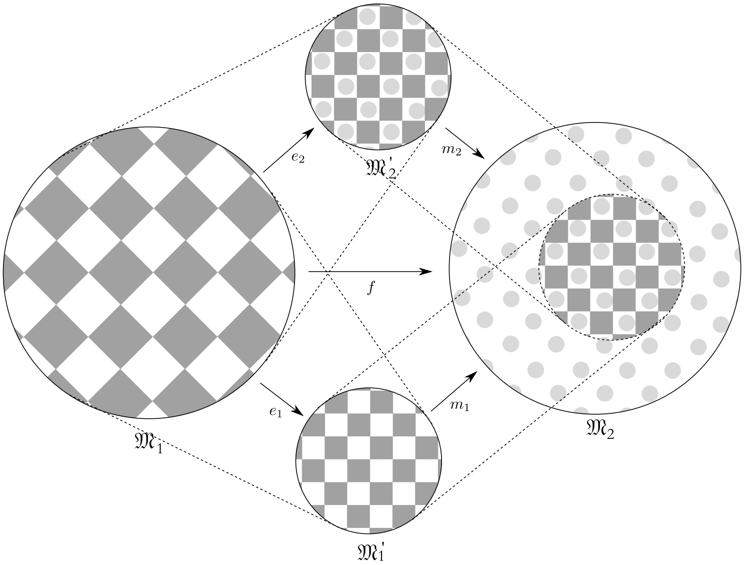


Figure 3.1: ***Two factorisations of the morphism*** $f : \mathfrak{M}_1 \to \mathfrak{M}_2$***.***

cease to hold when h is applied. Parts of $\mathfrak{M}_2$ that are outside h's image may also be entirely different.

As a general principle, we can often envisage a transformation $f : \mathfrak{M}_1 \to \mathfrak{M}_2$ as composed of two factors: some kind of internal change e_1 in $\mathfrak{M}_1$ that turns it into a model $\mathfrak{M}_1'$, and then an insertion m_1 of $\mathfrak{M}_1'$ inside $\mathfrak{M}_2$. Alternatively, we can view f as first taking $\mathfrak{M}_1$ to a *part* of $\mathfrak{M}_2$ by a transformation e_2, and then embedding this part into $\mathfrak{M}_2$ by an identity transformation m_2. The alternatives are illustrated in figure 3.1, for a model space whose models consist of simple patterns.

Here, the pattern in $\mathfrak{M}_1$ is included in $\mathfrak{M}_2$ by the morphism f, and we have envisaged f as being composed of first a rotation and scaling e_1, and then an insertion, or "pasting" of $\mathfrak{M}_1'$ into $\mathfrak{M}_2$. The other possibility is to factor f as $m_2 \circ e_2$, where e_2's codomain $\mathfrak{M}_2'$ is a part of $\mathfrak{M}_2$. Using this second path, $\mathfrak{M}_2'$'s pattern is not only pasted onto $\mathfrak{M}_2$'s but actually appears exactly as it is inside $\mathfrak{M}_2$. Another way to express this is that knowing about the part of $\mathfrak{M}_2$ that $\mathfrak{M}_2'$ is mapped onto tells us everything about what $\mathfrak{M}_2'$ is like, but *not* everything about $\mathfrak{M}_1'$, even though $\mathfrak{M}_1'$ is mapped onto the very same part by m_1.

In the category of sets $\mathcal{V}$, every function $f : X \to Y$ can be split into a surjection s and an injection i. This splitting is not unique, of course: we can let the surjection s take X to a subset of Y, or to a subset of X, or to some other set entirely. What we have is that if $f = i_1 \circ s_1$, and $f = i_2 \circ s_2$, where i_1 and i_2 are injections and s_1 and s_2 are surjections, the codomain of s_1 (and thus also the domain of i_1) must have the same cardinality as the codomain of s_2. In $\mathcal{V}$, this means that they are isomorphic:

$$\begin{tikzcd}
X \arrow[r, "s_1"] \arrow[d, "s_2"'] & Z_1 \arrow[d, "i_1"] \arrow[dl, shift left] \\
Z_2 \arrow[r, "i_2"'] \arrow[ur, shift left, "isom."] & Y
\end{tikzcd}$$

Some general taxonomy would be useful in order to be able to differentiate between these kinds of transformations for arbitrary model spaces, and it turns out that category theory supplies us with just the concepts that we need.

An *isomorphism* is a morphism $f : a \to b$ for which there is some morphism $f^{-1} : b \to a$ such that $f \circ f^{-1} = 1_b$ and $f^{-1} \circ f = 1_a$. If there is an isomorphism between the elements a and b, we say that they are *isomorphic* and write this as $a \simeq b$. This notion captures, in a wholly abstract way, what it is for two elements of $\mathsf{obj}_{\mathcal{M}}$ to have exactly the same structure. If we have that $a \simeq b \Rightarrow a = b$ for all $a, b \in \mathcal{C}$, the category $\mathcal{C}$ is called *skeletal*.

A *monic* (or *monomorphism*) is a morphism $f : b \to c$ such that, for any morphisms $g_1, g_2 : a \to b$, we have that

$$f \circ g_1 = f \circ g_2 \Rightarrow g_1 = g_2$$

or as it is put mathematically, that any application of f is *left cancellable* — whenever we apply f to some object, we can find out what object it was that we applied it to from just knowing the result. In the category of sets $\mathcal{V}$, the monics are exactly the injective functions, and intuitively we can see a monic as one that is *ontology-preserving.* It does not in general have to preserve structure, however, as we will see shortly.

An *epic* (or *epimorphism*) is the opposite of a monic: a morphism $f : a \to b$ for which it holds, for any morphisms $g_1, g_2 : b \to c$, that

$$g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$$

or in mathematicians' parlance, that f is *right cancellable.* In the category of sets, the epics are the surjective functions, but all that can be said in general is that the image of an epic covers so much of its codomain that any two different morphisms from it must differ at some point in that image. *As far as the morphisms are concerned*, an epic therefore is surjective, but for it to be surjective in some more substantial sense, we need to have enough morphisms available in the category.

An example of where an epic fails to be surjective is the category of algebras of a given signature. Any homomorphism h into a free algebra $\mathfrak{F}$ whose image contains $\mathfrak{F}$'s generating set is monic, since the image of any function from this set must fix the values of any homomorphisms from $\mathfrak{F}$. But h does not need to have an image that contains all of $\mathfrak{F}$ — it is sufficient that it covers enough of it, so that the algebraic operations themselves can be used to determine the other values.

It is a standard exercise in category theory to check that any isomorphism is both monic and epic. The opposite does not hold in general: a morphism may be both monic and epic, in which case we call it a *bimorphism*, and yet fail to be an isomorphism. A category for which monicity and epicity together imply isomorphism is called *balanced*; these include the category of sets, and more generally, any so-called *topos*, which is the categorical form of most logics.

Since these concepts are defined without reference to any internal structure of the category's objects, we can use them to characterise model spaces without going into what models are. Thus we hold that, for example, the morphisms in a model space $\mathcal{M}$ are to be defined so that $\mathfrak{M}_1 \simeq \mathfrak{M}_2$ iff $\mathfrak{M}_1$ and $\mathfrak{M}_2$ have the same structure. The isomorphisms form a group under composition as we expect them to, and thus this part of our theory coincides with earlier group theoretic accounts.

The existence of an isomorphism always expresses identity of structure. But monics and epics do not always correspond to the informal notions of "embedding" and "reduction" that we introduced in the last section. Take, for instance, the model space $\mathcal{T}$. While a monic homomorphism $m : \mathfrak{M}_1 \to \mathfrak{M}_2$ in $\mathcal{T}$ does have to take every element in $\mathfrak{M}_1$'s domain to a unique element in $\mathfrak{M}_2$'s, it can be shown that this does not preclude relations holding between elements in m's image that did not hold in $\mathfrak{M}_1$.

Theorem 3.2 : In $\mathcal{T}$, the monics are the injective homomorphisms, and the epics are the surjective homomorphisms.

Proof. The result about monics follows from the fact that $\mathcal{T}$ is a *construct* (see the next section) and that it has a free object over a singleton set through a standard result of category theory (Adámek et al., 2004, §8.29). Proving that surjective homomorphisms are epic is trivial. For the other direction, assume that $e : \mathfrak{M}_1 \to \mathfrak{M}_2$ is an epimorphism, and construct a model $\mathfrak{M}_3$, of which $\mathfrak{M}_2$ is a submodel, such that $D_{\mathfrak{M}_3} = D_{\mathfrak{M}_2} \cup \{*\}$, $R(*, \ldots, *)$ holds for all relations R in $\mathfrak{M}_3$, and $o(\ldots, *, \ldots) = *$ for any operation o in $\mathfrak{M}_2$. Let $f, g : \mathfrak{M}_2 \to \mathfrak{M}_3$ be homomorphisms such that $f(x) = *$ if $x \in e[D_{\mathfrak{M}_1}]$ and $f(x) = x$ otherwise, and $g(x) = *$ for all x. We must have that $f \circ e = g \circ e$, but since e is an epic, this means that $f = g$. But this can only hold if the image of e is the whole of $D_{\mathfrak{M}_3}$ □

What we need to properly capture embeddings and reductions are strengthenings of the notions of epic and monic. While the problem of finding a purely category-theoretic notion of embedding is far from solved (see Adámek et al., 2004, chs. 7, 8), there are several such strengthenings available. One that seems especially congenial for us is the notion of *strong* monic (or epic). It can be characterised as follows.

A *preorder* is usually defined as a transitive and reflexive binary relation, and an *order* as a preorder which is antisymmetric. Set inclusion, as well as parthood, are both examples of orders. We can form a preordered set from a category by letting $a \leqslant b$ iff there is some morphism $f : a \to b$, and we call the category $\mathcal{C}$ an *order* iff there is at most one morphism $f : a \to b$ for each pair of objects a and b.

Let an *inclusion system* for the model space $\mathcal{M}$ be a pair $\mathcal{E}_\mathcal{M}, \mathcal{I}_\mathcal{M}$ such that

(*i*) $\mathcal{E}_\mathcal{M}$ and $\mathcal{I}_\mathcal{M}$ are categories that contain the same models as $\mathcal{M}$, and no morphisms other than those in $\mathcal{M}$.

(*ii*) Every morphism in $\mathcal{E}_\mathcal{M}$ is an epimorphism in $\mathcal{M}$.

(*iii*) $\mathcal{I}_\mathcal{M}$ is an order, whose morphisms are monic in $\mathcal{M}$. The morphisms in $\mathcal{I}_\mathcal{M}$ are called the *canonical embeddings* with respect to this inclusion system.

If every morphism f in $\mathcal{M}$ can be factored as $f = m \circ e$, where $m \in \mathsf{hom}_{\mathcal{I}_\mathcal{M}}$ and $e \in \mathsf{hom}_{\mathcal{E}_\mathcal{M}}$, we say that $\mathcal{E}_\mathcal{M}, \mathcal{I}_\mathcal{M}$ is a *complete* inclusion system for $\mathcal{M}$. The category $\mathcal{I}_\mathcal{M}$ in such an inclusion system shares many of the important properties of inclusion or parthood, and this means that we can take it as an explication thereof. The morphisms in $\mathcal{E}_\mathcal{M}$ can then be *interpreted* as surjective, i.e. as taking their domain to the whole of their codomain, even if we do not necessarily have this for epics in general. The completeness condition guarantees that all transformations in $\mathcal{M}$ can be seen this way.

In a complete inclusion system, the categories $\mathcal{I}_\mathcal{M}$ and $\mathcal{E}_\mathcal{M}$ have a very useful property called *diagonalisation*. Given any $e : \mathfrak{M}_1 \to \mathfrak{M}_2$ in $\mathsf{hom}_{\mathcal{E}_\mathcal{M}}$ and any $m : \mathfrak{M}_3 \to \mathfrak{M}_4$ in $\mathsf{hom}_{\mathcal{I}_\mathcal{M}}$, and any two morphisms $f : \mathfrak{M}_1 \to \mathfrak{M}_3$ and $g : \mathfrak{M}_2 \to \mathfrak{M}_4$ in $\mathsf{hom}_\mathcal{M}$ such that the diagram

$$\begin{tikzcd}
\mathfrak{M}_1 \arrow[r, "e"] \arrow[d, "f"'] & \mathfrak{M}_2 \arrow[d, "g"] \\
\mathfrak{M}_3 \arrow[r, "m"'] & \mathfrak{M}_4
\end{tikzcd}$$

commutes, there is a unique morphism $h \in \mathsf{hom}_\mathcal{M}$ such that

$$\begin{tikzcd}
\mathfrak{M}_1 \arrow[r, "e"] \arrow[d, "f"'] & \mathfrak{M}_2 \arrow[d, "g"] \arrow[dl, "h"'] \\
\mathfrak{M}_3 \arrow[r, "m"'] & \mathfrak{M}_4
\end{tikzcd}$$

commutes. This is thus a necessary condition for the morphisms in $I_{\mathcal{M}}$ to be interpretable as inclusions: if $\mathfrak{M}_3$ is a part of $\mathfrak{M}_4$, then f takes $\mathfrak{M}_1$ to a part of $\mathfrak{M}_4$, and thus there has to be a morphism h such that $f = h \circ e$ and $g = m \circ h$, namely g itself. It is not a sufficient condition, however, and we cannot give a purely structural condition sufficient for a morphism to be an inclusion. This is due to the fact that an inclusion by its nature preserves identity, and identity is not a structural property — there is no way to ensure $a = b$ by only giving structural properties of a and b.

If the diagonalisation property holds for the classes E, M of morphisms, we say that the morphisms in E are orthogonal to those in M. It is worth noting that orthogonality in this sense is non-symmetric, since the morphisms f and g in the above diagram do not have to be reversible. A monomorphism that is orthogonal to the class of all epimorphisms is called *strong*, and strong monomorphisms are very well suited to be taken as explications for what we in the preceding section called embeddings of models. Indeed, this property follows from the preformal understanding of an embedding $m : \mathfrak{M}_1 \to \mathfrak{M}_2$ as being an isomorphism from $\mathfrak{M}_1$ to a part of $\mathfrak{M}_2$, since the inclusion morphisms in any complete inclusion system satisfy it, and isomorphisms preserve all structural properties.

Inclusion is one type of embedding, but usually not the only one. Strong monomorphisms also have the following properties that make them suitable for this task.

(*i*) The composition of two strong monomorphisms is again a strong monomorphism. This means that embeddability is transitive.

(*ii*) A monomorphism that is both strong and epic is an isomorphism, unlike monomorphisms in general.

(*iii*) Strong monomorphisms are *extremal*, which means that if $m : \mathfrak{M}_1 \to \mathfrak{M}_2$ is a strong monomorphism, and we can factor m as $m = f \circ e$ where e is epic, then e must be an isomorphism:

$$\begin{tikzcd}
\mathfrak{M}_1 \arrow[rr, "m"] \arrow[dr, "e"'] & & \mathfrak{M}_2 \\
& \mathfrak{M}_3 \arrow[ur, "f"'] &
\end{tikzcd}$$

> Since, as is quickly proved, the first factor in any factorization of a monomorphism must also be a monomorphism, e above is a monomorphism as well. The extremalness condition guarantees that there is no way to change the structure of $\mathfrak{M}_1$ essentially, and place it inside $\mathfrak{M}_2$, which is equivalent to how it is placed there by m. Or, in other words, $\mathfrak{M}_1$ is placed "as is" in $\mathfrak{M}_2$.

While the canonical monomorphisms are relative to an inclusion system, strongness depends only on the category. In $\mathcal{V}$, every monomorphism is strong, and in $\mathcal{T}$, strong monomorphisms coincide with embeddings, which is another reason for us to adopt them as an explication of this concept.

Theorem 3.3 : In $\mathcal{T}$, a homomorphism is a strong monomorphism iff it is a model embedding.

Proof. Let $m : \mathfrak{M}_1 \to \mathfrak{M}_2$ be a strong monomorphism. From the monomorphism condition, it follows that m is an injection. Let m' be the same function as m, but defined on the model $\mathfrak{M}_2^{sub}$ which is the submodel of $\mathfrak{M}_2$ generated by $m[D_{\mathfrak{M}_1}]$. Now, this morphism m' has to be an epimorphism, and thus, by the strongness condition, there must be a unique morphism $h : \mathfrak{M}_2^{sub} \to \mathfrak{M}_1$ such that $m' \circ h = 1_{\mathfrak{M}_2^{sub}}$ and $h \circ m' = 1_{\mathfrak{M}_1}$. But this means that m must be an isomorphism onto a submodel of $\mathfrak{M}_2$, and thus an embedding.

$$\begin{tikzcd}
\mathfrak{M}_1 \arrow[r, "m'"] \arrow[d, equal, "1_{\mathfrak{M}_1}"'] & \mathfrak{M}_2^{sub} \arrow[dl, "h"] \arrow[d, "incl."] \\
\mathfrak{M}_1 \arrow[r, "m"'] & \mathfrak{M}_2
\end{tikzcd}$$

For the other direction, let $m : \mathfrak{M}_1 \to \mathfrak{M}_2$ be an embedding, let $e : \mathfrak{M}'_1 \to \mathfrak{M}'_2$ be an epic, and let $f : \mathfrak{M}'_1 \to \mathfrak{M}_1$ and $g : \mathfrak{M}'_2 \to \mathfrak{M}_2$ be homomorphisms such that $m \circ f = g \circ e$. We need to show that there is a unique homomorphism h such that $f = h \circ e$ and $g = m \circ h$. Since m is an embedding, it is an isomorphism i onto a submodel $\mathfrak{M}_2^{sub}$ of $\mathfrak{M}_2$. We can define h such that $h(x) = i^{-1}(g(x))$, for all $x \in D_{\mathfrak{M}_1}$. This is well defined since the image of g must coincide with that of m because e is an epic and the original diagram commutes, and it is a homomorphism because it is the composition of a homomorphism and an isomorphism. □

There is also a form of epimorphism that will be of metaphysical interest. Returning to fig. 3.1, we have noted that there are two ways to factor the transformation f. The first is as an epimorphism followed by a strong monomorphism, but we can also see f as something stronger than an epimorphism, followed by a monomorphism that may be non-strong, which is illustrated in the factorisation $f = m_1 \circ e_1$.

The general epimorphism concept can often be interpreted as a generalisation of the method of identifying parts of a structure by means of an equivalence relation. This does not guarantee that the parts identified have the same structural relationships, and thus it is only compatible with a model's structure in very simple cases, such as when the model is a set. Intuitively, an epimorphism can identify any parts, and not only ones that are congruent. For a more appropriate conception, we shall again make use of the property of being strong, although in this case, being a strong epic rather than a monic.

Analogously to the case with monics, we call an epic *strong* when it is orthogonal to all monics. For $\mathcal{T}$, the following holds.

Theorem 3.4 : In $\mathcal{T}$, a homomorphism is a strong epimorphism iff it is a reduction.

Proof. Let $e : \mathfrak{M}_1 \to \mathfrak{M}_2$ be a strong epimorphism, and write e as $e = m \circ e'$, where $e' : \mathfrak{M}_1 \to \mathfrak{M}_1^{sub}$, $i : \mathfrak{M}_1^{s}ub \to \mathfrak{M}_2$, and $\mathfrak{M}_1^{sub}$ is a submodel of $\mathfrak{M}_1$, such that m is an injection:

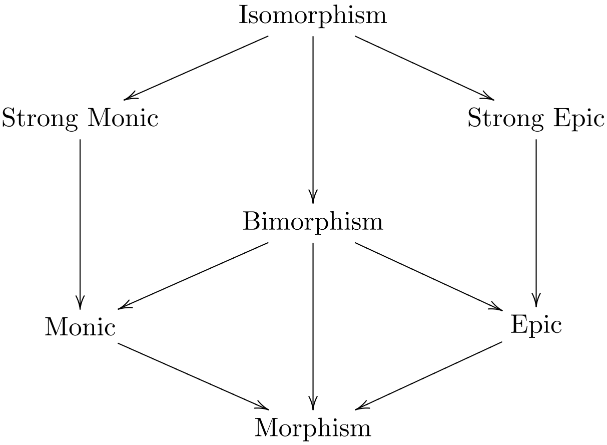
$$\begin{tikzcd}
\mathfrak{M}_1 \arrow[r, "e"] \arrow[d, "e'"'] & \mathfrak{M}_2 \arrow[d, equal, "1_{\mathfrak{M}_2}"] \\
\mathfrak{M}_1^{sub} \arrow[r, "m"'] & \mathfrak{M}_2
\end{tikzcd}$$

Since injections are monomorphisms, we can apply the strongness condition to form a homomorphism $h : \mathfrak{M}_2 \to \mathfrak{M}_1$ such that the diagram commutes, and $e \circ h = 1_{\mathfrak{M}_2}$. Since h has to preserve relations, we have that the reduction condition holds.

To prove the other direction, assume that $e : \mathfrak{M}_1$ is a reduction. We need to show that whenever there is a monomorphism (i.e. an injective function, in our case) $m : \mathfrak{M}_1' \to \mathfrak{M}_2'$, and homomorphisms $f : \mathfrak{M}_1 \to \mathfrak{M}_1'$ and $g : \mathfrak{M}_2 \to \mathfrak{M}_2'$ such that these all commute, there is a unique $h : \mathfrak{M}_2 \to \mathfrak{M}_1'$ such that $f = h \circ e$ and $g = m \circ h$. Let e^{inv} be a function from $\mathfrak{M}_2$ to $\mathfrak{M}_1$, such that $e \circ e^{inv} = 1_{\mathfrak{M}_2}$ (such a homomorphism exists because e is a reduction). We can then let $h = f \circ e^{inv}$. □

In $\mathcal{V}$, just as all monomorphisms are strong, all epimorphisms are strong. In general, we take the notion of strong epimorphism as an explication of the preceding section's concept of reduction.

To summarise, the types of morphism we have introduced can be ordered as follows, where the arrows represent entailment.



3·3 Model Space Mappings

In the last section, we characterised relationships between models in terms of their categorical properties, and more specifically with respect to ways to factor transformations between different models. But that this is possible is a substantial assumption. In particular, for a model space $\mathcal{M}$, we can have that there is *no* way to split its morphisms into subcategories $\mathcal{E}_{\mathcal{M}}$ and $\mathcal{I}_{\mathcal{M}}$ that make up a complete inclusion system. For example, it may be that the part of $\mathfrak{M}_2$ that f takes $\mathfrak{M}_1$ to is unable to "stand on its own" in that it presupposes other parts of $\mathfrak{M}_2$ which are not included in the image of f.

A related problem appears when we consider combinations of models. In $\mathcal{T}$, we can always embed two models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ inside a third model $\mathfrak{M}_3$, and we can even do this canonically: we just let the domain of $\mathfrak{M}_3$ be the union of the domains of $\mathfrak{M}_1$ and $\mathfrak{M}_2$, and define the relations and functions accordingly. But in general, it may be that some possibilities are *exclusive* in the sense that only one of them can be actual.

What are we to do then? To begin with, we may note that these problems appear because of a lack of objects in $\mathcal{M}$. The usual way to handle such a lack in a mathematical structure is to embed that structure in a larger one. In our case, we may embed $\mathcal{M}$ inside a larger model space $\mathcal{M}'$ which has the missing objects. From the point of view of $\mathcal{M}$, such models are *impossible*, i.e. they do not correspond to ways things can be. The process can thus be compared to the practice of introducing impossible worlds to deal with semantics for nonnormal modal logics. From the perspective of $\mathcal{M}'$, however, there is nothing impossible about the added models. Our interpretation of possibility for models is as relativistic as the one we have used for theories. Just as in that case, we do not want to exclude the coherence of some absolute notion, but we do not want to presuppose it either.

Since a model space is a category, embeddings of model spaces are embeddings of categories. These are most succinctly characterised as a type of transformation between the categories themselves, or as it is called in category theory, a *functor*. Formally, a functor $F : \mathcal{C}_1 \to \mathcal{C}_2$ is a function from $\mathsf{obj}_{\mathcal{C}_1}$ to $\mathsf{obj}_{\mathcal{C}_2}$, together with a function from $\mathsf{hom}_{\mathcal{C}_1}$ to $\mathsf{hom}_{\mathcal{C}_2}$, such that $F(f \circ g) = F(f) \circ F(g)$ and $F(1_a) = 1_{F(a)}$ for all morphisms f, g and any object a in $\mathcal{C}_1$. This is usually expressed as the requirement that F has to preserve composition and identities.

A functor is called *faithful* iff it takes no two morphisms between the same objects to the same morphism. It is a *category embedding* iff it is injective on the objects, and a *full* category embedding iff it also is surjective on the morphisms. The notion of a faithful functor is strictly weaker than that of embedding, since a faithful functor still can identify objects, so long as no morphisms are identified in the same set of morphisms between objects a and b. A full embedding can be seen as a selection of some of the objects of a category, together with *all* the morphisms between these, and a *subcategory* as an embedding that takes every object and morphism to itself.

In an inclusion system, both $\mathcal{E}_{\mathcal{M}}$ and $I_{\mathcal{M}}$ are subcategories of $\mathcal{M}$, though in general neither is full. As we mentioned, it can be that not every morphism in $\mathcal{M}$ can be written in terms of the elements of such sets. We can still always fully embed $\mathcal{M}$ in a model space in which such factorisations are possible, for instance by using the so-called Yoneda embedding (Awodey, 2006, pp. 160–167), which reinterprets a category in terms of functors from that category to the category of sets.

How should such an embedding, and the additional models it introduces, be interpreted? Continuing the analogy with mathematical structures, we can see them as *ideal* models, i.e. idealisations of the models in $\mathcal{M}$. In the extended model space $\mathcal{M}'$, we are free to combine models as we wish, and also to speak about intersections of arbitrary sets of models. In this sense, $\mathcal{M}'$ can be seen as a kind of completion of $\mathcal{M}$.[7]

But, if $\mathcal{M}$ is a collection of ways something can be, what is $\mathcal{M}'$? An hypothesis is that the ideal models added by going from $\mathcal{M}$ to $\mathcal{M}'$ can be taken to be *aspects* of things. Since they, from the point of view of $\mathcal{M}$ cannot exist on their own, they are not fit to be seen as objects or things in the standard metaphysical sense. Yet, they represent things that can be *in common* among models, even if these things are not self-subsistent. $\mathcal{T}$ is complete in itself, so all aspects are models in this space. In the next chapter, we shall encounter the model space $\mathcal{N}$ for which this does not hold.

Embedding one category in another is an example of a reinterpretation of a model space. Another such example is given by the existence of a faithful functor F from $\mathcal{M}$ to another category $\mathcal{C}$, in which case $\mathcal{M}$ is called a *concrete category over* $\mathcal{C}$, and F is called a *forgetful* functor (since it "forgets" the possible extra structure that may be encoded in $\mathcal{M}$'s morphisms). If F takes $\mathcal{M}$ to the category $\mathcal{V}$, the pair $\mathcal{M}, F$ is called a *construct.*[8]

Many model spaces can be seen as constructs, since their models are built up from sets in some sense. $\mathcal{T}, F$, where F is a functor that takes each model to its domain, and each homomorphism to its underlying

[7] Not least in the sense that $\mathcal{M}'$, if we use the Yoneda embedding, is a so-called *complete* category.

[8] Since interpreting structures in terms of sets is so common in mathematics, it is usual in category theory to use the notion "concrete category" to denote what we have called a construct. Since we will be interested not only in concretising model spaces over $\mathcal{V}$, but over other categories as well, we have retained the more general interpretation.

function, is a construct. Another example is the model space $\mathcal{N}$ of the next chapter. The advantage of a construct, as will be explored in ch. 6, is that the ontologies of models become very transparent, since we can take $F(\mathfrak{M})$ to be the set of things existing in a model. This in turn means that there is a straightforward way to define inclusion systems on such models. Given any model space $\mathcal{M}$, let the inclusion system $\mathcal{E}_{\mathcal{M}}, I_{\mathcal{M}}$ be *induced* by the faithful functor $F : \mathcal{M} \to \mathcal{V}$ iff the morphisms $i \in \mathsf{hom}_{I_{\mathcal{M}}}$ are exactly those for which $F(i) \subseteq 1_{\mathsf{cod}(F(i))}$. This means that any inclusion i must be mapped to a function f such that $f(x) = x$, although in general these functions need not be defined on the entirety of $\mathsf{cod}(F(i))$.

Constructs furthermore have useful formal properties. For one thing, all morphisms in a construct $\mathcal{M}, F$ whose underlying functions are injective (i.e. the morphisms f such that $F(f)$ is injective) are monomorphisms. The converse does not hold, unless $\mathcal{M}$ has a so-called *free object* for some singleton set of models (Adámek et al., 2004, p. 144). Roughly, such a free object is a model that contains a single entity, and is included in any other model which contains that entity. In $\mathcal{T}$, the free models for singleton sets are those with singleton domains, where no fundamental relations hold.

In one sense, though, constructs may be *too* structured for certain applications. Consider models that are physical objects. Which sets are these to be identified with? Sets of space-time points? Sets of elementary particles? Sets of their properties? In each of these cases, controversial metaphysical assumptions have been made. In particular, identifying objects with sets means that *numbers* will be applicable directly to things in the world, since they are applicable to sets. This goes against the Fregean observation that numbers require not only an object, but a concept to place objects under (Frege, 1884, §§21–25).

A somewhat less demanding concretisation of a model space can be acquired by letting the functor F take $\mathcal{M}$ not to $\mathcal{V}$, but to some similar structure, such as a mereology — for example one of spacetime regions. A mereology can be defined as a model space $\mathcal{M}$ where $\mathsf{obj}_{\mathcal{M}}$ is a collection of possible (presumably concrete) things with an order relation $\leqslant$ that determines which things are part of which. Since, as is well-known from order theory, any ordered set can be embedded in the set of all subsets of some set in such a way that the order corresponds to set inclusion, one can see this as a part-way stop between the potential ontological vagueness of the bare model space, and the sometimes excessive ontological precision of a functor to $\mathcal{V}$.

Category embeddings, when viewed as morphisms in a higher-order category whose objects are categories themselves, fulfil the diagonalisation requirement that we imposed on embeddings in the last section. But the reduction concept also has interesting applications to entire model spaces. By the characterisation we have given, a reduction of a model space $\mathcal{M}$ to a model space $\mathcal{M}'$ would be a functor $R : \mathcal{M} \to \mathcal{M}'$ which is orthogonal to all monomorphic functors. Since, for the category of categories, monomorphisms are category embeddings (Adámek et al., 2004, p. 252), this means that if R is a reduction functor, any commutative diagram

$$\begin{tikzcd}
\mathcal{M}_1 \arrow[r, "R"] \arrow[d, "F"'] & \mathcal{M}_2 \arrow[d, "G"] \\
\mathcal{M}_3 \arrow[r, "M"'] & \mathcal{M}_4
\end{tikzcd}$$

where M a category embedding, must have a diagonal functor H such that

$$\begin{tikzcd}
\mathcal{M}_1 \arrow[r, "R"] \arrow[d, "F"'] & \mathcal{M}_2 \arrow[d, "G"] \arrow[dl, "H"'] \\
\mathcal{M}_3 \arrow[r, "M"'] & \mathcal{M}_4
\end{tikzcd}$$

commutes. What does this mean? A *congruence* on a category $\mathcal{C}$ is an equivalence relation on the objects together with a partial equivalence relation on morphisms, both of which are compatible with the categorical structure. The following theorem characterises reductions in terms of congruences.

Theorem 3.5 : $R : \mathcal{M}_1 \to \mathcal{M}_2$ is a reduction iff $\mathcal{M}_2$ is isomorphic to some category $\mathcal{M}_1'$ obtained by identifying congruent models in $\mathcal{M}_1$ under some congruence relation, i.e. iff $\ker R$ is a congruence on $\mathcal{M}_1$.

Proof. As we noted, extremalness follows from strongness. The result follows from a theorem of Bednarczyk et al. (2007), where one also can find the exact definition of what a congruence on the category of categories must be like. □

So a reduction of one model space to another is a functor that identifies models, but does not introduce any new ones. It is easily shown that the isomorphism relation $\simeq$ is a congruence on any category, and we may therefore speak of some model space which is the image a reduction R with $\ker R = \simeq$ as a *reduct* of the domain of R. Such a reduction identifies isomorphic models, and no others, and thus only disregards "differences without a difference", as we required for reductions. All reducts of the same model space are isomorphic.

Model spaces (and categories in general) that have a common reduct are called *equivalent.* There is a significant sense in which model spaces which are equivalent have the same structure, even though they may have different numbers of models. Another way to define such equivalence of categories is with two functors $F : \mathcal{C} \to \mathcal{C}'$ and $G : \mathcal{C}' \to \mathcal{C}$, such that applying $G \circ F$ or $F \circ G$ to any object of $\mathcal{C}_1$ or $\mathcal{C}_2$ returns us to an object isomorphic to the one we started with:

$$\begin{tikzcd}
\mathcal{M} \arrow[r, shift left, "F"] & \mathcal{M}' \arrow[l, shift left, "G"]
\end{tikzcd}$$

$$\begin{tikzcd}
\mathfrak{M}_1 \arrow[dr] \arrow[dd, leftrightarrow, bend right=60, "isom."'] & {\mathfrak{M}'}_2 \arrow[dl] \arrow[dd, leftrightarrow, bend left=60, "isom."] \\
\mathfrak{M}_1 \arrow[dr] & {\mathfrak{M}'}_2 \arrow[dl] \\
\mathfrak{M}_1 & {\mathfrak{M}'}_2
\end{tikzcd}$$

Because of this property, category equivalence is often described as *isomorphism up to isomorphism.* It is usually more important than category isomorphism, which takes into account the cardinalities of categories' object classes as well.

3.4 The Diversity of Model Spaces

Using the sketch of a general categorical theory of models of this chapter, we can characterise model spaces in a systematic manner. The next chapter will do so for the model space $\mathcal{N}$ of necessitarian models, which is the one that will be our primary focus in this book, but there are of course others as well, of varying use. This section will be devoted to these, in order to get a taste of how different kinds of model spaces can be described.

3.4.1 *Theory Space Models*

One of the most general forms of model space will be termed $\mathcal{Th}$, or the space of *theory space* models. Let the objects of this space be all theories, in the sense of the last chapter in which a theory A is a closure operator C_A on a set L_A of claims called the theory's language. Let the morphisms be the theory homomorphisms between these theories, by which we as before mean those functions $f : L_A \to L_B$ for which

$$p \in C_A(X) \Rightarrow f(p) \in C_B(f[X])$$

holds, for all $p \in L_A$ and $X \subseteq L_A$. It is quickly checked that the monomorphisms are exactly the injective homomorphisms. The following theorems characterise embeddings and reductions.

Lemma 3.6 : The epimorphisms in $\mathcal{Th}$ are the surjective homomorphisms.

Proof. Assume that $e : A \to B$ is an epimorphism, and let B' be a theory such that $L_{B'} = L_B \cup \{*\}$, $C_{B'}(X) = C_B(X)$ for all $X \subseteq L_B$ and $C_{B'}(X) = L_{B'}$ whenever $* \in X$. Let $f, g : B \to B'$ be morphisms such that $f(p) = *$ if $p \in e[L_A]$ and $f(p) = p$ otherwise, and $g(p) = *$. We must then have that $f \circ e = g \circ e$, and since e is epic, it follows that $f = g$. But this requires the image of e to be all of L_B. □

Theorem 3.7 : The embeddings in $\mathcal{Th}$ are the injective functions $m : A \to B$ such that $p \in C_A(X)$ iff $f(p) \in C_B(f[X])$.

Proof. Let m be a strong monic from A to B. Since it is a monomorphism, it is injective. Let B^{sub} be the subtheory of B onto which m maps A, let $m' : A \to B^{sub}$ be the function such that $m'(p) = m(p)$ for all $p \in L_A$, and let $i : B^{sub} \to B$ be the inclusion of B^{sub} into B. Then there is a morphism $h : B^{sub} \to A$ such that $h \circ m' = 1_A$ and $m' \circ h = 1_B^{sub}$, and thus m is an isomorphism onto a subtheory of B.

For the converse, assume that $m : A \to B$ fulfils the condition that $p \in C_A(X)$ iff $f(p) \in C_B(f[X])$, that $f : A' \to A$, $g : B' \to B$ are morphisms, and that $e : A' \to B'$ is an epimorphism such that $m \circ f = g \circ e$. Factor m as $m = i \circ m'$, where i is an inclusion and m' is epic. Then h can be defined as $h = m'^{-1} \circ g$, and this is well defined since the image of g must be the same as that of m. □

Theorem 3.8 : The reductions in $\mathcal{Th}$ are the surjective functions $e : A \to B$ such that $p \in C_A(X)$ iff $e(p) \in C_B(e[X])$.

Proof. Since a reduction is an epic, it is, by the preceding lemma, a surjection. To show that $e(p) \in C_B(e[X])$ if $p \in C_A(X)$, write e as $e = m \circ e'$, where $e' : A \to A^{sub}$, $i : A^{sub} \to B$, and A^{sub} is a subtheory of A, such that m is an injection. Then we can use the strongness condition to prove the existence of $h : B \to A$ such that $e \circ h = 1_B$. Since h, as a homomorphism, has to preserve consequence, we have that the reduction condition holds.

For the other direction, assume that $e : A \to B$ is a reduction. We need to show that whenever there is an injective morphism $m : A' \to B'$, and morphisms $f : A \to A'$ and $g : B \to B'$ such that these all commute, there is a unique $h : B \to A'$ such that $f = h \circ e$ and $g = m \circ h$. Let e^{inv} be a function from B to A, such that $e \circ e^{inv} = 1_{\mathfrak{M}_2}$ (such a morphism exists because e is a reduction). We can then let $h = f \circ e^{inv}$. □

This means that embeddings and reductions among theories work as we expect them to: an embedding places a theory exactly as it is inside another, and a reduction maps only A-equivalent claims to the same claims. Given any theory A, we let $\mathcal{Th}_A$ be the subcategory of $\mathcal{Th}$ that contains the theories *in* A (i.e. the strengthenings of A).

Theory space models are cheap: whenever we have a theory A, we have that theory's theory space, and thus also the model space $\mathcal{Th}_A$ of its theory space models. We can then use these models to give semantics for arbitrary theories, as we will show in the next chapter. The downside to them is that they do not provide a very useful notion of "possible world": traditionally, so-called *ersatz* possible worlds are assumed to be *maximal* consistent sets of sentences. This, however, works only for theories that have a logical structure close enough to a Boolean lattice, such as classical logic. For intuitionistic theories, a "possible world" (i.e. a possible state of mathematics) does not have to contain either a sentence or its negation, since it could be the case that proofs exist neither for p nor for $\neg p$.

3.4.2 *Matrix Models*

Slightly more structure than that needed by $\mathcal{Th}$ is required by the so-called *matrix models*, first investigated by Łukasiewicz, but made popular primarily through the works of Lindenbaum and Tarski. Let a *matrix model* be a pair $\mathfrak{M} = \langle \mathfrak{A}, D \rangle$, where $\mathfrak{A}$ is an algebra with a carrier set A of claims (see section 2.4) and D is a subset of A called the *designated values*. We refer to the space of all such models as $\mathcal{Mt}$. D is commonly called the *truth predicate*, since it is interpreted as the set of claims in A that are *true* in $\mathfrak{M}$.

It is fairly easy to define morphisms on the space $\mathcal{Mt}$: given two such models $\mathfrak{M}_1 = \langle \mathfrak{A}_1, D_1 \rangle$ and $\mathfrak{M}_2 = \langle \mathfrak{A}_2, D_2 \rangle$, a morphism h from $\mathfrak{M}_1$ to

$\mathfrak{M}_2$ is a homomorphism from $\mathfrak{A}_1$ to $\mathfrak{A}_2$, such that $h[D_1] \subseteq D_2$. The second condition guarantees that morphisms preserve which claims in A are taken to be true.

Whenever we can formalise a theory, in the sense of section 2.4, we can interpret the theory's consequence operator as having been specified through selecting a specified subset of $\mathcal{Mt}$ as a set of possible worlds or states of affairs. How to do this is discussed in the chapter on semantics.

3.4.3 *Coherence Models*

Both theoretical and matrix models are built from the same stuff the theories themselves are built from. Another way to build a model space from claims is employed in constructing a space of *coherence models* — henceforward $\mathcal{Ch}$. Let us define such a model as a pair $\mathfrak{M} = \langle B, K \rangle$, where B is a set of *potential beliefs* (whatever we take these to be), and K is a function from subsets of B to an ordered set D — the *degrees of coherence.* Such *coherence measures* have been much discussed lately in epistemology, beginning with the introduction of a simple probabilistic real-valued measure of coherence by Tomogi Shogenji (1999). Since then, most epistemologists seem to have taken D to be the real line (cf. Olsson, 2005), but there are also those who assume only the structure of a partial order (Bovens and Hartmann, 2003). Since the latter interpretation is compatible with the reasonable possibility that degree of coherence is vague, possessing only something like a stable core, and also invites interesting philosophical problems, that is the one we have used here.

An interpretation h of a theory A in a coherence model $\mathfrak{M} = \langle B, K \rangle$ can be taken to be a function from L_A to B. For any claim p in L_A, we say that p is *true* iff $h(p) \in X$, for some $X \subseteq B$ such that $K(X) > K(Y)$ for all subsets Y of B not logically equivalent to X. In other words, p is true iff p expresses a belief that is a member of the uniquely most coherent set of beliefs in B (where this uniqueness is assumed to hold up to logical equivalence). Without further assumptions on K, there is no guarantee that there are any beliefs that correspond to true sentences.

This situation can arise in two different ways:

- (*i*) B lacks a set of beliefs that is at least as coherent as the others, or, in the order theorist's terms, $K[B]$ lacks a top. In such a case, there are several sets of beliefs for which no set having a higher coherence can be found, but there is still no fact of the matter as to which of these is more coherent than the other. This can happen if D is only partially ordered, but it can also happen if B contains an infinite chain of sets of beliefs of greater and greater coherence. Thus assuming D to be the real line is not sufficient in order to exclude this possibility, and we also need conditions on B and C.

- (*ii*) $K[B]$ has a top, but there are several nonequivalent sets of beliefs that are mapped to this value by C, i.e. that are maximally coherent.

One way to exclude (*i*) is to take D to be the real line (or at least some linearly ordered set), and B to be finite, presumably since infinite sets of beliefs are not potential sets of beliefs one could *have.* (*ii*) is harder to avoid: it seems that we need substantial, perhaps empirical assumptions to impose on B, such as one that lets the empirical data we have (i.e. a subset of B) uniquely determine C. But in such a case, one might ask what role coherence plays, since the empirical data after all determines what is true or false.

Another possibility seems to be to drop the requirement that the set of true beliefs has to be determined up to logical equivalence. The problem with this is that it can allow both a claim and its negation to be true at the same time, so long as they belong to different maximally coherent sets. This, in turn, could be held to conflict with the meaning of "true", or "negation" but how such an argument should proceed is not completely clear. It is still something that a subjective idealist, or perhaps a dialethist, might want to argue for. Coherence models *are* idealistic in spirit, since they interpret the world as something made up from our own beliefs.

Finally, we can just accept that whether truths exist is dependent on what the world (i.e. the set of potential beliefs) is like. This way, we treat the question of whether there is any uniquely maximally coherent set of beliefs as an empirical matter. Truth or non-truth can be regained for a subset of our claims by redefining p to be true iff $h(p)$ is equivalent to some member of *all* sets of maximally coherent beliefs. Using this definition, we still have the empirical possibility of truth-value gaps, but we have at least excluded the possibility of gluts (i.e. claims that are both true and false), so long as we make sure that sets of contradictory beliefs can never have maximal coherence.

What should we take as the morphisms of $\mathcal{C}$? A reasonable interpretation of the concept is to let a morphism from $\mathfrak{M}_1 = \langle B_1, K_1 \rangle$ to $\mathfrak{M}_2 = \langle B_2, K_2 \rangle$ be a function $f : B_1 \to B_2$ such that $K_1(X) \leqslant K_1(Y) \Rightarrow K_2(f[X]) \leqslant K_2(f[Y])$, for all $X, Y \subseteq B_1$. This choice makes isomorphisms come out as expected, although we could, of course, also have made other choices.

3.4.4 *Concrete Models*

So far, the model spaces we have discussed have all been abstract mathematical structures. For a much more concrete example, and to show that model theory does not have to be a purely mathematical game, we may define a model space $\mathcal{L}$ of *Lego models*, such that $\mathsf{obj}_{\mathcal{L}}$ is the class of everything that can be built with nothing but an endless supply of a given type of brick; for simplicity we can take these to be the "standard" 2 × 4 bricks (fig. 3.2), in various colours.

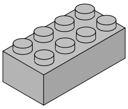


Figure 3.2: ***Building block of $\mathcal{L}$.***

Models of $\mathcal{L}$ are as concrete as one could possibly wish for: you can actually touch them![9] But they can still be taken to form a category,

[9] A perhaps amusing observation is that they, by the category theorists' terms,

if we decide on an interpretation for the morphism notion. It might be easiest to start from the isomorphisms here, which means that we must decide what is to count as a model's own structure, and what is to count as circumstantial. It is natural to include rigid translations of unattached parts among the structure-preserving operations, which means that a model $\mathfrak{M}_1$ is isomorphic to a model $\mathfrak{M}_2$ if $\mathfrak{M}_2$ can be obtained from $\mathfrak{M}_1$ by moving around the parts of $\mathfrak{M}_1$ without attaching or detaching any blocks.

But there is a second class of transformations that we should include as well. Let a *replacement* be the act (or operation, in the concrete sense of the word) of replacing one or more blocks by other blocks of the same colour, oriented the same way. Letting the isomorphisms include replacements means that we do not take the specific identity of a block as part of the structure, and this seems very reasonable. Combining the two forms of transformation we have mentioned, we therefore require the isomorphisms in $\mathcal{L}$ to be those operations that can be performed by composition of rigid translations and replacements. We can see that this definition also satisfies the category-theoretic definition of isomorphisms: for every isomorphism, there is an inverse transformation (also an isomorphism), such that composing these gets us back to the same model we started with.

Moving on to the embeddings, there is one natural way to define these, given the notion of isomorphism: we require an embedding $f : \mathfrak{M}_1 \to \mathfrak{M}_2$ be an isomorphism of $\mathfrak{M}_1$ to a *part* of $\mathfrak{M}_2$ (i.e. to a model that is a subcollection of the blocks of $\mathfrak{M}_2$). Fig. 3.3 below demonstrates two embeddings f and g of one $\mathcal{L}$-model in another. It also illustrates the importance of distinguishing between specific embeddings, and not collapsing them into a mere parthood relation: f and g are different *ways* that $\mathfrak{M}_1$ can be a part of $\mathfrak{M}_2$.

The step to characterisation of arbitrary morphisms can be taken through the observation that for any embedding f, the block $f(a)$ is

make up an *abstract* category. A concrete category, as we explained in the last section, is a category that has a faithful functor to some "underlying" category, or usually just a category whose objects are *sets*. So according to the category theory, sets are concrete, and physical things are abstract. I think this is an excellent example of the sense in which, as Russell put it, "logic is so very backwards as a science" (Russell, 1985, p. 59).

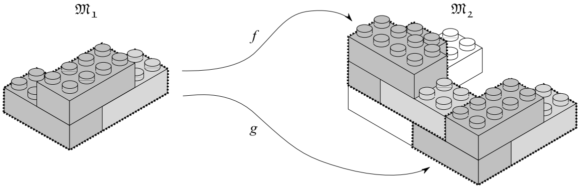


Figure 3.3: ***Two embeddings of*** $\mathfrak{M}_1$ ***in*** $\mathfrak{M}_2$.

attached in way α to block $f(b)$ iff a is attached in way α to block b.[10] This is similar to embeddings for MMT models: relations must hold in the image of an MMT embedding iff they hold in the preimage. But just as a homomorphism in that model space is "half" of an embedding (i.e. it guarantees that relations that hold in the preimage must hold in the image, but lacks the "only if" part), we can define a morphism in $\mathcal{L}$ as half a monic. The result can be summarised as follows:

> $f : \mathfrak{M}_1 \to \mathfrak{M}_2$ is a *morphism in* $\mathcal{L}$ iff f is a combination of (i) rigid translations of parts, (ii) replacements, and (iii) attachment of unattached parts, that results in some part of $\mathfrak{M}_2$, given $\mathfrak{M}_1$.

According to this definition, the operation of assembling a Lego model is therefore a morphism, but the morphisms are a wider class than this, since they include the replacements as well.

3·4·5 *Physical Models*

Another class of models, seemingly straddling the divide between the concrete and the abstract, are the physical models, by which we really mean the typical models of physical theories. Currently the most well-accepted of these theories is QFT (quantum field theory), for which a

[10] We leave it as an exercise for the reader to prove that there are exactly 46 exclusive ways to connect two 2 × 4 blocks — 25 with the blocks parallel, and 21 with them perpendicular.

model (a "$\mathcal{Qf}$ model") is a collection of operator-valued fields—one for each fundamental particle in the Standard Model. If QFT was *true*, and not only our best theory, then that would be what the actual world is.[11]

Now, operator-valued fields come with their own notions of morphism (or continuous maps, as they are called there), embedding and isomorphism, and we could of course just adopt these. This would however go against the sense in which the models in $\mathcal{Qf}$ are *physical*, and not only mathematical. It is common in a physical model to separate the variables that have a physical interpretation (the measurables) from those that do not (the "artifacts" of the model). This separation corresponds directly to a separation of mappings of physical models into those that the physical theory must be invariant under, and those that it need not be invariant under. For theories obeying special relativity (such as QFT), we find the Lorentz transformations in the first group, and for theories that include quantum mechanics (again, like QFT), it includes phase-shifts of the wave function.

For models intended for a physical theory, it is therefore reasonable to take the isomorphisms to be the mappings that the theory is invariant under. The *metaphysical* claim that only the invariant parts of the theory are real (for instance, that only spatiotemporal relations are real, and not absolute positions or times), then translates to the claim that the model space in question is skeletal.

Assuming isomorphisms in physical model spaces to express identity of all observables, we come to the question of embeddings. Here our previous method of interpreting these as isomorphisms to *parts* of models fails us. All fields fill out the whole of space-time, so they do not have fields as parts, in the geometrical sense of the term.

This does not have to be a disaster. Perhaps there just is no useful notion of embedding between $\mathcal{Qf}$ models that differs from the isomorphism. More likely, however, we just have to look at the field from another angle. A quantised field can also be seen as a superposition of *states* $\{f_i\}$, each a function of the space-time coordinates x, y, z, t, together with an assignment of non-negative integers to each state: the

[11] Of course, the world may be a collection of fields even if QFT is wrong, or at least incomplete, as seems to be the case.

number of particles of the type the field describes that are in that state. This construction is what is commonly referred to as "Fock space" after its inventor (Dirac, 1958, p. 139).

Using Fock space, we could potentially define an embedding as a function that may add particles in each such state, but may not remove any. By fiddling a little, we could get this to reduce to the case of isomorphism when it neither removes nor adds a particle in any state. The question of how general morphisms are to be determined is however still open. As the complexities of QFT are too great for us to be able to say anything well-motivated about its models in this book, we will not attempt to answer it here.

3.5 Models and Theories

From the examples of model spaces in the last section, we may draw some general conclusions. In all the spaces discussed to far, it is obvious that we had to make *choices* when we defined what was to count as a structure-preserving mapping: for $\mathcal{L}$, for instance, we chose to let the colour of the bricks count as part of a model's structure, but not their specific identities. For $\mathcal{Q}f$, we chose to regard models that differ only in their non-physical quantities as isomorphic. Both these cases should make it obvious that the structure, at least partly, is something we *lay down* on the models, in order to be able to grasp and categorise them more efficiently. Pragmatism enters in creating model spaces from collections of concrete objects, since they do not really come with a predefined structure.

The case is somewhat different for more abstract model spaces such as $\mathcal{M}t$, whose objects already are mathematical structures.[12] Pragma-

[12] I use the words "abstract" and "concrete" here without attempting to give any kind of definition. My intention is not to capture anything like their "common meaning", if there is such a thing — I just need a pair of words for distinguishing between objects whose structure are given with them, and objects where this is not

tism enters at an earlier stage here: when we choose what model space to represent some phenomena with. This problem could be seen as a metaphysical analogue to Carnap's principle of free choice of language for theories (Carnap, 1937). Just as, for Carnap, the selection of a language is a pragmatic affair, the selection of what model space to use must be guided by what kind of understanding we are after.

But Carnap's view was that language choice is not subject to questions about truth at all. In contrast to this, there is no significant difference between a theory and a language for us, since the adoption of a language involves a commitment to the inferences allowed in that language being truth-preserving. To see that this is applicable to model spaces as well, we may note that in a certain sense, a model space *is* a language, or more generally, a theory. Just as $\mathcal{Th}$ correlates every theory with a model space, every model space can be correlated with a theory. Where $\mathcal{M}$ is a model space, let $\mathcal{M}$'s *canonical theory* $Th(\mathcal{M})$ be the pair $\langle L_{Th(\mathcal{M})}, C_{Th(\mathcal{M})}\rangle$, where

$$L_{Th(\mathcal{M})} = \wp(\mathsf{obj}_{\mathcal{M}})$$

and

$$C_{Th(\mathcal{M})}(X) = \left\{p \in L_{Th(\mathcal{M})} \;\middle|\; \bigcap X \subseteq p\right\}$$

for all $X \subseteq L_{Th(\mathcal{M})}$

The motivation is this: in the canonical theory, each claim is a set p of models, which can be interpreted as the claim that the actual model $\mathfrak{A}$ is one of these. By holding all claims in a set X to be true, we say that the actual model is in all of the sets in X, or equivalently, that is lies in their intersection. The consequences of a set X of such claims are then the sets of models that contain that intersection.

Canonical theories are complete in the sense that every theory *in* them corresponds to a unique claim: if A is a canonical theory and B is a theory in A, then the intersection p of all claims in $\top_B$ is also a claim in A, and $C_A(\{p\}) = C_A(\top_B)$, so p and B are equivalent according to A.

The canonical theories are however only some of those that can be constructed from a model space $\mathcal{M}$. Let us call $A = \langle L_A, C_A\rangle$ a

the case.

subcanonical theory of $\mathcal{M}$ iff L_A is a subset of $\wp(\mathsf{obj}(\mathcal{M}))$ and $C_A = \{p \in L_A \mid \bigcap X \subseteq p\}$, just as in the canonical theory. It is easy to see that a subcanonical theory of $\mathcal{M}$ is always a subtheory of $Th(\mathcal{M})$. The difference is that all sets of models no longer correspond to claims expressible using the theory.

The subcanonical theories comprise a fairly large class. So large, indeed, that any theory is equivalent to a subcanonical theory of some model space.

Theorem 3.9 : For every theory A, there is a model space $\mathcal{M}$ and a subcanonical theory B of $\mathcal{M}$ such that A is isomorphic to B.

Proof. Assume that A is a theory, and let $\mathcal{M}$ be the space of theory space models of A. Let B be a theory $\langle L_B, C_B\rangle$ such that $L_B = \wp(\mathsf{obj}_{\mathcal{M}})$ and $\alpha \in C_B(\Gamma)$ iff $\bigcap \Gamma \subseteq \alpha$.[13] Then B is the canonical theory of $\mathcal{M}$, as defined above.

Now define a function $\varphi : L_A \to L_B$ such that

$$\varphi(p) = \{T \in \mathcal{T}_A \mid p \in \top_T\}$$

As in ch. 2, $\mathcal{T}_A$ is the theory space of the theory A. Let the theory B' be the subtheory of B whose language is the image of L_A under φ. If we then show that $X \vdash_A p$ iff $\varphi[X] \vdash_B \varphi(p)$, for any $X \subseteq L_A$ and $p \in L_A$, this means that B' is isomorphic to A, since it is easy to show that φ is injective, and it is by definition surjective. Furthermore, B' is a subcanonical theory of $\mathcal{M}$.

First we show that $X \vdash_A p$ implies $\varphi[X] \vdash_B \varphi(p)$. Assume that $X \subseteq L_A$, and $p \in C_A(X)$. What we need to show is that

$$(\forall T \in \mathcal{T}_A)(T \in \bigcap \varphi[X] \to T \in \varphi(p))$$

which is equivalent to

$$(\forall T \in \mathcal{T}_A)((\forall q \in X)(T \in \varphi(q)) \to T \in \varphi(p))$$

[13] We use small Greek letters for the claims in B here, and capital Greek letters for sets of such claims, in order to better separate the claims of theory B from those of theory A.

Assume an arbitrary theory $T \in \mathcal{T}_A$ such that $X \subseteq \top_T$. Then it follows by our assumption that $p \in \top_T$. But since $\varphi(p)$ contains all theories in A that contain p, we must have that $T \in \varphi(p)$ as well.

For the other direction, assume that

$$(\forall T \in \mathcal{T}_A)((\forall q \in X) T \in \varphi(q)) \to T \in \varphi(p))$$

Let T be the theory in A such that $\top_T = C(X)$. Then $X \subseteq \top_T$, and so by our assumption $T \in \varphi(p)$. This, however, holds iff $p \in \top_T$. □

Since the adoption of a model space is equivalent to the use of a theory, it is, unlike the Carnapian languages, not entirely immune to questions of truth or falsity. As we mentioned in the preceding chapter, every theory *when seen from its own viewpoint* is of course true, but when we embed theories or model spaces into others, it may very well be that the embedded theory A sees as true claims that are not true according to the theory B it is embedded in. In such a case, A may be false *according to B*.

Is there some model space in which *all* of the world's structure is representable, and which thus contains all others and can be used to settle questions such as these once and for all? Lacking a universal definition of "structure", it is unfortunately hard to see what this could mean. If there is such a thing as a well-defined category of all worlds, whose morphisms are the "true" structure-preserving mappings, then this category is such a model space, but this is just a reframing of the initial problem, rather than an actual answer. Furthermore, the considerations in sct. 2.5 tell strongly against the coherence of such a concept.

The general relationship between model and reality can be described as follows: the world is a thing, possibly with some kind of structure. Parts of this structure can be described by subsuming the world under a model space, i.e. by taking it to be an object in such a space. But these spaces, to be informative, must have the structures of their models *independently* specifiable — a model space such as the one of the preceding paragraph is impossible to work with for us. Therefore, we have no guarantee that we can have useful model spaces that are rich enough to capture all of the world's structure, and so it is pertinent for us to be able to work with different spaces, for different interests. A summary of some of the more important of the model spaces employed in this book is given in table 3.1.

Model space	***Type of models***	***Described in section***
$\mathscr{T}$	Tarskian models	3.1
$\mathscr{V}$	Thin models	3.1
$\mathscr{Th}$	Theory Space models	3.3
$\mathscr{Mt}$	Matrix models	3.3
$\mathscr{Ch}$	Coherence models	3.3
$\mathscr{Qf}$	Quantum Field theoretical models	3.3
$\mathscr{N}$	Necessitarian models	4.1
$\mathscr{PN}$	Probabilistically Necessitarian models	4.4

Table 3.1: ***Model spaces of this book.***

The model space that we will concern ourselves with the most is the space $\mathscr{N}$, described in the next chapter. I believe that this space is especially useful for metaphysics since, as we shall see in chapters 6 and 7, it allows very strong relations between a theory and the structure of its metaphysics to be derived. It also holds some interest as a space in which many traditional metaphysical concepts can be represented, and so may function as a bridge between traditional metaphysics and the model-theoretic version of it that I advocate.

Chapter 4
Necessitarian Metaphysics

Here we introduce *necessitarian metaphysics*, which will take a center stage in the later parts of this book. A necessitarian metaphysic is a kind of model space in which models are sets of entities with relations of *necessitation* defined among them. The structure of this necessitation is roughly that of a multiple-conclusion consequence relation, so it is by nature nondeterministic. We discuss axioms for these relations, and prove a characterisation theorem that shows that we also can view a necessitarian metaphysic as a selection of *possible worlds*.

Section 2 deals with the category-theoretic aspects of necessitarian metaphysics. We show that embeddings and reductions work as we expect them to, and we also discuss the question of identity vis-a-vis necessary coexistence. Section 3 attempts to tie this discussion to more traditional metaphysical concerns. In particular, we show how to express several metaphysical concepts such as parthood and causality in terms of necessitation relations, and also how to work with objects that have more structure than the primitive *entities* that we have based necessitarian metaphysics on.

In the final section, we introduce a modification of the necessitation relation which is crucial for capturing *probabilistic* necessitation. Axioms are given, and we prove that probabilistic necessitation relations are interpretable in terms of standard Kolmogorovian probability theory.

4.1 Necessitation Relations and Possible Worlds

The idea of "necessary connection" may be as old as philosophy itself, and although one could fairly say that not very much progress has been made in clarifying what is means for one thing to necessitate another, both scientists and philosophers often have use for a distinction between those relations that we say hold necessarily, and those that we say hold only contingently. For example, many sciences make a difference between laws and accidental generalisations, or causal relationships and relationships of mere statistical correlation.

We do not have to stipulate anything transcendental or non-empirical about this concept: purported causal laws can be shown to be spurious correlations, for instance, by exhibiting circumstances in which the causal effect is screened off (Pearl, 2009, ch. 2). Claimed universal generalisations can be shown false by finding counterexamples. We do not mean to exclude any of these concepts of necessitation, although we do not want to limit ourselves to them at the outset either. Necessitation may involve something as simple as a statistical relationship, or something as "deep" as a higher-order relation between universals, depending on which metaphysics we have.

The simplest form of necessitation is the singular, deterministic one, which lets us say that an entity a necessitates another entity b when it is impossible that a should exist without b, or, in Leibnizian terms, that b exists in all possible worlds where a exists. We write this relationship as

$$a \mathrel{\circ\!\!\rightarrow} b$$

Some properties of $\mathrel{\circ\!\!\rightarrow}$ follow easily from the Leibnizian characterisation. For one thing, it must be a preorder, i.e. a reflexive, transitive binary relation. Unfortunately, there are some cases of necessitation that we may want to be able to model that cannot be expressed using only $\circ\!\!\rightarrow$. An example is joint causality, where c_1 and c_2 are sufficient for the effect e *together*, but not individually.

It could seem, at first, that these limitations can be avoided by accepting a sufficiently strong mereology, and saying that $\circ\!\!\rightarrow$ holds between the mereological sum or fusion $c_1 + c_2$ of c_1 and c_2, and e. But this only pushes the problem further back. What is it that makes the existence of c_1 and c_2 entail that of $c_1 + c_2$? What we *want* to say is that if c_1 and c_2 both exist, then e exists, and to do this involves several things jointly necessitating another at some point.

There is thus another relation, which we will term *joint deterministic necessitation*, and write as

$$X \rightarrowtail b$$

where X is a set of entities and b is a single entity.[1] Taking the necessitation of b by X to be the condition that b exists in all worlds in which all entities in X exist, $\rightarrowtail$ can be seen to satisfy the following axioms, analogous to those that a logical consequence relation satisfies:

$$\begin{array}{rl} (\textit{Reflexivity}) & \text{if } b \in X \text{ then } X \rightarrowtail b \\ (\textit{Monotonicity}) & \text{if } X \rightarrowtail b \text{ and } X \subseteq Y\text{, then } Y \rightarrowtail b \\ (\textit{Cut}) & \text{if } X \rightarrowtail a \text{ and } X \cup \{a\} \rightarrowtail b\text{, then } X \rightarrowtail b \end{array}$$

All of these follow directly from the definition. But we still cannot quite capture all the things we might want to call cases of necessitation. Many kinds of causation, for example, are often taken to be non-deterministic. For this kind of generality, we need a relation of *joint nondeterministic necessitation*, or as we frequently will call it, just *necessitation.* We write

[1] Using set-theoretic terminology here is a notational convenience: the relation holds between the entities *in* X and b, and not between the set X itself and b – that would reinstate the same problems that occur with defining joint necessitation to hold between sums of entities, which we discussed in the preceding paragraph.

$$X \mathrel{>\!\!\in} Y$$

if every possible world in which all the entities that are in the set X exist also contains some entity that is in the set Y, and we read this as the statement that *the X's necessitate some Y*.

With this relation we can also express what it means for a collection of entities to make up a possible world. The intuitive idea is that what is required for a set X of entities to be the contents of a possible world is for these entities to require nothing except themselves for their existence, so that it is possible for the elements of X to exist, and nothing else. This property can be expressed as the condition that X makes up a possible world iff, for any subset $Y \subseteq X$, and any set Z of entities whatsoever, $Y \mathrel{>\!\!\in} Z$ implies that $Z \cap X \neq \varnothing$. A more succinct characterisation is given by the equivalent condition $X \not\mathrel{>\!\!\in} X^C$, where X^C is the set of all possible entities *not* in X.

It is clear that everything that can be expressed using a singular or deterministic necessitation relation can be expressed using an indeterministic one as well, so indeterministic necessitation relations are well suited to play the part of primitives for us. Let a *necessitarian metaphysic* be a pair $\mathcal{M} = \langle E, \mathrel{>\!\!\in} \rangle$, where E is a set that we call the *set of possible entities*, and $\mathrel{>\!\!\in}$ is a nondeterministic necessitation relation on E. Just as with the deterministic necessitation relation, we can give axioms for $\mathrel{>\!\!\in}$.

(*Overlap*) if $X \cap Y \neq \varnothing$ then $X \mathrel{>\!\!\in} Y$

(*Dilution*) if $X \mathrel{>\!\!\in} Y$, $X \subseteq X'$ and $Y \subseteq Y'$, then $X' \mathrel{>\!\!\in} Y'$

(*Set cut*) if $X \cup Y \mathrel{>\!\!\in} Y^C \cup Z$ for all $Y \subseteq E$, then $X \mathrel{>\!\!\in} Z$

(*Non-triviality*) $\varnothing \not\mathrel{>\!\!\in} \varnothing$

These all follow from the intended interpretation: Overlap is motivated by noting that the overlap of two sets is sufficient for the existence of everything in one of them to guarantee the existence of something in the other, and Dilution holds because if all of X' exists, then everything in every subset of X' exists as well, and if something in Y exists, then it must exist in every set containing Y. Set cut can be motivated as follows: assume that *not* every world that contains all of X also contains something in Z. Then, there must be some world ω for which this is true, and since ω is a world, we have $\omega \not\succ\!\!\in \omega^C$. But because $X \subseteq \omega$, and $Z \cap \omega = \varnothing$, we must also have that $X \cup \omega = \omega$, and $Z \cup \omega^C = \omega^C$. It follows that there is some set Y (namely, ω) for which $X \cup Y \not\succ\!\!\in Y^C \cup Z$. Non-triviality simply ensures that we do not have $X \succ\!\!\in Y$ for all $X, Y \subseteq E$, and is an assumption made to make our proofs easier.

Let a *partition* of a set Z be a pair of sets $\langle Z_1, Z_2 \rangle$ such that $Z_1 \cap Z_2 = \varnothing$ and $Z_1 \cup Z_2 = Z$. As before, let a *world* be a set ω of entities such that $\omega \not\succ\!\!\in \omega^C$. Set cut can then also be written in the forms

(*Set cut**) if $X \cup Y_1 \succ\!\!\in Y_2 \cup Z$, for all partitions $\langle Y_1, Y_2 \rangle$ of Y, then $X \succ\!\!\in Z$

(*World cut*) if $X \cup \omega \succ\!\!\in \omega^C \cup Z$, for all worlds ω, then $X \succ\!\!\in Z$

(*World cut**) if $\omega \succ\!\!\in \omega^C \cup Z$, for all worlds ω that contain X, then $X \succ\!\!\in Z$

Theorem 4.1 : Given Dilution, Set cut is equivalent to Set cut* and World Cut. Given Dilution and Overlap, it is equivalent to World cut*.

Proof. For the left-to-right direction of Set cut*, assume that Set cut holds, and that $X \not\succ\!\!\in Z$. Then, by Set cut, there is a set Y such that $X \cup Y \not\succ\!\!\in Y^C \cup Z$. Let Y' be any set, and Y_1', Y_2' the partition of Y' such that $Y_1' = Y' \cap Y$ and $Y_2' = Y' \cap Y^C$ By dilution, we must have that $X \cup Y_2' \not\succ\!\!\in Y_1' \cup Z$ as well. The other direction follows trivially by taking $Y = E$.

For world cut, it is only the left-to-right direction that needs proof, since if $X \cup Y \succ\!\!\in Y^C \cup Z$ for all Y, it naturally holds for those Y that are worlds as well. But assume that Y is *not* a world, i.e. that $Y \succ\!\!\in Y^C$. Then, by dilution, the same must hold for $X \cup Y$ and $Y^C \cup Z$ as well, for any X, Z.

Just as the left-hand side of World cut follows from that of Set cut, the left-hand side of World cut* follows from that of World cut. Thus we only need to prove that if ω is a world, and $X \nsubseteq \omega$, $X \cup \omega \mathrel{>\!\!\!\in} \omega^C \cup Z$. But if $X \nsubseteq \omega$, X must overlap ω^C, so the necessitation follows by the Overlap axiom. □

These rules are sometimes easier to apply than the standard version of Set cut. When the necessitation relation is compact (i.e. when $X \mathrel{>\!\!\!\in} Y$ iff $X' \mathrel{>\!\!\!\in} Y'$ for some finite subsets $X' \subseteq X$ and $Y' \subseteq Y$, Set cut is also equivalent to the much simpler axiom[2]

(*Entity cut*) if $X \cup \{e\} \mathrel{>\!\!\!\in} Z$ and $X \mathrel{>\!\!\!\in} \{e\} \cup Z$ for some entity e, then $X \mathrel{>\!\!\!\in} Z$

Theorem 4.2 : Given Overlap and Dilution, Entity cut is equivalent to Set cut for compact necessitation relations.

Proof. Assume Set cut to hold, and that $X \mathrel{\not>\!\!\!\in} Z$. Then there is a set Y such that $X \cup Y \mathrel{\not>\!\!\!\in} Y^C \cup Z$. Because of Dilution, it follows that for all sets $Y_1' \subseteq Y$ and $Y_2' \subseteq Y^C$, $X \cup Y_1' \mathrel{\not>\!\!\!\in} Y_2' \cup Z$ as well, and Entity cut follows by taking $\{e\} = Y_1' \cup Y_2'$.

In the other direction, assume Entity cut, and again assume $X \mathrel{\not>\!\!\!\in} Z$. From Overlap, we have that $X \cap Y = \varnothing$. For any two partitions $\langle Y_1, Y_2 \rangle$ and $\langle Y_2', Y_2' \rangle$ of the sets Y and Y', let $Y \preccurlyeq Y'$ iff $Y_1 \subseteq Y_1'$ and $Y_2 \subseteq Y_2'$. Call a partition $\langle Y_1, Y_2 \rangle$ such that $X \subseteq Y_1$, $Z \subseteq Y_2$ and $Y_1 \mathrel{\not>\!\!\!\in} Y_2$ an *extension* of X, Z. For an arbitrary increasing sequence $\sigma = \langle Y_1^i, Y_2^i \rangle_{i=1}^{\infty}$ of extensions of X, Z, let the *limit* of such a sequence be the partition

$$\lim \sigma = \langle \bigcup_i Y_1^i, \bigcup_i Y_2^i \rangle$$

By compactness, if $Y_1 \mathrel{>\!\!\!\in} Y_2$, then there are finite sets $Y_1^{fin} \subseteq Y_1$ and $Y_2^{fin} \subseteq Y_2$ such that $Y_1^{fin} \mathrel{>\!\!\!\in} Y_2^{fin}$. Letting σ be a finite series of increasing extensions of X, Z, it is obvious that $\lim \sigma$ must be such an

[2] A compact nondeterministic necessitation relation is also known as a *Scott consequence relation*, see Scott, 1971.

extension as well, and compactness allows us to extend this to infinite series. Thus every series of extensions has an upper bound, and it follows by Zorn's lemma that the set of *all* extensions of X, Z must have a maximal element. Let $\langle W_1, W_2 \rangle$ be such an element.

Now let $W = W_1 \cup W_2$. We show that $W = E$. Assume that e is an entity that is *not* in W. Then we must have that $W_1 \cup \{e\} \mathrel{\succ\!\!\in} W_2$ and $W_1 \mathrel{\succ\!\!\in} \{e\} \cup W_2$, since otherwise e would be in either W_1 or W_2. But then it follows, by Entity cut, that $W_1 \mathrel{\succ\!\!\in} W_2$, contrary to assumption, so $W = E$. □

Entity cut is easy to motivate: if some Y exists in every world in which both the X's and e exist, and either e or some Y exist in every world in which the X's exist, some Y must be in every world where the X's are, for either the X's necessitate some Y, or they necessitate e, which together with X necessitate some Y. It would thus be nice if we could limit ourselves to compact necessitation relations. Unfortunately, this is not possible. Take *mereological* necessitation as an example. Given a set of entities, a metaphysics may postulate the existence of a whole that has these as parts. But consider space-time, as made up from points: no finite set of space-time points makes up a volume of space-time, but we may very well want to allow that any such volume consists of points nevertheless.

The real interest in the three axioms Overlap, Dilution and Set cut lies not only in the fact that they hold in our intended interpretation, but that they are *complete* with regard to it: given a set E, *any* choice of sets of possible entities as the possible worlds corresponds to a unique necessitation relation. Let a *possible world system* Ω on a set of entities E be a selection of subsets of E, to be taken as a specification of which combinations of entities can make up a world. We can then show:

Theorem 4.3 (*Representation of necessitation relations*) : Let E be a set of possible entities. Then every possible world system Ω on E determines a unique necessitation relation $\mathrel{\succ\!\!\in}$, and every necessitation relation $\mathrel{\succ\!\!\in}$ on E determines a unique possible world system through the correspondence that $W \in \Omega$ iff $W \not\mathrel{\succ\!\!\in} W^C$, for every $W \subseteq E$.

Proof. Let Ω be a set of subsets of E. Let the necessitation relation $\mathrel{>\!\!\!\in}_\Omega$ *characterised* by Ω be the relation

$$X \mathrel{>\!\!\!\in}_\Omega Y \text{ iff } (\forall\omega \in \Omega)(X \subseteq \omega \rightarrow Y \cap \omega \neq \varnothing)$$

We need to show that a binary relation on $\wp(E)$ is a nondeterministic necessitation relation (i.e. that it fulfils the axioms Overlap, Dilution, and Set cut) iff it is characterised by a set of possible worlds Ω. The right-to-left direction is mostly a matter of verification, and we have given the outlines of a proof in the discussion above. For the left-to-right direction, let $\mathrel{>\!\!\!\in}$ be a necessitation relation, and let $\Omega(\mathrel{>\!\!\!\in})$ be those subsets $W \subseteq E$ such that $W \mathrel{\not>\!\!\!\in} W^C$. We show that $\Omega(\mathrel{>\!\!\!\in})$ is one-to-one and onto.

For injectivity, assume that $\mathrel{>\!\!\!\in}_1$ and $\mathrel{>\!\!\!\in}_2$ are different necessitation relations. We then wish to find some W such that $W \mathrel{\not>\!\!\!\in}_1 W^C$ but $W \mathrel{>\!\!\!\in}_2 W^C$, or vice versa. Assume, without loss of generality, that there are $X, Y \subseteq E$ such that $X \mathrel{>\!\!\!\in}_1 Y$ but $X \mathrel{\not>\!\!\!\in}_2 Y$. We must have that $X \cap Y = \varnothing$, or we would have $X \mathrel{>\!\!\!\in}_2 Y$. By Set cut, it follows that there must be some W such that $X \cup W \mathrel{\not>\!\!\!\in}_2 W^C \cup Z$. But we must have that $X \subseteq W$ and $Y \subseteq W^C$, for otherwise W would overlap Y, or W^C would overlap X, so $W \mathrel{\not>\!\!\!\in}_2 W^C$. On the other hand, by Dilution, we must also have that $W \mathrel{>\!\!\!\in}_1 W^C$, so $\Omega(\cdot)$ is one-to-one.

To prove surjectivity of $\Omega(\cdot)$, assume that Ω' is any possible world system. Then $\mathrel{>\!\!\!\in}_{\Omega'}$ is a necessitation relation, and $\Omega(\mathrel{>\!\!\!\in}_{\Omega'})$ is the set of possible worlds

$$\Omega'' = \left\{ W \subseteq E \;\middle|\; (\exists\omega \in \Omega')(W \subseteq \omega \wedge W^C \cap \omega = \varnothing) \right\}$$

Some quick set-theoretical rearrangement shows that $\Omega'' = \Omega'$, so Ω' must be in the image of $\Omega(\cdot)$. It follows that $\Omega(\cdot)$ is a one-to-one correspondence with inverse $\mathrel{>\!\!\!\in}_{(\cdot)}$. □

The theorem, as well as the axioms Overlap, Dilution and Set cut, are taken from Shoesmith & Smiley's book on multiple-conclusion logics (Shoesmith and Smiley, 1978), and the structure of a necessitation relation is equivalent to that of such a logic. Multiple-conclusion logic originated with Gentzen's introduction of *Sequenzen* in his thesis (Gentzen, 1934) and Carnap's of *involutions* in *The Formalization of*

Logic (Carnap, 1943). Gentzen's work was purely proof-theoretical, and he therefore viewed the disjunctive conclusions as nothing but a useful notational apparatus, while Carnap's point of view was semantic, which made him arrive at multiple-conclusion consequence relations, or something equivalent to them, as absolutely necessary for capturing the semantics of propositional logics. Our reason for adopting the structure of a multiple-conclusion consequence relation is however not Carnap's, since our relation of necessitation transmits *existence*, and Carnap's transmits *truth*. It is also not Gentzen's: we really *need* the multiple conclusions for the extra representative power they give, as we will show later in this section.

Say that the necessitation relation $\mathrel{\succ\!\!\in}$ *extends* the binary relation R on $\wp(E)$ if $R \subseteq \mathrel{\succ\!\!\in}$. As the following theorem shows, any binary relation on $\wp(E)$ can be uniquely extended to a minimal necessitation relation.

Theorem 4.4 : If $\mathbf{N}$ is a set of necessitation relations that extend a binary relation R, then $\mathrel{\succ\!\!\in} = \bigcap \mathbf{N}$ is a necessitation relation that extends R.

Proof. As usual, Dilution and Overlap are easy to prove. For Set cut, assume that $X \not\mathrel{\succ\!\!\in} Y$. Then there must be some W such that $X \subseteq W, Y \subseteq W^C$, and $W \not\mathrel{\succ\!\!\in} W^C$, because this has to hold for all members of $\mathbf{N}$, and the intersection of these relations cannot have necessitations that hold but do not hold in any of them individually. If $\mathbf{N}$ is empty, $\mathrel{\succ\!\!\in}$ is the intersection of all necessitation relations on E. That there is such a relation is proved in the next theorem. □

We refer to the least necessitation relation containing R as the *closure* $Cl(R)$ of R or the necessitation relation *generated* by R. We can use this operator to define a *minimal* necessitation relation $Cl(\varnothing)$ on any set of possible entities. This relation, which captures the necessitations common to *all* nondeterministic necessitation relations, is uniquely determined by the following property.

Theorem 4.5 : If $\mathrel{\succ\!\!\in}$ is the minimal necessitation relation on E, then $X \mathrel{\succ\!\!\in} Y$ iff $X \cap Y \neq \varnothing$.

Proof. The right-to-left direction is simply the axiom Overlap. In the other direction, we need to prove that $X \mathrel{>\!\!\!\!\Rrightarrow} Y$ as defined is a necessitation relation. Overlap and Dilution are trivial, and Set cut follows because if $X \not\mathrel{>\!\!\!\!\Rrightarrow} Z$, then $X \cap Z = \varnothing$, and we can take $Y = Z^C$ or $Y^C = X^C$. Because of how $\mathrel{>\!\!\!\!\Rrightarrow}$ has been defined, it follows that $X \cup Y \not\mathrel{>\!\!\!\!\Rrightarrow} Y^C \cup Z$. □

Any necessitation relation $\mathrel{>\!\!\!\!\Rrightarrow}$ gives rise to a number of related concepts. First of all, we can allow ourselves to extend it to single elements by declaring $X \rightarrowtail b$ to be $X \mathrel{>\!\!\!\!\Rrightarrow} \{b\}$, and $a \mathrel{\circ\!\!\!\rightarrow} b$ to be $\{a\} \mathrel{>\!\!\!\!\Rrightarrow} \{b\}$. We can also introduce the following derived relation:

> The X's *distributively necessitate* some Y (in symbols $X \mathrel{\gg\!\!\!\!\Rrightarrow} Y$) iff, for every $x \in X$, $\{x\} \mathrel{>\!\!\!\!\Rrightarrow} Y$. In the intended interpretation, this means that $X \mathrel{\gg\!\!\!\!\Rrightarrow} Y$ iff some element of Y exists in every possible world where some element of X exists.

Distributive necessitation has a preorder structure, but not generally that of an order, since two sets of entities may necessitate one another without being identical. The reason why we have chosen $\mathrel{>\!\!\!\!\Rrightarrow}$ as our primitive relation here and defined others in terms of it is that the opposite would have been impossible. For every deterministic necessitation relation, there are several nondeterministic ones extending it. That necessitation cannot be written in terms of the distributive variant can be shown by noting that no distributive necessitation relation can exclude the empty set as a possible world, but any relation $\varnothing \mathrel{>\!\!\!\!\Rrightarrow} X$ where $X \neq \varnothing$ does.

Extending the terms "deterministic" and "singular", we say that $\mathrel{>\!\!\!\!\Rrightarrow}$ is deterministic iff it is the closure of a deterministic necessitation relation $\rightarrowtail$, and singular iff it is the closure of a singular necessitation relation $\mathrel{\circ\!\!\!\rightarrow}$.These properties correspond to intuitive ones on the set of possible worlds.

Lemma 4.6 : If $\mathrel{>\!\!\!\!\Rrightarrow}$ is deterministic, $X \mathrel{>\!\!\!\!\Rrightarrow} Y$ iff $X \mathrel{>\!\!\!\!\Rrightarrow} \{y\}$ for some $y \in Y$.

Proof. Assume that $X \succ\!\Rrightarrow \{y\}$. Then by Dilution $X \succ\!\Rrightarrow Y$ for all Y that contain y. In the other direction, assuming that $\succ\!\Rrightarrow$ is deterministic, we have that there is a deterministic relation $\rightarrowtail$ such that $X \succ\!\Rrightarrow Y$ iff $X \succ\!\Rrightarrow Y$ holds in all extensions of $\rightarrowtail$. But these extensions all satisfy $X \rightarrowtail y$ for some $y \in Y$. □

Lemma 4.7 : If $\succ\!\Rrightarrow$ is singular, $X \succ\!\Rrightarrow Y$ iff $\{x\} \succ\!\Rrightarrow \{y\}$ for some $x \in X$ and $y \in Y$.

Proof. Parallel to the preceding lemma. □

Theorem 4.8 : $\succ\!\Rrightarrow$ is deterministic iff Ω is closed under arbitrary non-empty intersections. It is singular iff Ω is closed under arbitrary non-empty intersections and unions.

Proof. Let us start with the deterministic case, left-to-right. Assume that $\mathcal{W}$ is a set of subsets of entities such that $W \not\succ\!\Rrightarrow W^C$ for all $W \in \mathcal{W}$. Since $(\bigcap \mathcal{W})^C = \bigcup \mathcal{W}^C$, we have that if $\bigcap(W) \succ\!\Rrightarrow (\bigcap \mathcal{W})^C$, there must be some element y in $\bigcup \mathcal{W}^C$ such that $\bigcap(W) \succ\!\Rrightarrow \{y\}$. But if this holds, there must have been some $W \in \mathcal{W}$ such that $W \succ\!\Rrightarrow W^C$ as well, contrary to assumption.

In the other direction, assume that $X \succ\!\Rrightarrow Y$. Hence all worlds ω that contain X contain some Y as well. So let ω' be the intersection of all worlds that contain X. Since ω' by assumption is a world as well, we have that $X \succ\!\Rrightarrow \omega'$ as long as $\omega' \neq \varnothing$, and thus $X \succ\!\Rrightarrow \{y\}$ for every $y \in \omega'$. On the other hand, if we were to have $\omega' = \varnothing$, then we would have $X = \varnothing$ as well, and thus $\varnothing \succ\!\Rrightarrow \varnothing$, contradicting the non-triviality axiom.

For singular necessitation, closure under intersections follows in the same way as for deterministic. Let $\mathcal{W}$ be a set of subsets of entities such that $W \not\succ\!\Rrightarrow W^C$ for all $W \in \mathcal{W}$. Then we must have that if $\bigcup \mathcal{W} \succ\!\Rrightarrow (\bigcup \mathcal{W})^C$, then $\bigcup \mathcal{W} \succ\!\Rrightarrow \bigcap \mathcal{W}^C$, so by assumption there are $x \in \bigcup \mathcal{W}$ and $y \in \bigcap \mathcal{W}^C$ such that $\{x\} \succ\!\Rrightarrow \{y\}$. But then we could not have had that $W \not\succ\!\Rrightarrow W^C$ for all $W \in \mathcal{W}$, for x has to be in some $W \in \mathcal{W}$, and y in all W^C where $W \in \mathcal{W}$.

Conversely, let $X \succ\!\!\Rrightarrow Y$. That $X \succ\!\!\Rrightarrow \{y\}$ for some $y \in Y$ follows in the same way that it did for deterministic necessitation. Now assume, for contradiction, that $\{x\} \not\succ\!\!\Rrightarrow \{y\}$ for all $x \in X$. Then, for every $x \in X$, there is some world ω_x such that $x \in \omega_x$, but $y \notin \omega_x$. But if the worlds are closed under unions, there must be a world that contain *all* the x's as well, which is the union of the ω_x. But this world cannot contain y either, so we cannot have $\{x\} \not\succ\!\!\Rrightarrow \{y\}$ for all $x \in X$. □

Thus, if we were to limit ourselves to necessitation relations whose right-hand side is determined, we would be unable to treat systems of possible worlds where overlap between worlds does not mean that the overlapping parts make up worlds themselves. An example which requires this is the traditional notion of a *substance*: if s is a substance, we may have that it must have some property in every possible world, since there can be no such thing as a bare substance, but we may also have that there is no single property that it has in all worlds. Such a substance would be impossible to represent if we were to limit ourselves to deterministic necessitations. If we were to accept only singular necessitations, on the other hand, we would be unable to represent joint necessitation: the case where a and b together necessitate c, but neither a nor b can do this on their own.

The versatility of nondeterministic necessitation relations allows us to express several important concepts using them. Let us call a world system *essentially possibilistic* if $E \notin \Omega$, i.e. if the set of all possible objects do not make up a world. Another way to write the same condition is $E \succ\!\!\Rrightarrow \varnothing$. This is the case if there are any incompatible possible objects. An *in*essentially possibilistic world system contains only things that *could* be actual together, or, in other worlds, things which together make up a possible world. Thus the difference between essentially and inessentially possibilistic world systems corresponds to a difference between the concept of *metaphysic*, and that of *possible world.* Only in some specific kinds of metaphysics do they coincide.

For any essentially possibilistic world system, the following hold, for all $X, Y \subseteq E$:

$\varnothing \succ\!\!\!\in X$ iff some X exists in every possible world. In the singular case $\varnothing \rightarrowtail a$, a is a *necessary existent.*

$X \succ\!\!\!\in \varnothing$ iff no possible world contains all the X's, i.e if the X's are incompatible. In the singular case where $\{a\} \succ\!\!\!\in \varnothing$, a is an *impossible object.*

$Y^C \succ\!\!\!\in X$ iff some Y exists in any world containing none of the X's, i.e. if the *lack* of X's means that some Y must exist.

$X \cup \{y\} \succ\!\!\!\in \varnothing$ iff y exists in no world all the X's exist in, i.e. if the existence of the X's *excludes* the existence of y.

The proofs of these are in general fairly straightforward, and we have omitted them. In the following, we will generally assume the necessitation relations under investigation to be essentially possibilistic, since this allows us larger freedom of expression. For reasons of simplicity, we will also assume metaphysics to contain no impossible objects, i.e. no object a for which $\{a\} \succ\!\!\!\in \varnothing$.

4.2 The Model Space $\mathcal{N}$

Necessitarian metaphysics can be taken to be model spaces in the following sense. For each necessitarian metaphysic $\mathcal{M} = \langle E, \succ\!\!\!\in \rangle$ we define a *model* to be a possible world $\omega \in \Omega$, together with the parts of $\succ\!\!\!\in$ that lie in ω. More specifically, a model $\mathfrak{M} \in \mathcal{N}$ is a set $W \subseteq E$ for which $W \not\succ\!\!\!\in W^C$, together with a relation $\succ\!\!\!\in_W$ such that $X \succ\!\!\!\in_W Y$ iff $X \succ\!\!\!\in Y$, for all $X, Y \subseteq W$.

A necessitarian model is a thus set of entities, together with a necessitation relation on the subsets of this set. It is a type of necessitarian metaphysic structurally, although one in which the set of all entities *does* make up a possible world. Thus, the necessitation relation of a model in $\mathcal{N}$ is always inessentially possibilistic. The interpretation is different as well: the entities in the set E of a necessitarian metaphysic are taken to be merely possible, while the elements in a model's set of entities are assumed to exist, if that model is the actual one.

Let a *submodel* of a necessitarian model $\mathfrak{M}_1 = \langle E_1, \mathrel{>\!\!\in}_1 \rangle$ be a necessitarian model $\mathfrak{M}_2 = \langle E_2, \mathrel{>\!\!\in}_2 \rangle$ such that

(i) $E_2 \subseteq E_1$,

(ii) $\mathrel{>\!\!\in}_2 = \mathrel{>\!\!\in}_1 \cap \wp(E_1)^2$, and

(iii) $E_2 \not\mathrel{>\!\!\in}_1 (E_1 \backslash E_2)$

The last condition guarantees that, at least as far as $\mathfrak{M}_1$ is concerned, the entities of $\mathfrak{M}_2$ may exist on their own.

Any necessitarian metaphysics determines a set of necessitarian models – one for each possible world. For the other direction, say that a two necessitarian models are *compatible* iff their necessitation relations agree on the subsets that contain entities in their intersection, and that a set $\mathcal{X}$ of necessitarian models is *closed* if it contains all submodels of each of the models it contains. Then any closed set $\mathcal{X}$ of pairwise compatible necessitarian models determines a necessitarian metaphysic, namely the one where the possible worlds are exactly those that are the sets of existent entities of the models in $\mathcal{X}$. A necessitarian metaphysic thus defined always has a necessitation relation that is an extension of those of its models, although in general this extension will not be minimal.

Since $\mathrel{>\!\!\in}$ constitutes a structure on a set of possible entities, it is natural to use this structure to characterise the elements of $\mathsf{hom}_{\mathcal{N}}$. Thus we define a morphism $f : \mathfrak{M}_1 \to \mathfrak{M}_2$, where $\mathfrak{M}_1 = \langle E_1, \mathrel{>\!\!\in}_1 \rangle$ and $\mathfrak{M}_2 = \langle E_2, \mathrel{>\!\!\in}_2 \rangle$ as a function $f : E_1 \to E_2$ such that

$$X \mathrel{>\!\!\in}_1 Y \Rightarrow f[X] \mathrel{>\!\!\in}_2 f[Y]$$

for all $X, Y \subseteq E_1$.

$\mathcal{N}$ is clearly a construct, just like the Tarskian model space $\mathcal{T}$. The natural forgetful functor F that takes it to $\mathcal{V}$ is the one that takes every model to its set of entities, and every morphism to its underlying function. Because of this, every morphism that has an injective underlying function is a monic. That the converse also holds, and that the epics are the surjective morphisms, are proven in the following theorems.

Theorem 4.9 : The monics of $\mathcal{N}$ are exactly the morphisms with injective underlying functions.

Proof. Only the left-to-right direction needs proving. But this follows from the existence of a free singleton model $\langle \{*\}, \mathrel{>\!\!\!\Rrightarrow} \rangle$, where $\{*\} \mathrel{>\!\!\!\Rrightarrow} \{*\}$ but no other necessitations hold (Adámek et al., 2004, §8.29). □

Theorem 4.10 : The epics of $\mathcal{N}$ are exactly the morphisms with surjective underlying functions.

Proof. Completely parallel to the corresponding theorem for Tarskian models. □

This also means that monics are not embeddings, and epics are not reductions in $\mathcal{N}$. Although injective or surjective, they may introduce all kinds of new necessitations that did not hold in their domain. As expected, strong monics and epics are what we are after.

Theorem 4.11 : A monic or epic f in $\mathcal{N}$ is strong iff $f[X] \mathrel{>\!\!\!\Rrightarrow} f[Y] \Rightarrow X \mathrel{>\!\!\!\Rrightarrow} Y$.

Proof. Let m be a strong monic $m : \mathfrak{M}_1 \to \mathfrak{M}_2$. Define $m' : \mathfrak{M}_1 \to \mathfrak{M}_2^{sub}$, where $\mathfrak{M}_2^{sub}$ is the submodel of $\mathfrak{M}_2$ whose set of entities consist of the image of $\mathfrak{M}_1$ under m. Then m' has to be an epimorphism, and since m is strong, there is a morphism $h : \mathfrak{M}_2 \to \mathfrak{M}_1$ such that $m' \circ h = 1_{\mathfrak{M}_2^{sub}}$ and $h \circ m' = 1_{\mathfrak{M}_1}$. Thus m is an isomorphism onto a submodel of $\mathfrak{M}_2$, and this means that their necessitation relations must coincide here.

Let $m : \mathfrak{M}_1 \to \mathfrak{M}_2$ be an embedding (i.e. an injective function such that $m[X] \mathrel{>\!\!\!\Rrightarrow} m[Y] \Leftrightarrow X \mathrel{>\!\!\!\Rrightarrow} Y$), let $e : \mathfrak{M}_1' \to \mathfrak{M}_2'$ be an epic, and let

$f : \mathfrak{M}_1' \to \mathfrak{M}_1$ and $g : \mathfrak{M}_2' \to \mathfrak{M}_2$ be morphisms such that $m \circ f = g \circ e$. We need to show that there is a unique morphism h such that $f = h \circ e$ and $g = m \circ h$. Since m is an embedding, it is an isomorphism i onto a submodel $\mathfrak{M}_2^{sub}$ of $\mathfrak{M}_2$. We can define h such that $h(x) = i^{-1}(g(x))$, for all $x \in E_{\mathfrak{M}_1}$. This is well defined since the image of g must coincide with that of m because e is an epic and the original diagram commutes, and it is a morphism because it is the composition of a morphism and an isomorphism.

For the epic case, let $e : \mathfrak{M}_1 \to \mathfrak{M}_2$ be a strong epimorphism, and write e as $e = m \circ e'$, where $e' : \mathfrak{M}_1 \to \mathfrak{M}_1^{sub}$, $i : \mathfrak{M}_1^s ub \to \mathfrak{M}_2$, and $\mathfrak{M}_1^{sub}$ is a submodel of $\mathfrak{M}_1$, such that m is an injection. Since injections are monomorphisms, we can apply the strongness condition to form a morphism $h : \mathfrak{M}_2 \to \mathfrak{M}_1$ such that the whole diagram commutes, and $e \circ h = 1_{\mathfrak{M}_2}$. Since h as well as e have to preserve the necessitation relation, we have that $X \mathrel{>\!\!\in} Y \Leftrightarrow e[X] \mathrel{>\!\!\in} e[Y]$.

Finally, assume that $e : \mathfrak{M}_1 \to \mathfrak{M}_2$ is a reduction. We need to show that whenever there is a monomorphism $m : \mathfrak{M}_1' \to \mathfrak{M}_2'$, and morphisms $f : \mathfrak{M}_1 \to \mathfrak{M}_1'$ and $g : \mathfrak{M}_2 \to \mathfrak{M}_2'$ such that these all commute, there is a unique $h : \mathfrak{M}_2 \to \mathfrak{M}_1'$ such that $f = h \circ e$ and $g = m \circ h$. Let e^{inv} be a function from $\mathfrak{M}_2$ to $\mathfrak{M}_1$, such that $e \circ e^{inv} = 1_{\mathfrak{M}_2}$ (such a morphism exists because e is a reduction). We can then let $h = f \circ e^{inv}$. □

The fact that $\mathcal{N}$ is a construct, with forgetful functor F, allows us to define a notion of identity preservation. Let a morphism f be *identity preserving* if $(F(f))(x) = x$, for all $x \in F(\mathsf{dom}(f))$. This means that the morphism's underlying set-theoretical function is an identity function. It follows that all identity preserving morphisms are monics, but they do not in general have to be strong. This is due to the fact that entities in $\mathcal{N}$ only have their properties in relation to a model, just as the elements of a domain in a model in $\mathcal{T}$. Not every identity function between two Tarskian models is an embedding, and the same holds for $\mathcal{N}$.

The notion of identity preservation allows us to say that two models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are *compatible* iff any identity preserving monic between them is strong. It is quickly checked that this definition is equivalent to the one we gave at the start of this section, and this means that we can see a necessitarian metaphysics as a subcategory of $\mathcal{N}$ wherein any two models are compatible. The entire hierarchy of $\mathscr{N}$, necessitarian metaphysics, and necessitarian models is depicted in fig. 4.1.

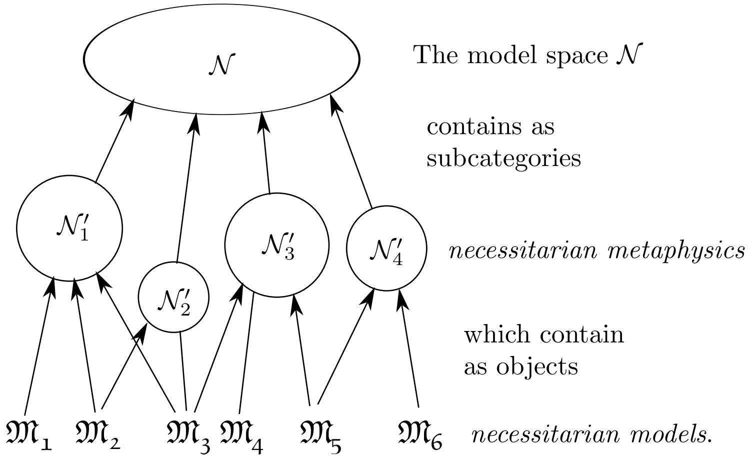

$\mathcal{N}$
The model space $\mathcal{N}$
contains as subcategories
$\mathcal{N}'_1$
$\mathcal{N}'_2$
$\mathcal{N}'_3$
$\mathcal{N}'_4$
necessitarian metaphysics
which contain as objects
$\mathfrak{M}_1$ $\mathfrak{M}_2$ $\mathfrak{M}_3$ $\mathfrak{M}_4$ $\mathfrak{M}_5$ $\mathfrak{M}_6$
necessitarian models.

Figure 4.1: ***Hierarchy of $\mathscr{N}$ and its parts***

.

Within a specific necessitarian metaphysics $\mathscr{N}'$, we can define canonical monics to be those morphisms that are identity preserving. In general, these will not be part of a complete inclusion system, though we can as before always extend $\mathscr{N}'$ to a model space in which factorisation of any morphism into an epic and a canonical monic is possible.

The strongest structural relationship between models—that of isomorphism—holds between models when they are of equal cardinality, and the same structure of necessitation relations hold in them. Reductions (i.e. strong epics) express a slight weakening of this concept. Say that two objects a, b in a model space are *equivalent* iff $a \circ\!\!\rightarrow b$ and $b \circ\!\!\rightarrow a$, and write this condition as $a \sim b$. Then a reduction is a surjective function that maps only equivalent objects to the same value.

Another way to express the relation $a \sim b$ is to note that if $a \sim b$, then a and b exist in exactly the same possible worlds. But this notion is clearly relative to which specific metaphysics we are envisaging; in the whole of $\mathscr{N}$ there are necessitarian metaphysics both where a are b

equivalent, and also metaphysics where there are not.

This is relevant to the question of whether necessary coexistence is interpretable as identity. On the abstract level of $\mathcal{N}$, there is no unambiguous notion of necessity, so holding necessary coexistence to imply identity does not make sense unless we specify a metaphysic. This, in turn, fixes the network of necessitations, and thus also the set of possible worlds.

On what interpretation of necessity *would* necessary coexistence imply identity? One example is first-order logical necessity—at least for the interpretation of identity used in that language. Another may be metaphysical necessity, although this naturally depends on the metaphysics. If there is such a thing as a *true* model space, then the question has a unique truth value. Otherwise, the only thing we can say is "that depends on what you mean by *necessary*". From the point of view of $\mathcal{N}$, a specific necessitarian metaphysics supplies an answer to that question.

As we noted in the last chapter, the choice of model space is indeed to a large part one of which conventions to adopt, and this naturally holds for necessitarian metaphysics as well. If $a \sim b$ in the metaphysic we have settled on, we are *allowed* to treat a and b as identical. *Are* they identical? If we say so. Identity is just as convention-laden as metaphysics; for any terms "*a*" and "*b*", there are contexts where they are replaceable *salva veritate*, and contexts where they are not. Of course, we can say that identity needs substitutability in all *transparent* contexts, but this gives us nothing so long as we do not specify exactly which contexts are transparent or opaque in a non-circular way. In fact, such specifications are conventional as well: by deciding to treat a context as opaque, we are saying nothing more than that $\varphi(a) \leftrightarrow \varphi(b)$ should not be inferable from $a = b$, where $\varphi(\cdot)$ is the context in question.

The relativity of identity has been noted several times before, most famously by Peter Geach (1967). According to Geach, all claims of identity involve a *sortal*, so that "*a* is the same as *b*" always will invite the question: the same *what*? This is not quite what I am saying here. Identity as we see it *is* relative, but to a convention rather than a sortal. In this, we are really more close to Quine in his post-ontological relativity writings. Is this *gavagai* the same as that one? That depends on whether "gavagai" refers to rabbits or undetached rabbit parts, and as Quine held, this is for the interpreter to decide.

4.3 Metaphysical Interpretations

The preceding sections introduced necessitarian models in an abstract setting, and we should now say something about how to interpret their features in terms of a more easily recognisable concepts. While such interpretations will be studied more formally in the next chapter, some largely informal remarks may help in making them more concrete.

As defined, a necessitarian metaphysics is a model space $\mathcal{M} \subseteq \mathcal{N}$ whose models are subsets of a possible entity set E, with necessitation relations that are subrelations of the same necessitation relation. It is thus a sort of specification of the ways the subject matter of a theory, and in the limit, the entire world, *could* be. By holding the necessitation relation fixed, we focus on only the necessary features of our world, and by requiring the actual world to be variable, we factor out the contingent.

Many metaphysical theories can be interpreted as necessitarian metaphysics, in the sense that we can express them in terms of necessitation, possible entities and possible worlds. Another way to say the same thing is that a necessitarian metaphysics is a kind of *representation* of a metaphysical theory. As such, we do not have to interpret it literally. Although we have let a world be a set of possible entities, for example, *the set itself* should not be taken to be that world. It would actually be better to say that the world is represented by a set, and if that world is the actual one, then the things that exist are those that are *in* the set, and no others. When interpreting $\mathcal{N}$'s fundamental concepts, we should allow ourselves some latitude.

Section 2.5 stated that we will neither assume nor rule out any absolute notion of necessity, and our usage of necessitation relations is not intended as such an assumption (or ruling out, of course). For now, all we need to know is the intended interpretation: the X's necessitate some Y iff any possible world that contains all the X's also contains at least one Y. These possible worlds are representable as models in $\mathcal{N}$.

Whenever $X \mathrel{>\!\!\!\Rrightarrow} Y$, this may hold for several reasons, or, as we shall put it, it may have different *bases.* Let us call an instance of a necessitation relation $\mathrel{>\!\!\!\Rrightarrow}$ (i.e. a pair X, Y such that $X \mathrel{>\!\!\!\Rrightarrow} Y$) a *necessitation.* Some examples of bases for necessitations are:

Semantic necessitation. The weakest necessitation relation that can hold in any $\mathcal{N}$ model is the minimal one according to which $X \mathrel{>\!\!\!\Rrightarrow} Y$ iff $X \cap Y \neq \varnothing$, which corresponds to the possible-world system where *any* set of possible entities makes up a unique possible world. This kind of necessitation follows simply from our semantics of the necessitation relation, so it does not have any metaphysical "punch", in that it cannot be used to distinguish $\mathcal{N}$ models from each other. If *all* necessitations in our metaphysic are semantic, then there is really no necessitation going on at all, and it is a "Tractarian" metaphysic, where we can never infer the existence or non-existence of one entity (in Wittgenstein's case a *Sachverhalt*) from the existence of another.

Mereological necessitation. The classical theory of mereology is that created by Leśniewski (1916) and Leonard and Goodman (1940) as a theory of the relation of *parthood.* At least on a certain reading (an *extensional* one), it is obvious that if an entity exists, then all its parts must exist as well. It is also a principle of classical mereology that if any entities exist, their *mereological sum*—the entity that overlaps all of those entities, and none others—must exist. Both these principles are easily captured by imposing the following condition on the metaphysic $\mathcal{M} = \langle E, \mathrel{>\!\!\!\Rrightarrow} \rangle$:

> For every non-empty $X \subseteq E$, there is a unique entity $\hat{X}$ such that
>
> $$X \rightarrowtail \hat{X}$$
>
> $$\hat{X} \mathrel{\circ\!\!\rightarrow} x \text{ for all } x \in X$$

We can define x to be a *part* of y iff $y = \widehat{\{x, y\}}$, and we will also generally write $x + y$ for the sum $\widehat{\{x, y\}}$. It quickly proved that this theory is equivalent to that of atomistic classical mereology: it defines a Boolean algebra of subsets of the set E with the empty set removed, and any such algebra is isomorphic to an atomistic mereology, as was proved by Tarski (1956, ch. II). We will call any necessitarian metaphysic for which this axiom holds *unrestrictedly mereological.*

Necessitation relations also allow us to define weaker forms of mereology. The simplest ones are those that do not fulfil what is commonly referred to as *unrestricted composition* (that any non-empty set of entities has a sum). We do this by changing "for every non-empty $X \subseteq E$" to "for every $X \in \mathcal{S}$", where $\mathcal{S}$ is some set of subsets of E. One such limitation, which is quite reasonable, is to limit composition to entities that are compatible in the sense that there is some possible world that contains all of them. When $\mathcal{M}$ fulfils the axiom for the class $\mathcal{S} = \{X \subseteq E \mid X \not\mathrel{\succ\!\!\Rrightarrow} \varnothing\}$, we say that $\mathcal{M}$ is *restrictedly mereological*, or simply just *mereological*, since this will be the case most useful for us in later chapters. Adding a restricted mereological structure to a metaphysics does not introduce impossible objects, unlike the imposition of an unrestricted mereology.

The generality of necessitation relations also makes it possible to relax the extensional aspects of classical mereology. What if, for instance, I must have some heart to exist, but there is no specific heart that I must have, since I do not go out of existence by a heart transplant? In such a case, we would have a mereological necessitation relation that fulfilled $\{me\} \mathrel{\succ\!\!\Lleftarrow} \{heart_1, heart_2, \ldots\}$ without fulfilling any of $me \mathrel{\circ\!\!\rightarrow} heart_1$, $me \mathrel{\circ\!\!\rightarrow} heart_2$, etc.

In fact, we can be even more general. Mereological structure is very useful to have in a metaphysic, but it is common for the *existence* of sums to be more important than their uniqueness. We can thus call $\hat{X}$

a sum of X if it fulfills the criteria for being a mereological sum, but may be non-unique in doing so. It is simple to prove that if a and b are sums of X, then $a \sim b$, so such sums are necessarily coexistent. If sums are non-unique, $\hat{\cdot}$ is not really a function, but since all sums must be equivalent, we can treat it as one nevertheless: it does not matter which of the sums of X we choose to represent it, since they all have the same place in the necessitation structure.

Causality. Unlike what is the case in mereology, there is no formal "classical" theory of causality. This does not stop many such theories from being interpretable in terms of necessitation, however. Those easiest to represent are the deterministic ones, which always are singular. Mackie's version, from *The Cement of the Universe* (Mackie, 1974), in which the natural language word "cause" is interpreted as "*INUS* condition", is a theory of this type. According to Mackie, c_1 is an *INUS* condition of e iff c_1, together with some other (unspecified) conditions $c_2, c_3, \ldots$, are sufficient (but usually not necessary) for e, and the $c_2, c_3, ...$ are insufficient for e on their own. We can write this simply as $\{c_1, c_2, c_3, \ldots\} \rightarrowtail e$, together with $\{c_2, c_3, \ldots\} \not\rightarrowtail e$.

An historically important class of theories is the regularist one, where causality is a relation holding between events (or possibly facts, tropes, or something else), such that if e_1 causes e_2, then e_1 precedes e_2, events of e_2's type generally follow events of e_1's type, and e_1 and e_2 are continuous in space and time. Strengthening the second clause to "events of e_2's type invariably follow events of e_1's type", we arrive at the deterministic, regularist notion of causation described by Hume.

One advantage of our necessitation framework is that it readily allows indeterministic causation as well. If e_1 can have either of the effects e_2 or e_3, but must have at least one one of them, we have the necessitation $e_1 \succ\!\!\Lleftarrow \{e_2, e_3\}$, but neither $e_1 \succ\!\!\Lleftarrow e_2$ nor $e_1 \succ\!\!\Lleftarrow e_3$. If e_1 *precludes* e_2, then causality is a basis for the necessitation $\{e_1, e_2\} \succ\!\!\Lleftarrow \varnothing$, i.e. the incompatibility of e_1 and e_2.

More specific types of nondeterministic causality are allowed by letting the necessitation relation be probabilistic, as will be described in the next section. Using such relations, we can capture not only general facts such as "e_1 causes e_2 or e_3", but also those of the form "e_1 gives an $x\%$ chance of e_2".

Ontological dependence. Some metaphysical entities are thought to depend on others. A *non-transferable trope*, for instance, depends for its existence on a specific thing, of which it is a property (cf. Martin, 1980). For Kripke, *person a's existence* presupposes *person a's origin*, since he takes someone's specific origin to be essential to that person (Kripke, 1981, pp. 110–115).

A more complex notion is that of *generic dependence.* A *substance*, as we mentioned earlier in this chapter, may require the instantiation of some property in order to exist (to avoid the problem of "bare substances"), but it may not need the instantiation of any *specific* property. Object a's redness may require a to have some exact shade of red, but not necessarily any specific shade. It is clear that these cases are also representable as ones of necessitation, although ones which are non-singular.

Later on, we will also encounter *dichotomous* metaphysics, in which the set of possible entities is partitionable into pairs, and every world must contain exactly one entity from each such pair. An example is where we require every truth to have something that makes it true (a *truthmaker*), and use a language which conforms to the laws of classical logic. Since either p or $\neg p$ must hold in every world, every world must contain a truthmaker for either p or for $\neg p$. It is hard to say what grounds this "must", except perhaps ontological dependence: the *wholeness* implicit in being a world requires some truthmaker for every statement to exist, but *different* worlds can still have different truthmakers in them.

Necessitation relations used in scientific or commonsense models usually have instances based on several of these phenomena, since necessitations *combine*: as we noted, for any set of necessitations, there is a smallest necessitation relation that includes all of these. Differently put, if we have that $R(X, Y)$, for some sets $X, Y \subseteq E$ and some binary relation R on $\wp(E)$, we can extend this relation in a unique way to a necessitation relation $\mathrel{>\!\!\!\Rrightarrow}_R$ that fulfils the axioms *Overlap*, *Dilution* and *Set Cut.* Such an extension does not add any "real" necessitations to the model, but only semantic ones.

Necessitation relations and possibility of worlds are two ways to view the structure of a necessitarian metaphysic. But just as we started with the inference structure of a theory, and laid an algebraic structure on it in section 2.4, we can also start with a necessitation relation, or a part of one, and impose an algebra (total, as in chapter 2, or possibly partial) on that. One example is a formalisation of mereology as an algebra, with the operations of mereological sum, product (the overlap, if any, between two things) and complement (the sum of everything that does not overlap what we take the complement of). This would be an example of a partial algebraisation: the product of a and b is only defined when a and b overlap, and if we do not have unrestricted composition, neither do they always have a sum.

Formally, such an algebraisation works much like the algebraisations of the preceding chapter. First, we need a notion of *structurality*: let $\mathfrak{A}$ be an algebra whose carrier set A is a subset of the possible entity set E of a necessitarian metaphysic $\mathcal{M}$. We then say that $\mathfrak{A}$ is *structural in* $\mathcal{M}$, or an *algebraisation of* $\mathcal{M}$, iff

> If $X \mathrel{>\!\!\!\Rrightarrow} Y$, then $\varepsilon[X] \mathrel{>\!\!\!\Rrightarrow} \varepsilon[Y]$, for any $X, Y \subseteq A$ and any endomorphism ε on $\mathfrak{A}$.

Just as with an algebraic theory, the algebraisation of a necessitation relation allows us to view necessitation as based on the structure of the model, rather than on individual objects' intrinsic properties. This distinction could be useful for the separation of *natural law* and singular causality. A suggestion would be that the *laws* are those necessitation relations that are invariant under certain endomorphisms, although this naturally requires the world to have an algebraic structure. Different theories of natural law would then correspond to different algebraisations.

So much for the necessitation relations. We come now to the other part of an $\mathcal{N}$ model: the entities. These could be said to differ somewhat from some of the things that are commonly called *objects*, in that we have taken what entities exist to fully determine the identity of a possible world. But some (see Dodd, 2001; Lewis, 2001) would say that more is required: we need to know not only what things there are, but also what these things are like, and how they are related. Being told that the rose r exists in world ω does not suffice to settle the question of whether ω contains a red rose or not, since we also need to determine whether r is red or white, for instance. According to this line of thought, r *by itself* does not determine its colour, since it might have had a different colour (for instance, by having been dyed) and still existed.

Yet, there is also a sense in which existence *must* be enough: if I am told exactly what things exist in a world, and what things do not exist, I am told everything that is possible to know about that world. This is clearest when we approach the problem in terms of predicate logic. Classically, *any* sentence can be rewritten as either an existential sentence (i.e. a sentence that begins with an existential quantifier), or a negative existential sentence (one that begins with a negated existential quantifier), and this holds for several extensions of classical predicate logic as well, such as some modal logics. If what sentences are true in a world determine what world it is, then what things exist in it must determine it as well.

The solution to this problem consists in the recognition that we have used words like “object”, “thing” and “entity” interchangably here, and the notion of *entity* that is embodied in necessitarian model theory cannot do the work of all these as it stands, since it is purely extensional. For a notion of object that can have different properties in different possible worlds, we may take some inspiration from trope theory. According to trope theory, a property is a *particular*, unique to the thing that has it, so that my humanness and my friend’s humanness are distinct entities. What makes both into *humanness* is an equivalence relation of *exact similarity* that holds between them. Furthermore, what makes my humanness and my two-leggedness into properties of the *same* object is the equivalence relation *compresence* (sometimes called *concurrence)* that holds between any two tropes that are properties of the same object. [3]

We can use these ideas to let an ordinary object o, with all its properties, correspond to a maximal compresent set $c(o)$ of entities,

[3] The word “trope” (or this usage of it, at least) comes from D.C. Williams’s article *On the Elements of Being* (Williams, 2004).

spread out over several possible worlds. The interpretation of this is that if $c(o)$ is an object-set, and ω a possible world, then the tropes that are in $c(o) \cap \omega$ determine what properties o has in world ω, and if $c(o) \cap \omega = \varnothing$, then o does not exist in ω at all. Given a property P, we let $c(P)$ be a maximal set of similar tropes. The predication $P(o)$ can then be taken to be true iff $c(o) \cap c(P) \cap \alpha \neq \varnothing$, i.e. if the entities of o that are in the actual world overlap those that are the tropes of the property P. We can also say that o is P *essentially* iff $c(o)$ overlaps $c(P)$ in every possible world in which o exists.

This can be extended to n-place relations as well. The instantiation of a relation is, however, a more complicated thing than that of a property, since there is only one way for an object to instantiate a property, but there are two ways for a pair of objects to instantiate a binary relation ($R(o_1, o_2)$ and $R(o_2, o_1)$), six ways for a triple of objects to instantiate a ternary relation, etc.

Just as we did in the case of properties (or 1-place relations), we associate each relation R with a maximal set $c(R)$ of similar tropes. Each trope in such a relation is then taken to be the instantiation *in a certain place* by the object that trope belongs to. Thus, for any natural number i, we say that the trope t is an ith-place trope iff t is the instantiation of some object o in the place i of R. Call these sets of tropes $N_1, N_2, \ldots$. These are clearly disjoint, and if there are no infinite-arity relations, they exhaust the class of possible tropes. The sets $N_1, N_2, \ldots$ then partition the set $c(R)$ into the n non-overlapping subsets $c_1(R), \ldots, c_n(R)$. Using these sets, we can define what it means for a relation to hold in ω by taking $R(o_1, \ldots, o_n)$ to be true iff

$$c_i(R) \cap o_i \cap \omega \neq \varnothing$$

for all i from 1 to n. It is easily checked that this way of defining relations in terms of tropes lets the holding of $R(o_1, o_2)$ be something else than that of $R(o_2, o_1)$.[4]

Now, both similarity and compresence can be taken to be supervenient on their relata, i.e. we have that if t_1 is similar to t_2, this holds in any possible world in which t_1 and t_2 exist, and if t_1 and t_2 are properties of the same object, they are so in any world in which they exist.

[4] For another way to handle relations as tropes, see Bacon, 1995, ch. 2.

Both of these actually follow from similarity and compresence having been defined directly on the set of possible entities, rather than relative to some world.

How should we then interpret sets such as $c(o)$, for an object o, or $c(P)$, for a property P? Are these among the possible entities? They do not have to be: $c(o)$ is used in our definition of truth for $P(o)$, but the tropes themselves are doing all the actual work by being bases for the similarity and compresence relations. From these, all the objects and the properties follow, and what c does is the purely semantic function of associating a singular term or a predicate with some possible entities. Instead of taking c to be a function to sets of entities, we could just as well have taken it to be a relation to the entities themselves, and thereby have avoided mentioning sets at all.

Another way to represent objects with contingent properties is use intensional semantics in the vein of Carnap (1956) and Montague (1970, 1973). We can let an object o be a partial function $o : \Omega \to E$ such that $o(\omega) \in \omega$, whenever o is defined at ω. Intuitively, $o(\omega)$ says what entity, if any, o is in the world ω. A *necessary existent* would be an object which is a total function. A *relation* R can be defined simply as its extension, i.e. a set of the n-tuples of possible entities that satisfy the relation.

Given these notions, we can define the predication $P(o_1, \ldots, o_n)$ to be *true* in world ω iff all of $o_1, \ldots, o_n$ are defined at ω and

$$\langle o_1(\omega), \ldots, o_n(\omega) \rangle \in P$$

We say that o has property P *necessarily* iff $\langle o(\omega) \rangle \in P$ for all worlds $\omega \in \Omega$ where o is defined, and that o has P *contingently* iff $\langle o(\alpha) \rangle \in P$, but $\langle o(\omega) \rangle \notin P$ for some other world ω where o is defined, where α is the actual world.

A function o is a version of what Carnap calls an *individual concept* (Carnap, 1956, p. 40). It allows us to identify the bearer of the name "o" in different worlds, and may be seen as a meaning specification of that name. Since a predicate P is defined directly on the possible entities, and not relativised to a world (unlike an individual concept), everything true about an entity e remains true in all worlds. However, different entities correspond to the same object in different worlds, and this is how we may have that the same object has different properties in different worlds.

In neither the trope-based solution, nor in the intensional one, do we need to have something in our set of possible entities that corresponds one-to-one with the "ordinary" object o. Despite this, talk about ordinary objects can be paraphrased into talk about possible entities. Thus we hold that even objects of a more Aristotelian flavour are representable in necessitarian metaphysics, and that requiring worlds to be determined by what entities exist in them imposes no significant limitation. As a model space, $\mathcal{N}$ is a way to speak and think about the world, and it seems to be a way that has enough structure for it to be useful for the purposes at hand.

4.4 Probabilistic Necessitation

While necessitation in its various forms definitely has uses in science, probabilistic relationships are even more common. It is therefore important to indicate how the above metaphysics can be extended in that direction, and this is the aim of the current section. Unfortunately, we will not be able to delve into any depths regarding these metaphysics, but since they are useful for quantum mechanics, which we want to be able to say something about later, we will at least give an outline of what they could be like.

Let a *probabilistically necessitarian metaphysic* be a pair $\langle E, N \rangle$, where E is set of possible entities, and N is a function from $\wp(E)^2$ to the real interval $[0, 1]$ called *probabilistic necessitation.* The intended interpretation of N is that $N(X, Y) = \pi$ iff any world that contains all the X's has a chance π to contain some Y. If $N(X, Y) = \pi$ holds, we also write $X \overset{\pi}{\rightarrowtail} Y$ to highlight the connection with nonprobabilistic necessitation.

It is clear that N must have other properties than $\rightarrowtail$, and we would like it to be a generalisation of such a relation in the sense that

$$X \overset{1}{\mathrel{>\!\!\!\Rrightarrow}} Y \text{ iff } X \mathrel{>\!\!\!\Rrightarrow} Y$$

for some *non*probabilistic necessitation relation $\mathrel{>\!\!\!\Rrightarrow}$. Thus, when the necessitation is *certain* (or at least almost certain), it should conform to the laws of nondeterministic necessitation. This allows us to define a *possible world* in the same way as before. We let the possible worlds induced by a probabilistic necessitation relation be the set

$$\Omega = \left\{ W \subseteq E \;\middle|\; W \not\overset{1}{\mathrel{>\!\!\!\Rrightarrow}} W^C \right\}$$

as before. Thus a possible world is a set of entities, such that when these exist, we cannot be certain that something else exists as well. There is one thing that is worth noting here: probability 1 is not usually taken to be the same as certainty (the traditional word for it is *almost certainty*). However, for the applications we will have for our probabilistic necessitation relations, the difference will be negligible. It does, however, have a few interesting consequences, such as the fact that no world can contain an infinity of non-necessary, probabilistically independent objects.

When *un*certain, probabilistic necessitation should conform to the laws of probability theory. Here we make life considerably simpler for us if we assume the metaphysic in question to be mereological. In this case, it means that for every non-empty set of entities X such that $X \not\overset{1}{\mathrel{>\!\!\!\Rrightarrow}} \varnothing$, there is an entity $\widehat{X}$ such that

$$X \overset{1}{\mathrel{>\!\!\!\Rrightarrow}} \widehat{X}$$

$$\{\widehat{X}\} \overset{1}{\mathrel{>\!\!\!\Rrightarrow}} \{x\} \text{ for all } x \in X$$

Thus, a sum $\widehat{X}$ of some set of compatible entities X is an entity that we can be sure exists iff all the entities in X exist. As before, we write $x + y$ for the sum $\widehat{\{x, y\}}$. With sums, we can define the following important concept. Let the *cross-sum* $\otimes\mathcal{X}$, where $\mathcal{X}$ is a set of sets of entities, be a set of entities that contains a sum $\widehat{X}$ for each consistent set X containing at least one entity from each set in $\mathcal{X}$:

$$\otimes\mathcal{X} \underset{def}{=} \left\{ \widehat{Y} \;\middle|\; Y \subseteq E \text{ and } Y \cap X \neq \varnothing \text{ for all } X \in \mathcal{X} \right\}$$

A cross-sum is thus a set of sums of entities that contains something from each set in a set of sets. We write the cross-sum of the sets X and Y as $X \otimes Y$.

We also have use for the following relations on $\wp(E)$, which we will call *weak orthogonality*, $\forall$-*equivalence* and $\exists$-*equivalence*:

$$X \perp Y \underset{def}{=} (\forall x \in X)(\forall y \in Y)(\{x, y\} \overset{1}{\succ\!\!\!\in} \varnothing)$$

$$X \overset{\forall}{\sim} Y \underset{def}{=} (\forall y \in Y)(X \overset{1}{\succ\!\!\!\in} \{y\}) \text{ and } (\forall x \in X)(Y \overset{1}{\succ\!\!\!\in} \{x\})$$

$$X \overset{\exists}{\sim} Y \underset{def}{=} (\forall x \in X)(\{x\} \overset{1}{\succ\!\!\!\in} Y) \text{ and } (\forall y \in Y)(\{y\} \overset{1}{\succ\!\!\!\in} X)$$

In our intended interpretation, $X \perp Y$ means that no X can coexist together with any Y. $X \overset{\forall}{\sim} Y$ holds iff any world that contains all X's also contains all Y's, and vice versa, and $X \overset{\exists}{\sim} Y$ means that the worlds that contain some X coincide with those that contain some Y. The last of these could also be expressed as the condition that X and Y distributively necessitate one another.

Using this array of concepts, we can give axioms for N:

(*Necessitation*)	$\overset{1}{\succ\!\!\!\in}$ is a nondeterministic necessitation relation
(*Equivalence*)	if $X \overset{\forall}{\sim} X'$ and $Y \overset{\exists}{\sim} Y'$, then $N(X,Y) = N(X',Y')$
(*Additivity*)	if $Y \perp Z$ and Y and Z are non-empty, $N(X, Y \cup Z) = N(X,Y) + N(X,Z)$
(*Conditionalisation*)	$N(X, \{\hat{Y}\} \otimes Z) =$ $N(X, \{\hat{Y}\})\, N(X \cup Y, Z)$

The first of these follows from our intention to have the relation $\overset{1}{\succ\!\!\!\in}$ conform to the rules of nondeterministic necessitation. Equivalence is required for us to be able to assign probabilities to sets of worlds, rather than to just pairs of sets of entities. Additivity gives us the additivity of probability across non-overlapping sets of worlds, and conditionalisation makes possible the interpretation of N in terms of *conditional* probability.

The main advantage of the axioms is that they allow us to interpret the probabilistic necessitation relation N as we intended. For this purpose, it is useful to, whenever Ω is the set of possible worlds, define the two functions

$$\Omega^{\forall}(X) = \{\omega \in \Omega \mid (\forall x \in X)\ x \in \omega\}$$

$$\Omega^{\exists}(X) = \{\omega \in \Omega \mid (\exists x \in X)\ x \in \omega\}$$

i.e. $\Omega^{\forall}(X)$ is the set of those worlds that contain all entities in X, and $\Omega^{\exists}(X)$ is the set of worlds that contains some of the entities in X. Using these, we may check that the relations $\perp$, $\overset{\forall}{\sim}$ and $\overset{\exists}{\sim}$ behave as we would expect them to.

Theorem 4.12 : If $\langle E, N\rangle$ is a necessitarian metaphysic, the following hold.

(i) $X \perp Y$ iff $\Omega^{\exists}(X) \cap \Omega^{\exists}(Y) = \varnothing$.

(ii) $X \overset{\forall}{\sim} Y$ iff $\Omega^{\forall}(X) = \Omega^{\forall}(Y)$.

(iii) $X \overset{\exists}{\sim} Y$ iff $\Omega^{\exists}(X) = \Omega^{\exists}(Y)$.

Proof.

(i) Assume that $X \perp Y$, and for contradiction that there is some world ω such that $X \cap \omega \neq \varnothing$ and $Y \cap \omega \neq \varnothing$. Then there are $x \in X$ and $y \in Y$ such that $\{x, y\} \subseteq \omega$. But for all $x \in X$, $y \in Y$, we have that $\{x, y\} \overset{1}{>\!\!\!\Rrightarrow} \varnothing$, and so, by Dilution, we must have that $\omega \overset{1}{>\!\!\!\Rrightarrow} \omega^C$ as well, contrary to assumption. Conversely, let $\Omega^{\exists}(X) \cap \Omega^{\exists}(Y) = \varnothing$, and let $x \in X$ and $y \in Y$ be arbitrary. Since $\{x, y\}$ are in no world, we must have that $\{x, y\} \cup W \overset{1}{>\!\!\!\Rrightarrow} W^C \cup \varnothing$ for any set $W \supseteq \{x, y\}$, but then it follows by Set Cut that $\{x, y\} \overset{1}{>\!\!\!\Rrightarrow} \varnothing$.

(*ii*) Let $X \overset{\forall}{\sim} Y$, and assume that $\omega \in \Omega^{\forall}(X)$. Then $X \subseteq \omega$, and since $X \overset{1}{\rightarrowtail} \{y\}$ for all $y \in Y$, we must have that $\omega \overset{1}{\rightarrowtail} \{y\}$ for all $y \in Y$ as well. But ω is a world, and thus contains all the things it necessitates, so we must have $Y \subseteq \omega$ as well, and thus also $\omega \in \Omega^{\forall}(Y)$. In the other direction, let $\Omega^{\forall}(X) \subseteq \Omega^{\forall}(Y)$. It is only those y that are outside X that we have to be concerned about here. For all worlds ω containing X, we have that $X \cup \omega \overset{1}{\rightarrowtail} \{y\}$, for all $y \in Y$, since y has to be in ω. But from Dilution, it then follows that $X \cup \omega \overset{1}{\rightarrowtail} \omega^{C} \cup \{y\}$, and by World cut* that $X \overset{1}{\rightarrowtail} \{y\}$.

(*iii*) Let $X \overset{\exists}{\sim} Y$ and $\omega \in \Omega^{\exists}(X)$. Since $x \in \omega$, for some $x \in X$, and $\{x\} \overset{1}{\rightarrowtail} Y$, we must have that there is some $y \in \omega$ as well. Conversely, suppose that $\Omega^{\exists}(X) \subseteq \Omega^{\exists}(Y)$, and let $x \in X$. We then have, for all worlds ω that contain $\{x\}$, that $\omega \overset{1}{\rightarrowtail} Y$, and by Dilution that $\omega \overset{1}{\rightarrowtail} \omega^{C} \cup Y$. The theorem follows by World cut*.

□

The following theorem gives some useful properties of $\Omega^{\forall}$ and $\Omega^{\exists}$. Since the proofs consist in straightforward verifications using set theory, we have omitted them.

Theorem 4.13 : $\Omega^{\forall}$ and $\Omega^{\exists}$ fulfil the following, for all $X, \subseteq E$, and any set $\mathcal{X}$ of such subsets:

(*i*) $\Omega^{\forall}(\bigcup \mathcal{X}) = \bigcap_{X \in \mathcal{X}} \Omega^{\forall}(X)$.

(*ii*) $\Omega^{\exists}(\bigcup \mathcal{X}) = \bigcup_{X \in \mathcal{X}} \Omega^{\exists}(X)$.

(*iii*) $\Omega^{\exists}(\{\hat{X}\}) = \Omega^{\forall}(X)$.

(*iv*) $\Omega^{\exists}(\otimes \mathcal{X}) = \bigcap_{X \in \mathcal{X}} \Omega^{\exists}(X)$.

$\Omega^\forall$ and $\Omega^\exists$ are useful for giving succinct definitions of the nondeterministic necessitation relations $\mathrel{\succ\!\!\Rrightarrow}$ and $\mathrel{\rightarrow\!\!\!\!\succ\!\!\Rrightarrow}$:

$$X \mathrel{\succ\!\!\Rrightarrow} Y \text{ iff } \Omega^\forall(X) \subseteq \Omega^\exists(Y)$$
$$X \mathrel{\rightarrow\!\!\!\!\succ\!\!\Rrightarrow} Y \text{ iff } \Omega^\exists(X) \subseteq \Omega^\exists(Y)$$

Generalising this, we want to be able to interpret probabilistic necessitation so as to fulfil

$$X \overset{\pi}{\mathrel{\succ\!\!\Rrightarrow}} Y \text{ iff } P(\Omega^\exists(Y) \mid \Omega^\forall(X)) = \pi$$

whenever $P(\Omega^\forall(X)) > 0$, where P is a probability function on the set of possible worlds, i.e. a function on a σ-algebra of sets that fulfils the three Kolmogorov axioms

(i) $P(\Delta) \geqslant 0$ for all $\Delta \subseteq \Omega$.

(ii) $P(\Omega) = 1$

(iii) If $\Delta_1, \Delta_2, \ldots$ is a countable sequence of non-overlapping subsets of Ω, then
$$P(\bigcup_i \Delta_i) = \sum_i P(\Delta_i)$$

As usual, we define the conditional probability function

$$P(\Delta|\Gamma) \underset{def}{=} \frac{P(\Delta \cap \Gamma)}{P(\Gamma)}$$

whenever $P(\Gamma) > 0$, and leave it undefined otherwise. We can use the Kolmogorov axioms to explicate the connection between probabilistic necessitation and regular probability. The following theorem shows that the intended interpretation of N is possible.

Theorem 4.14 (*Representation of probabilistic necessitation*) : Given a probabilistic metaphysics $\langle E, N\rangle$, with set of worlds Ω, we can define a probability space $\langle \mathfrak{A}, P\rangle$ on Ω where $\mathfrak{A}$ is a σ-algebra over subsets of Ω and P a probability measure on $\mathfrak{A}$, such that $\mathfrak{A}$ is uniquely generated by the sets of the form $\Omega^\forall(X)$ and $\Omega^\exists(X)$, for $X \subseteq E$, and

$$P(\Omega^{\exists}(Y) \mid \Omega^{\forall}(X)) = N(X, Y)$$

whenever $P(\Omega^{\forall}(X)) > 0$.

Proof. We start by showing that the sets of type $\Omega^{\exists}(X)$ and $\Omega^{\forall}(X)$ generate a σ-algebra uniquely. For this, it is sufficient that they are closed under pairwise intersection, and thus make up what is called a *π-system*; we can then use Dynkin's lemma to show that the σ-algebra generated by these sets is uniquely determined (Fremlin, 2000, §136). But closure under pairwise intersection follows from theorem 4.13, since $\Omega^{\exists}(X) \cap \Omega^{\exists}(Y) = \Omega^{\exists}(X \otimes Y)$, and $\Omega^{\forall}(X) \cap \Omega^{\forall}(Y) = \Omega^{\forall}(X \cup Y)$, and it is also easily checked that $\Omega^{\exists}(X) \cap \Omega^{\forall}(Y) = \Omega^{\exists}(X \otimes \{\hat{Y}\})$.

Now, $\Omega^{\forall}(X)$ can be written as $\Omega^{\exists}(\{\hat{X}\})$ whenever $X \neq \varnothing$ and $X \overset{1}{\not\succ\!\!\Rrightarrow} \varnothing$. But if $X \overset{1}{\succ\!\!\Rrightarrow} \varnothing$, we have that $\Omega^{\forall}(X) = \varnothing = \Omega^{\exists}(\varnothing)$, so the only set that is not of the form $\Omega^{\exists}(X)$ is $\Omega^{\forall}(\varnothing) = \Omega$. Use the generated algebra to define a probability measure as

$$P(\Omega^{\exists}(X)) = N(\varnothing, X)$$

$$P(\Omega^{\forall}(\varnothing)) = 1$$

These are well-defined because of the Equivalence axiom. The first Kolmogorov axiom holds trivially since N always takes values in $[0, 1]$. The second holds by definition, since we have taken $P(\Omega^{\forall}(\varnothing)) = 1$. *Pairwise* additivity follows easily from the additivity axiom for probabilistic necessitation, since, as we proved in theorem 4.13, $X \perp Y$ iff $\Omega^{\exists}(X)$ and $\Omega^{\exists}(Y)$ are disjoint. But since P is bounded, its values for unions of countable sequences of disjoint sets are determined by the values of the unions of their finite subsequences.

Finally, we wish to show that whenever $P(\Omega^{\forall}(X)) > 0$,

$$N(X, Y) = \frac{P(\Omega^{\exists}(Y) \cap \Omega^{\forall}(X))}{P(\Omega^{\forall}(X))}$$

Assuming first that $X \neq \varnothing$, we rewrite the right-hand numerator in terms of $\Omega^{\exists}$:

$$\begin{aligned}\frac{P(\Omega^{\exists}(Y) \cap \Omega^{\forall}(X))}{P(\Omega^{\forall}(X))} &= \frac{P(\Omega^{\exists}(Y) \cap \Omega^{\exists}(\{\hat{X}\}))}{P(\Omega^{\exists}(\{\hat{X}\}))} \\ &= \frac{P(\Omega^{\exists}(Y \otimes \{\hat{X}\}))}{P(\Omega^{\exists}(\{\hat{X}\}))} \\ &= \frac{N(\varnothing, Y \otimes \{\hat{X}\})}{N(\varnothing, \{\hat{X}\})} \\ &= \frac{N(\varnothing \cup X, Y)\; N(\varnothing, \{\hat{X}\})}{N(\varnothing, \{\hat{X}\})} \\ &= N(X, Y)\end{aligned}$$

where the next-to-last equality follows from the Conditionalisation axiom. In the case where $X = \varnothing$, we have

$$\frac{P(\Omega^{\exists}(Y) \cap \Omega^{\forall}(\varnothing))}{P(\Omega^{\forall}(\varnothing))} = \frac{P(\Omega^{\exists}(Y) \cap \Omega)}{P(\Omega)} = P(\Omega^{\exists}(Y)) = N(\varnothing, Y)$$

which proves that the definitions we have adopted make everything come out as expected. □

A probabilistic necessitation relation thus determines a probability distribution on the set of possible worlds. Whether the converse holds as well (i.e. whether any probability distribution on a set of possible worlds can be written as a probabilistic necessitation relation on the entities) is an open question. In any case, we do not have the elegant one-to-one correspondence between necessitation relations and sets of possible worlds that we have with nonprobabilistic necessitation, since two different probabilistic necessitation relations can give rise to the same probability distribution on worlds. The reason for this is that the result of conditioning on a null set in standard, Kolmogorovian probability theory is undefined, while it is not so for a probabilistic necessitation relation. To fully capture the richness of this relation's structure, we would have to use a representation in terms of primitive conditional probability instead, such as that of Rényi (1955).

A *model* in a probabilistically necessitarian metaphysics is a world, just as in regular necessitarian metaphysics. More precisely, let the model space $\mathcal{PN}$ have as objects the class of models $\langle E, N\rangle$ where N is a probabilistic necessitation relation such that $N(E, \varnothing) < 1$. This condition works as a consistency requirement, since it guarantees that the entities in a model have a non-zero chance of occurring together.

A *necessitarian metaphysic* will in general have $N(E, \varnothing) = 1$, since some of its possible entities are incompatible. Take, for example, a typical quantum-mechanical experiment in which we measure the spin of a particle. However we measure it, we will get one of the answers "up" or "down", but we will never get *both* of them. Thus any world in which the experiment occurs has a certain chance to also contain an "up" observation, but if it does, we can be sure that it does not also contain a "down" observation in the same experiment.

As morphisms between probabilistically necessitarian models we may take those functions $h : \langle E_1, N_1\rangle \to \langle E_2, N_2\rangle$ for which

$$N_1(X, Y) \leqslant N_2(h[X], h[Y])$$

for all $X, Y \subseteq E_1$. However, since only a fairly small part of this book will deal with probabilistic necessitation, we will not go into what this choice will mean for embeddings and reductions.

Chapter 5
Semantics

This chapter is devoted to the relationship between a theory and its models, which we have called *semantics.* A semantics consists of an assignment of semantic values to interpretations of a theory's claims, where an *interpretation* is a kind of function from a theory to a model. Semantics generally involves both reinterpretation and modality, and we call those semantics that consist in just reinterpretation *Bolzanian*, and those that consist in just modality *Leibnizian.* The most important terms in this section are *soundness* and *completeness.* These concepts are broadened slightly to accommodate many-valued and probabilistic theories as well.

Section 3 discusses historically important kinds of semantics, among which are matrix and Tarskian semantics. We also give a kind of universal semantics, by using a theory's own theory space to make models for it, and show that this is sound and complete.

The rest of the chapter concerns semantics for necessitarian metaphysics. A central class of these is made up by so-called *truthmaker semantics*, which can be seen as a generalisation of the traditional correspondence theory of truth. We give a number of theorems on these, which clarify how truthmaker theory is connected to the more general concept of a claim's truth conditions.

5.1 Tying Theory to Reality

Semantics, as we mentioned in section 1.5, is for us the study of relationships between theories and model spaces, rather than the study of meaning in general. Not just *any* such relationship is, however, of interest for us. What we concentrate on are those relationships relevant to the truth and falsity of claims.

The basic entity in our version of semantics is the *interpretation*, roughly conceived as a method of assigning entities in or parts or features of a model to claims in a theory.[1] Formally, we take it to be associated with every interpretation h a theory $\mathsf{dom}(h)$ called its *domain* and a model $\mathsf{cod}(h)$ called its *codomain.* It thus has the structure of a morphism in a category, although it naturally does not make up any category on its own.

By a *semantics* $\mathcal{S}(A \overset{H}{\mapsto} \mathcal{M})$ for a theory A in the model space $\mathcal{M}$, we shall understand a binary function from a set H of interpretations and the language L_A of A to a set V of *semantic values*, such that $\mathsf{dom}(h) = A$ and $\mathsf{cod}(h) \in \mathcal{M}$, for every $h \in H$. For us, the most important semantic values will be *truth* and *falsity*, and we will call a semantics *bivalent* if it assigns either t or f to all combinations of interpretations and claims.

The idea is that while an interpretation says how claims are mapped onto a model, the semantics interprets the results of these mappings in terms of semantic values such as truth or falsity. For this purpose, it is imperative that when we know an interpretation h, and the model that h interprets the theory in, the assignments of semantic values to claims should follow more or less directly. Unfortunately, I do not quite know how to make this condition entirely precise. An illustration may help.

In the next section, we will discuss Tarskian semantics. According to Tarski's theory of truth, open formulae can be assigned sets of sequences of elements of a domain, and the sets of such sequences as-

[1] In general, if we do not limit ourselves only to claims, but consider parts of claims as well (such as individual words in a language), an interpretation will also associate these with parts of the model. A well-known example here is *reference*, through which singular terms in a language are assigned objects in a model. But as we have disregarded sub-sentence structure here, we will bypass these complications.

signed to complex open and closed formulae can obtained recursively from the assignments to their parts. Where D is a domain (i.e. a thin Tarskian model, as we also called it in chapter 3), it is natural to take an assignment of sets of sequences of objects in D to the open formulae of a language $\mathcal{L}$ to be an interpretation of $\mathcal{L}$ in D. As Tarski proved, all closed formulae will be assigned either the class of *all* sequences of objects in D, or the empty class.

Now, which of these cases should we take to correspond to truth, and which should correspond to falsity? *This* is exactly the problem of determining the semantics, given the interpretations and the models. Its solution is partly dependent on convention. The one Tarski makes is that the sequence S should be in the set assigned to the formula $P(x_1, \ldots, x_n)$ precisely when the n first elements of S, in the order they appear there, *satisfy* this formula. The meaning of his term "satisfaction", together with his material criterion of adequacy on truth definitions, then forces us to say that a sentence is true iff it is satisfied by all sequences, and false iff it is satisfied by none.

We could, on the other hand, just as well have made the opposite convention, and said that an open formula is to be interpreted as the set of sequences that do *not* satisfy it (we may say that these sequences are the "falsifiers" of the formula). We must then say that a sentence is true iff it is assigned the empty set, since truth, on this picture, corresponds to absence of falsifiers rather than the presence of satisfiers. In any case, however, a description (or in some cases a stipulation) of the meanings of the terms involved will settle how semantic values are to be assigned as well. This is roughly what we mean by the condition that knowledge of the semantics should be inferable from knowledge of the interpretations, together with knowledge of the models.

Given a bivalent semantics $\mathcal{S}(A \overset{H}{\mapsto} \mathcal{M})$, we write $h \models p$ when $\mathcal{S}(h, p) = t$, and if X is a set of claims, we write $h \models X$ iff $h \models p$ for all $p \in X$. For a *theory* A, we have that A is true iff its consequence operator preserves truth. This means that we should interpret $h \models A$ as the claim that for all $X \subseteq L_A$, $h \models X \Rightarrow h \models C(X)$. These are our versions of the notion of *truth-in-a-model.*

Each semantics $\mathcal{S}(A \overset{H}{\mapsto} \mathcal{M})$ gives rise to a *semantic* consequence relation $\models_{\mathcal{S}}$ on $\wp(L_A) \times L_A$ through the definition

$$X \models_{\mathcal{S}} p \underset{def}{=} (\forall h \in H)(h \models X \to h \models p)$$

We say that a semantics $\mathcal{S}(A \overset{H}{\mapsto} \mathcal{M})$ is *sound* iff it satisfies the condition

$$\text{if } X \vdash_A p \text{ then } X \models_{\mathcal{S}} p$$

and *complete* iff it satisfies the converse implication

$$\text{if } X \models_{\mathcal{S}} p \text{ then } X \vdash_A p$$

for all sets of claims $X \subseteq L_A$ and all claims $p \in L_A$. A sound and complete semantics is thus, as usual, one in whose theory p is a consequence of the set X of claims iff all interpretations that make the claims in X come out true are such that p comes out true according to them as well. If $\mathcal{S}$ is a sound and complete semantics for a theory A, we say that A's consequence operator is *given* by $\mathcal{S}$, or that A is *characterised* by $\mathcal{S}$.[2]

Related to the concept of soundness is that of validity. Let F be a theory (for instance, a logic) that we use as a framework. A *claim* $p \in L_F$ is *valid* according to $\mathcal{S}(F \overset{H}{\mapsto} \mathcal{M})$ iff $h \models p$ for all interpretations h, and since this property only depends on the semantics (and the model space, but the semantics determines the model space), we write this as $\models_{\mathcal{S}} p$. We count a set X of claims as valid iff $\models_{\mathcal{S}} p$ for all $p \in X$, and a theory A in F as valid iff A's consequence operator is truth-preserving in all interpretations. We write these as $\models_{\mathcal{S}} X$ and $\models_{\mathcal{S}} A$, respectively. In the limiting case, F itself is valid iff $\mathcal{S}$ is sound.

Semantics come in different forms, and there seems to be two fundamentally different ways to interpret what it means to be logically valid. Let us call a semantics $\mathcal{S}(A \overset{H}{\mapsto} \mathcal{M})$ with a set H of interpretations *Bolzanian* if $\mathsf{cod}(h) = \mathsf{cod}(h')$ for all $h, h' \in H$. In such a semantics, all

[2] A note of caution: what we have called *completeness* is sometimes referred to as *strong completeness*. We have omitted the word "strong", since the other, "weak" form of completeness is of little use unless one takes logic to be concerned primarily with logical truth, rather than consequence.

the variation is done through quantifying over interpretations while the model is kept constant. The other extreme is where choosing a model also determines the interpretation, i.e. no *re*interpretation of terms is allowed.[3] Here, consequence corresponds directly to preservation of truth in all models. Since our models are representations of parts or aspects of possible worlds, we call such a semantics *Leibnizian*, although Leibniz himself tended to see logical consequence as dealing with concepts first, and only secondarily with possible worlds (cf. Ishiguro, 1990, p. 48).

If a semantics is *both* Leibnizian and Bolzanian, it determines the semantic values of all claims in its theory on its own, and truth will coincide with validity. Interesting examples are hard to come by. Even if Peano arithmetic, for example, should have an intended semantics in terms of the natural numbers, such a semantics cannot be complete, by Gödel's theorem. Tarskian semantics in general may perhaps best be taken as neither Bolzanian nor Leibnizian, since they work by reinterpreting terms, but also allows the domain of quantification to vary freely. Semantics whose theories' consequence operators are analytic or stronger are all Leibnizian, since the notion of *following in virtue of meaning* naturally requires the meanings of terms to stay constant.

Why be interested in non-Leibnizian semantics? The Tarskian explication of logical consequence points to one reason: we may be interested in what follows from the meanings of *some* words, but not all, such as when we keep the meanings of the logical constants fixed, but allow the nonlogical terms to vary. One could also envisage classing some terms of a theory's language as *physical*, for instance, and then defining the physical consequences of X to be those which follow from X in virtue of the meanings of the physical words.

Another reason for interest in non-Leibnizian semantics could be *semantic vagueness*. Perhaps the meanings of some claims in a theory

[3] One could hold that we need to do more than to require an interpretation to be uniquely determined by the model if we are to rule out reinterpretation of terms. However, in the absence of a theory of meaning, there seems to be no principled way to do this. Consider, for example, the claim "the largest person in this room is over two metres tall". It does not seem to be reasonable to maintain that it must be *the same* thing that makes this claim true in all the models of the room where it is true.

A are not determinate enough for us to be able to assign semantic values to them unequivocally, given a certain model space. We can still use that model space to define consequence for A by letting the set of interpretations in the semantics be the allowed sharpenings of A's claims (this is a type of supervaluationist treatment of vagueness). Each sharpening will then correspond to a theory stronger than A.

Still, the semantics of primary interest for us are the Leibnizian ones: Etchemen dy (1990) even makes the case that *all* logical consequence is purely modal, and has nothing to do with reinterpretation. Even Tarski himself slips into using modal language when giving his famous theory of what it means for a sentence X to be a consequence of the class K of sentences:

> (F) If, in the sentences of the class K and in the sentence X, the constants—apart from purely logical constants—are replaced by any other constants (like signs being everywhere replaced by like signs), and if we denote the class of sentences thus obtained from K by 'K'', and the sentence obtained from X by 'X'', then the sentence X' *must be* true provided only that all the sentences of the class K' are true. (Tarski, 1936, emphasis added).

I believe that we despite this *should* take Tarski's theory of logical consequence to at least be very much in the Bolzanian vein. Two paragraphs later, he states that if (F) were to be sufficient and necessary for consequence, we would have solved all problems pertaining to this concept, since the only possible difficulty would be with the usage of "true", and that had been answered by his own theory of truth. If he really had wanted to attach some modal force to his *must*, he would surely not have said this, since modal terms were seen as no less fraught with difficulty then than they are now.

It is telling that not even Tarski managed to stay clear of using a modality-laden term such as "must be", and it is doubtful that a purely Bolzanian notion of consequence can be materially adequate. Their primary problem is that they tend to make the validity of logics dependent on what exists: Bolzano's original version, for instance (Bolzano, 1837), did not allow the domain of quantification to vary, so sentences such as "there are at least n things" became truths of logic for all n such that n is less than or equal to the actual number of objects there are.

As we shall see in the next section, it is not obvious that Tarski's own semantics is entirely free from this problem either.

Leibnizian semantics will thus be taken as our primary field of interest. For these, there are several useful ways to describe soundness and completeness in terms of relations between theories and models. Since models and interpretations are correlated one-to-one in such a semantics, we will write $\mathfrak{M} \models p$ when $h \models p$ and h is the unique interpretation whose codomain is $\mathfrak{M}$. We will also drop the reference to a set of interpretations when giving the semantics itself, and write $\mathcal{S}(A \mapsto \mathcal{M})$ rather than $\mathcal{S}(A \overset{H}{\mapsto} \mathcal{M})$.

Within a theory space, the truth of a theory in a model is the same thing as the truth in that model of all its theoretical truths:

Lemma 5.1 : If $\mathfrak{M} \models F$ and A is any theory in F, then $\mathfrak{M} \models A$ iff $\mathfrak{M} \models \top_A$.

Proof. What we need to show is that

$$(\forall X \subseteq L_A)(\mathfrak{M} \models X \to \mathfrak{M} \models C_A(X))$$

iff $\mathfrak{M} \models \top_A$. For the left-to-right direction, assume that the l.h.s. holds, and take X to be the empty set. Then we have that $\mathfrak{M} \models C_A(\varnothing)$, but since $C_A(\varnothing) = \top_A$, $\mathfrak{M} \models \top_A$. For the other direction, we assume that $\mathfrak{M} \models \top_A$ and try to show that $(\forall X \subseteq L_A)(\mathfrak{M} \models X \to \mathfrak{M} \models C_A(X))$. So take any X, and assume that $\mathfrak{M} \models X$. Then $\mathfrak{M} \models X \cup \top_A$, and as we have assumed F to be true in $\mathfrak{M}$, we furthermore have that $\mathfrak{M} \models C_F(X \cup \top_A)$. But this is the same as $\mathfrak{M} \models C_A(X)$, which we sought. □

Now, assuming that we have a Leibnizian semantics $\mathcal{S}(F \mapsto \mathcal{M})$, let $[\![p]\!]$ be the set of models in $\mathcal{M}$ in which p is true, let $[\![X]\!]$, where X is a set of claims, be the set of models where every claim in X is true, and when A is a theory in F, let $[\![A]\!]$ be the set of models where A's consequence operator is truth-preserving. Borrowing some terminology from mainstream model theory, we call a subclass $\mathcal{X}$ of $\mathcal{M}$ such that $\mathcal{X} = [\![B]\!]$ for some theory B in A *elementary.*

The theory space $\mathcal{T}_F$ is ordered by F-entailment and $\wp(\mathcal{M})$ is ordered by the subset relation. This entails that the following theorem holds.

Theorem 5.2 : If $\mathcal{S}(F \mapsto \mathcal{M})$ is sound then $[\![\cdot]\!]$ is monotone and if it furthermore is complete, $[\![\cdot]\!]$ is an embedding.

Proof. Assume that A and B are arbitrary theories in F and that $A \preccurlyeq B$. Let $\mathcal{M}$ be an arbitrary model. By the preceding lemma, $\mathfrak{M} \models A$ iff $\mathfrak{M} \models \top_A$ and $\mathfrak{M} \models B$ iff $\mathfrak{M} \models \top_B$. But since $A \preccurlyeq B$, we have that $\top_B \subseteq \top_A$, and thus we must have that $[\![\top_A]\!] \subseteq [\![\top_B]\!]$.

For the completeness part, we need to prove that $[\![A]\!] \subseteq [\![B]\!] \leftrightarrow A \preccurlyeq B$ implies that $(\forall \mathfrak{M} \in \mathcal{M})(\mathfrak{M} \models X \rightarrow \mathfrak{M} \models p) \rightarrow X \vdash_F p$, for all $X \subseteq L_A$ and $p \in L_A$. Take X and p to be arbitrary. Assume that $(\forall \mathfrak{M} \in \mathcal{M})(\mathfrak{M} \models X \rightarrow \mathfrak{M} \models p)$. Since we have assumed soundness, we have that the models in which $C_F(X)$ is true are exactly those in which X is true. This means that the condition is equivalent to $\mathfrak{M} \models C_F(X) \rightarrow \mathfrak{M} \models C_F(\{p\})$. But any closed set of claims in F is the set of truths of a theory in F, so we set $C_F(X) = \top_A$ and $C_F(\{p\}) = \top_B$, and arrive at $(\forall \mathfrak{M} \in \mathcal{M})(\mathfrak{M} \models \top_A \rightarrow \mathfrak{M} \models \top_B)$. By the preceding lemma, this is equivalent to the condition that $[\![A]\!] \subseteq [\![B]\!]$, which by the embedding requirement in turn is equivalent to $A \preccurlyeq B$. But $\top_A = C_F(X)$ and $\top_B = C_F(\{p\})$, which means that $C_F(\{p\}) \subseteq C_F(X)$, and this is equivalent to $X \vdash_F p$. □

It is also enlightening to view the matter from the perspective of a model space's *canonical* theory. Remember that $M = Th(\mathcal{M})$ is the theory $\langle L_M, C_M \rangle$ such that $L_M = \wp(\mathcal{M})$, and

$$C_M(X) = \left\{ p \in L_M \;\middle|\; \bigcap X \subseteq p \right\}$$

The following holds:

Theorem 5.3 : $\mathcal{S}(A \mapsto \mathcal{M})$ is sound and complete iff $[\![\cdot]\!]$ is a translation of A into $Th(\mathcal{M})$.

Proof. A translation is a theory homomorphism $h : A \to Th(\mathcal{M})$ such that $h[C_A(X)] = C_{Th(\mathcal{M})}(h[X]) \cap h[L_A]$ for all $X \subseteq L_A$. By the last theorem, soundness and completeness are equivalent to the condition that $X \vdash_A p$ iff $[\![X]\!] \subseteq [\![p]\!]$. But in $Th(\mathcal{M})$, consequence coincides with set inclusion, and any set of such sets is equivalent to the intersection of them. This means that $[\![X]\!] \subseteq [\![p]\!]$ iff $p \in C_{Th(\mathcal{M})}$, for all $X \subseteq L_A$ and $p \in L_A$, and thus A's consequence relation coincides with that of $Th(\mathcal{M})$, on the image of $[\![\cdot]\!]$. □

Thus, a sound and complete semantics can also be seen as a set of *translations* of one theory into another. This is an interpretation of semantics according to which the subject concerns relations between theories, rather than relations between theories and the world. The realisation that all of semantics can be interpreted this way is mostly due to Sellars (1963), who makes the point in discussing Carnap's *Introduction to Semantics.*

The lesson we should draw here, I believe, is that semantics, and in extension metaphysics, rather than being about some occult connection between language and world, furnishes us with a specific way of looking at theories—of interpreting them. It allows us to take a *metaphysical stance*, to borrow (and slightly mutilate) one of van Fraassen's phrases (van Fraassen, 2002). In taking such a stance, we are able to ask questions like: what makes these claims true? How come this inference is truth-preserving? And most importantly of all: given that this theory is true, what could the world be like?

5.2 Probabilistic and Many-valued Semantics

Section 2.4 introduced two generalisations of the theory concept, viz. many-valued and probabilistic theories. Both of these require further comments on which kinds of semantics are required to capture their consequence operators. Starting out with many-valued theories, it is evident that semantics as we have characterised it (and as it is usually characterised) is concerned solely about *truth.* Since a claim is equivalent to the assignment of truth to it, we could write that p is *true* in $\mathfrak{M}$ under the semantics $\mathcal{S}$ as

$$\mathfrak{M} \models_{\mathcal{S}} t : p$$

This invites us to use a similar way of assigning other semantic values. In general, whenever $\mathcal{S}(A \overset{H}{\mapsto} \mathcal{M})$ assigns p the value v in the model $\mathfrak{M}$, we write

$$\mathfrak{M} \models_{\mathcal{S}} v : p$$

We also write $\mathfrak{M} \models_{\mathcal{S}} v : X$, where X is a set of claims, for the assignment of v to each and every one of these in $\mathfrak{M}$. Let a *many-valued semantics* be a semantics $\mathcal{S}(A \overset{H}{\mapsto} \mathcal{M})$ where A is a many-valued theory, and the set of semantic values V that $\mathcal{S}$ assigns combinations of interpretations and claims is the same as set of semantic values of A. If $\mathcal{S}$ is Leibnizian and $\mathfrak{M} \models_{\mathcal{S}} v : p$, we say that $\mathfrak{M}$ *satisfies* the assignment $v : p$. The important concepts of soundness and completeness generalise automatically to the many-valued case. Soundness means that if $\boldsymbol{X} \vdash v : p$, then any model that satisfies all assignments in $\boldsymbol{X}$ will satisfy $v : p$ as well, and completeness that if all models that satisfy $\boldsymbol{X}$ also satisfy $v : p$, then $\boldsymbol{X} \vdash v : p$.

Not only are many-valued semantics necessary for a proper understanding of traditional many-valued logic, but regular truth-centered single-valued semantics is also unable to fully capture even two-valued theories. Say that a model space $\mathcal{M}$ is *appropriate* for A iff the subject matter of A is a model $\mathfrak{A}$ in $\mathcal{M}$, and call $\mathfrak{A}$ the *actual model* of A's subject. Furthermore, if $\mathcal{M}$ is a model space appropriate for A, and $\mathcal{S}(A \mapsto \mathcal{M})$ is a Leibnizian semantics for A in $\mathcal{M}$, we call $\mathcal{S}$ appropriate iff $\mathfrak{A} \models_{\mathcal{S}} p$ iff p is actually true, for all $p \in L_A$.

An appropriate semantics is one that "gets actual truth right". But this property is not very accessible. How can we identify the model $\mathfrak{A}$, apart from the characterisation of it as the model in which all the claims in A that actually are true, are true, and no others? How can we tell if $\mathfrak{A} \models_{\mathcal{S}} p$ for all actual truths p, except by asking whether there is *some* model $\mathfrak{M}$ for which this holds? It is not as if we have any independent access to models, apart from through our theories.[4]

This means that the notions of soundness and completeness are more practically useful than appropriateness. A sound and complete semantics for A gets both the theoretical truths of A and the inference connections postulated by A right, and it does this *without* us having to identify which specific model $\mathfrak{M}$ is the actual one. In *many-valued* semantics, completeness furthermore allows us to derive something that is very close to appropriateness. As we said, one of the problems with this concept is that there seems to be no independent way to decide if the actual model is in a model space or not. The most we really can ask for is that there should be *some* model $\mathfrak{M}$, such that $\mathfrak{M} \vDash_{\mathcal{S}} p$ iff p is actually true, for all $p \in L_A$. Such a model is one that *could* be the actual model for all we can know, since there is no semantic way to discriminate between it and the actual one. The following theorem shows that this property follows from completeness, so long as the semantics in question is two-valued, and the theory is consistent with actual truth.

Theorem 5.4 : Let $\mathcal{S}(A \mapsto \mathcal{M})$ is a complete Leibnizian bivalent semantics and A a bivalent theory. Let $true_A$ be the set of claims in L_A that are actually true, let $false_A$ be $L_A \backslash true_A$, and assume that

$$t : false_A \cap C_A(t : true_A \cup f : false_A) = \varnothing$$

so that A does not allow us to infer the truth of any false claim from the assignment of t to all actual truths, and f to all actual non-truths. Then there is some $\mathfrak{M} \in \mathcal{M}$ such that $\mathfrak{M} \vDash_{\mathcal{S}} t : p$ iff p is true, for all $p \in L_A$.

Proof. Because of the assumed consistency with actual truth, we have an assignment $f : p$ such that $t : true_A \cup f : false_A \not\vdash_A t : p$, for any $p \in false_A$. By completeness, there must then be some model $\mathfrak{M}$ that satisfies the set $t : true_A \cup f : false_A$ of assignments such that

[4] A different way of expressing this point is as the slogan *theory precedes metaphysics.* This is a principle that I believe no philosopher who calls herself a naturalist should deny.

$\mathfrak{M} \not\models t : false_A$. Any such model makes true all actual truths, and no others. □

The many-valuedness of the semantics is necessary here. To see why, assume that $A = \langle L_A, C_A \rangle$ is a theory in which

$$\begin{aligned} L_A = \{p_1 &= \text{snow is white}, \\ p_1 &= \text{grass is red}, \\ p_3 &= \text{violets are black}\} \end{aligned}$$

and $C(X) = X$ for all $X \subseteq L_A$, since neither of p_1, p_2 or p_3 follow from any of the others. Let $\mathcal{M}$ consist of all triples $\langle v_1, v_2, v_3 \rangle$ of truth-values *except* $\langle t, f, f \rangle$, and let $\mathcal{S}(A \mapsto \mathcal{M})$ assign the claim p_i the value v_i any model. This means that, for any model $\mathfrak{M} = \langle v_1, v_2, v_3 \rangle$, we have that $\mathfrak{M} \models p_i$ iff $v_i = t$.

$\mathcal{S}$ is sound (trivially) and also complete. Whenever $\mathfrak{M} \models X \Rightarrow \mathfrak{M} \models p_i$, we have that $X \vdash_A p_i$, since the only cases in which $\mathfrak{M} \models X$ are those in which $\mathfrak{M} \models p_i$ for all $p_i \in X$. But $true_A = \{p_1\}$, and thus there is no model in which *only* the actually true claims in L_A are true: in both $\langle t, t, f \rangle$ and $\langle t, f, t \rangle$, something else is true as well.

The reason for this is the lopsidedness of regular consequence. If A is a true theory, so that its consequence operator preserves actual truth, then a complete semantics must have some model in which all the actual truths are true. But single-valued semantics cannot guarantee the existence of a model where all actual falsehoods are untrue.

The way this is usually handled it through conventional stipulation. If the theory A contains at least one actually true and one actually false claim in its language, satisfies *Ex Falso Quodlibet*, and we furthermore require that the set of actual truths has to be *maximal* in A, so that no claim could be added to it without formal inconsistency ensuing, we can avoid speaking about assignments for sound and complete semantics. *If* the theory is standard, then not every claim in its language can be true. Thus, there is no model $\mathfrak{M}$ such that $\mathfrak{M} \models L_A$, but by completeness, there *is* a model $\mathfrak{M} \models true_A$. But nothing *outside* $true_A$ can be true in this model either, since we then could draw the conclusion that *all* claims would be true in $\mathfrak{M}$, and we have already assumed that there is no such model.

However, the requirement that $true_A$ should be maximal in *any* theory A is naturally not possible unless we limit the range of sets of claims that can constitute A's language severely. Furthermore, maximalness is not the natural choice for many theories. Consider a theory M of intuitionistic mathematics, for instance. Here, a statement p is true iff we have an effective way of obtaining a proof of p. Interpreting *false*, again, as *not-true*, p is false if we do not have such a way. But here, there is no reason that $true_M$ should be maximal: that holds only in the very special case where all questions in M have been settled.[5]

Thus many-valued semantics provide a genuine generalisation of the single-valued kind. Another generalisation is connected with probabilistic consequence. Our intention here is to be able to read

$$X \vdash^{\pi} p$$

as "whenever the claims in X are true, there is a chance π of p being true". For this purpose, let a *probabilistic semantics* $\mathcal{S}(A \mapsto \Sigma)$, where $A = \langle \mathfrak{S}_A, Ev_A \rangle$ is a probabilistic theory and Σ is a probability space $\langle H_\Sigma, \mathfrak{S}_\Sigma, P_\Sigma \rangle$, be a function from $L_A \times H_\Sigma$ to a set V of semantic values. In Σ, H_Σ is a set of interpretations of A in models of a model space $\mathcal{M}_\Sigma$, $\mathfrak{S}_\Sigma$ is a σ-algebra of subsets of H, and P_Σ is a probability measure on $\mathfrak{S}_\Sigma$.

The intended interpretation of these concepts is that $P_\Sigma(X)$, where X is in H, is the probability that the *correct* interpretation of A is one of those in H. In the Leibnizian case, we can also use the concept of a *probabilistic model space* $\Sigma = \langle \mathcal{M}_\Sigma, \mathfrak{S}_\Sigma, P_\Sigma \rangle$, where $\mathfrak{S}_\Sigma$ is defined directly on the model space $\mathcal{M}_\Sigma$, and $P_\Sigma(\mathcal{X})$ is interpretable as the *proportion* of all models of $\mathcal{M}_\Sigma$ that are in $\mathcal{X}$.

As before, we concentrate on Leibnizian semantics. Call a such a semantics $\mathcal{S}(A \mapsto \Sigma)$ *structurally sufficient* if $[\![B]\!]$ is in $\mathfrak{S}_\Sigma$, for all $B \in \mathfrak{S}_A$. A probabilistic semantics which is *not* structurally sufficient

[5] The proper condition on the set of truths in an intuitionistic logic seems to be that it should be a *prime filter*, i.e. a set closed under consequence, such that $p \vee q \in true$ entails that $p \in true$ or $q \in true$. For classical, Boolean logic, prime filters coincide with maximal filters. However, they come apart for weaker logics.

will be unable to attach probabilities to some closed sets of claims in its theory, and thus be unfit for use. We say that a structurally sufficient Leibnizian semantics $\mathcal{S}(A \mapsto \Sigma)$ is *sound* iff

$$p \in C_A^\pi(X) \;\Rightarrow\; P(\llbracket p \rrbracket \mid \llbracket X \rrbracket) = \pi$$

and *complete* iff

$$P(\llbracket p \rrbracket \mid \llbracket X \rrbracket) = \pi \;\Rightarrow\; p \in C_A^\pi(X)$$

for all $X \subseteq L_A$ such that $P(\llbracket X \rrbracket) > 0$. Interestingly, for Leibnizian probabilistic semantics, completeness *implies* soundness. Since P is assumed to be defined on all subsets of $\mathcal{M}_\Sigma$ that are in the image of $\llbracket \cdot \rrbracket$, and $X \vdash_A^\pi p$ and $X \vdash_A^{\pi'} p$ implies that $\pi = \pi'$, the probabilistic structure of Σ must determine that of A. This, in turn, guarantees that this structure must conform to that of the probabilistic metaphysics.

Leibnizian probabilistic semantics give us a kind of interpretation of probability which is neither strictly frequentistic nor subjective. In a way, it could be described as a *modal* frequency interpretation, since it gives us that $P(\llbracket p \rrbracket \mid \llbracket X \rrbracket)$ is the frequency of p-models among the X-models. It is, however, not an *actualist* frequency interpretation, since it involves more models than the actual one. In this, it is similar to an *hypothetical* frequency interpretation, on which $P(Y|X)$ is the proportion of X's that *would* be Y's, given that the X's go on forever. An example of an interpretation of this type is von Mises's, according to which $P(X)$ is the limiting relative frequency of X's in a given collective (von Mises, 1981).

5.3 Varieties of Semantics

As we described it in section 3.4, the space $\mathcal{Th}_A$ has as models the theories in A. For each such theory B, let h_B be a unique interpretation.[6] Let the *theory space semantics* for A be the function that maps any interpretation h and claim p to *truth* iff $p \in \top_{\mathsf{cod}(h)}$. A theory space semantics is always Leibnizian. It is also sound and complete, as the following theorem shows.

Theorem 5.5 : For any theory A, A's theory space semantics is sound and complete.

Proof. What we need to show is that

$$p \in C_A(X) \text{ iff } (\forall M \in \mathcal{Th}_A)(X \subseteq \top_M \to p \in \top_M)$$

But the set of truths of the theories in A are exactly the sets of claims that are closed under C_A, since theories correspond one-to-one with closed sets in L_A. Thus we only need to show that $p \in C_A(X)$ iff $(\forall Y \in \mathcal{CS}(A))(X \subseteq \top_Y \to p \in \top_Y 7)$, where $\mathcal{CS}(A)$, as before, is the set of subsets of L_A closed under C_A. This, in turn, follows from the fact that C_A is a closure operator, and that every closure operator is interdefinable with its set of closed subsets this way. □

Theory space semantics are thus ubiquitous, but they also afford little enlightenment beyond what is given by the theory itself. Since we have imposed no restrictions on the structure of the theory, dependence on meaning can never be ruled out, and this is why theory space semantics must be Leibnizian. Adding one such restriction—that A must be formalisable—allows us to employ matrix models instead, and matrix semantics for connecting A with these.

[6] It does not really matter what the interpretation *is* here, but to make the discussion more intuitive, we can take it to be an identity function from L_A to L_B, to highlight the fact that we do not allow reinterpretation of claims. Although this makes the interpretations identical as set-theoretical functions (since all theories in A have the same language), they are still not identical as interpretations, since their codomains differ.

Recall that the space $\mathcal{M}t$ of matrix models contains as objects pairs $\mathfrak{M} = \langle \mathfrak{A}_{\mathfrak{M}}, D_{\mathfrak{M}} \rangle$, where $\mathfrak{A}_{\mathfrak{M}}$ is an algebra, and $D_{\mathfrak{M}}$ is a subset of the carrier of $\mathfrak{A}_{\mathfrak{M}}$ called the *designated value set.* Let an interpretation h of A in a matrix $\mathfrak{M}$ be a homomorphism from $\mathfrak{A}$ to $\mathfrak{A}_{\mathfrak{M}}$, where $\mathfrak{A}$ is the algebra that A is formalised by. Let the *matrix models of* A be the subcategory $\mathcal{M}t_A$ of $\mathcal{M}t$ containing those matrices which have algebras of the same signature as $\mathfrak{A}$, and for which $h^{-1}[D_{\mathfrak{M}}]$ is closed under C_A for all interpretations $h : A \to \mathfrak{M}$.

The *matrix semantics* $\mathcal{S}(A \mapsto \mathcal{M}t_A)$ for A is the function that assigns h, p the value *truth* iff $h(p) \in D_{\mathsf{cod}(h)}$, i.e. iff h takes p to a value that is designated in the model it interprets A in. Since the semantics postulates several interpretations for each model, and also several models, it is neither Bolzanian nor Leibnizian. The soundness and completeness of such semantics are well-known from the algebraic logic literature; a simple proof is given below.

Theorem 5.6 : If A is formalisable, then A's matrix semantics is sound and complete.

Proof. The formalisability of A means that there is an algebra $\mathfrak{A}$ on L_A such that C_A commutes with all endomorphisms on this algebra. Let $X \nvdash_A p$. Then there is a closed set of claims D, such that $X \subseteq D$, but $p \notin D$. $\langle \mathfrak{A}, D \rangle$ is then a matrix model of X which is not a model of p, under the identity interpretation. For the other direction, assume that $X \subseteq D$ but $p \notin D$ for some matrix model $\mathfrak{M} = \langle \mathfrak{A}_{\mathfrak{M}}, D \rangle$ of A, and let h be any interpretation of A in $\mathfrak{M}$. Then $h^{-1}[D]$ is a closed set of claims in L_A which, by assumption, contains $h^{-1}[D]$ but not $h^{-1}[\{p\}]$. But this means that we must have $X \nvdash_A p$. □

The requirement that A must be formalisable, and in particular the structurality condition, is necessary for the proof to go through. This is what allows atomic claims to take on the meanings of any others, and thus also what makes it possible to allow the reinterpretation of terms in a non-Leibnizian semantics.

Both theory space semantics and matrix semantics can be seen as generalisations of the truth-table semantics for classical logic. More generality is afforded by Tarskian semantics, which usually are associated with first-order predicate logic. Let us first discuss semantics whose domain is the full Tarskian model space $\mathcal{T}_{\mathcal{L}}$ for a first-order language $\mathcal{L} = \langle L, f_1, \dots, f_n, P_1, \dots, P_m \rangle$. The models of $\mathcal{T}_{\mathcal{L}}$ are first-order structures $\mathfrak{M}$ of the same signature as $\mathcal{L}$.

Now, L is not the language of any theory, since most elements of L are open formulae, and thus incapable of being true or false. But it is only the sentence-part L^{sent} of L that is of interest for questions of soundness and completeness, although we need the open formulae to define truth recursively. Let an *assignment* s in the model $\mathfrak{M}$ be a function from the variables of $\mathcal{L}$ into $\mathfrak{M}$'s domain D. Extend each such assignment to an assignment s' from the terms of $\mathcal{L}$ to the elements of M, such that $s'(\xi) = s(\xi)$ for each variable ξ and $s'(f_i(\tau_1, \dots, \tau_n)) = g_i(s'(\tau_1), \dots, s'(\tau_n))$ for all terms $\tau_1, \dots, \tau_n$, where f_i is the i:th function symbol of $\mathcal{L}$, g_i is the i:th function of $\mathfrak{M}$, and n is the arity of f_i (and g_i).

Where s' is such an extended assignment, write $s'[a/\xi]$ where ξ is a variable and a is an element of D for the extended assignment that is exactly like s' except for assigning a to ξ. For each extended assignment s', define *satisfaction* under s' to be a relation $\models_{s'}$ on $\mathcal{M}_L \times L$ that fulfils the following conditions for all formulae $\varphi, \psi \in L$:

$\mathfrak{M} \models_{s'} P_i(\tau_1, \dots, \tau_n)$ iff $\langle s'(\tau_1), \dots, s'(\tau_n) \rangle \in Q_i$, where P_i is the i:th predicate of $\mathcal{L}$, Q_i is the i:th relation in $\mathfrak{M}$, and n is the arity of P_i and Q_i.

$\mathfrak{M} \models_{s'} \tau_1 = \tau_2$ iff $s'(\tau_1) = s'(\tau_2)$.

$\mathfrak{M} \models_{s'} \neg\varphi$ iff $\mathfrak{M} \not\models_{s'} \varphi$.

$\mathfrak{M} \models_{s'} \varphi \wedge \psi$ iff $\mathfrak{M} \models_{s'} \varphi$ and $\mathfrak{M} \models_{s'} \psi$.

$\mathfrak{M} \models_{s'} (\forall\xi)\varphi$ iff $\mathfrak{M} \models_{s'[a/\xi]} \varphi$ for all $a \in D$.

For each $\mathfrak{M} \in \mathcal{T}_{\mathcal{L}}$, we define an interpretation h to be a function from sentences in S to the set of assignments in $\mathfrak{M}$ that satisfy them, and we define the Tarskian semantics for $\mathcal{L}$ to be the function that maps each such interpretation–sentence pair h, p to *true* iff $h(p)$ is the set of all assignments on $\mathfrak{M}$.

Since we have only one interpretation per model, this Tarskian semantics is Leibnizian. This may seem surprising, since Tarskian semantics regularly is taken to involve reinterpretation of terms. But we have already noted that the full notion of Tarskian model has some unexpected features, such as nonextensionality. Let us, for comparison, see what semantics on the space $\mathcal{V}$ of thin Tarskian models may be like.

Since a model in $\mathcal{V}$ is just a set, we do not have to relativise thin Tarskian models to language signatures, unlike what we have to with their full versions. The *semantics* naturally still needs to be relativised, though. Let an *extension specification* for a first-order language $\mathcal{L}$ in a thin model M be a function *ext* from the predicates and function symbols of $\mathcal{L}$ to sets of tuples of elements of M such that every n-place predicate P_i is taken to a set of n-tuples $ext(P_i)$ and every n-ary function symbol f_i is taken to an n-ary function on M.

For any extension specification *ext* in the model M, let an assignment on *ext* be a function s_{ext} from the terms of $\mathcal{L}$ to the elements of M such that $s(f_i(\tau_1, \ldots, \tau_n)) = ext(f_i)(\tau_1, \ldots, \tau_n)$ for all terms $\tau_1, \ldots, \tau_n$. Let the assignment s_{ext} satisfy the atomic formula $P_i(\tau_1, \ldots, \tau_n)$ iff $\langle s_{ext}(\tau_1), \ldots, s_{ext}(\tau_n)\rangle \in ext(P_i)$, and define the recursive clauses for $\neg$, $\wedge$ and $\forall$ as before.

Now, for each extension specification *ext*, define a unique interpretation h from the formulae in L to the sets of assignments on *ext* that satisfy them. Let the *thin semantics* for $\mathcal{L}$ be the function from the set of all these interpretations, and the sentences in the subset L^{sent} of L, that takes the value *true* iff $h(p)$ is the set of all assignments on its extension specification *ext*.

This version of Tarskian semantics is neither Bolzanian nor Leibnizian, and it may probably be held to lie fairly close to Tarski's intentions. Predicates are interpreted as relations on M, and function symbols as functions on M. Both this and the full version of Tarskian semantics are sound and complete as semantics for first-order logic.

There is one further modification we can make, however, which may take us even closer to what Tarski could have meant. Let the *universal model* be the class $\mathfrak{V}$ of everything that actually exists.[7] Define an ex-

[7] We need to assume here, of course, that the concept of a "class of everything

tension specification for $\mathcal{L}$ to be a function from the predicates, function symbols *and the universal quantifier* in $\mathcal{L}$ to subsets of $\mathfrak{V}$, such that the universal quantifier is taken to a set $D \in \mathfrak{V}$ called the extension specification's domain, and the predicates and function symbols are taken to relations and functions on D.

We define assignments and satisfaction as before and associate each extension specification with an interpretation in terms of assignments on that extension specification. The *minimal semantics* is the function that takes the value *true* for the interpretation h and the sentence p iff $h(p)$ is the set of all assignments on h's extension specification. Minimal semantics is fully Bolzanian: the model is always the same (viz. the universal class) and only the interpretation of the nonlogical constants and the universal quantifier varies.

An advantage of minimal semantics is that it seemingly does not require us to talk about non-existent entities, since everything that is in a model actually must exist. On the other hand, it *does* require us to have a *class of everything*, and since this class, on pain of contradiction, cannot itself exist, it is not obvious that we have got rid of *all* reference to non-existents. At the very least, we need a theory for how we are to avoid reference to $\mathfrak{V}$, and given such a theory, we may ask why we cannot use it to handle other non-existents as well.

The avoidance of talk about non-actual (i.e. nonexistent) things also comes at a steep price. We have already mentioned the problems that Bolzano's own definition of logical consequence runs into, which make logical validity become dependent on what actually exists. How does minimal Tarskian semantics avoid that problem? How does Tarski himself avoid it, if his own intention was that his semantics should be minimal in this way?

The truth is that minimal semantics works because of a combination of the strength of classical Platonistic set theory and the weakness of first-order logic. To begin with, the only logical predicate in FOL is identity, so the reason that $(\exists x)\,Unicorn(x)$ does not come out as logically false is that there is an interpretation of the predicate $Unicorn(x)$ that takes it to makes it mean the same thing as our word "porcupine", for example. In short, the only thing that we can talk about in FOL is

that exists" is consistent. $\mathfrak{V}$ thus cannot itself be something that exists.

set-theoretic structure. But here the strength of mathematical Platonism comes into play: any possible set-theoretic structure actually exists! So anything that FOL has the resources to say anything about has its existence guaranteed by the existence of sets. Allowing the domain to vary among subsets of this class then stops $(\exists^{!n} x)(x = x)$ from being logically false for any $n \in \mathbb{N}$.[8]

It seems clear that if we are to investigate questions such as the ontological commitments of set theory, we cannot avail ourselves of a semantics such as the minimal one, which presupposes the existence of sets if it is to work. It is also the case that if we want to study the metaphysics of other theories, with material or physical consequence operators, the decision to treat identity alone as having a fixed meaning cannot be maintained. One way or another, we will have to talk about things that do not exist but might have. Whether this involves us in any commitments to *possibilia* is itself a question of semantics. In fact, it is only in a certain semantics in which descriptions or names work by *referring* that it does so. But we do not have to interpret them this way, as Quine showed by shaving off Plato's beard in *On what there is.*[9]. Another type of semantics in which talking about X's does not automatically incur any commitment to them is the one described in the remainder of this chapter.

[8] It is interesting to note that Tarski criticises Carnap's definition of logical consequence in *The Logical Syntax of Language* as too dependent on peculiarities and limitations of one's formal language (Tarski, 1936). But if minimal semantics captures his own intentions, he is himself vulnerable to the criticism that for him, logical consequence becomes hostage to questions of ontology, and in particular to the existence of sets. Since Tarski himself most certainly was *never* a Platonist (Feferman and Feferman, 2004, p. 52), it is not clear to me that one should attribute the minimal interpretation to him either.

[9] It is unfortunately a common belief among philosophers that Kripke showed this approach to names to be untenable. This is far from the case, however. Kripke's arguments are fallacious, as was ably explained by Dummett in the second edition of *Frege: Philosophy of Language* (Dummett, 1981, pp. 112–146)

5.4 Necessitarian Semantics

As we have mentioned, necessitarian metaphysics allow us to study the connection between theory and reality in detail. Let $\mathcal{S}(A \mapsto \mathcal{M})$ be a bivalent Leibnizian semantics from the theory A to a necessitarian metaphysics $\mathcal{M}$. We showed in section 5.1 that if $\mathcal{S}$ is sound, then $X \vdash_A p \Rightarrow [\![X]\!] \subseteq [\![p]\!]$, and if $\mathcal{S}$ is complete, then $[\![X]\!] \subseteq [\![p]\!] \Rightarrow X \vdash_A p$.

Since models, in a necessitarian metaphysics, correspond one-to-one to worlds (which are sets of entities), we will simplify the discussion slightly by using $[\![p]\!]$ to refer to not only the set of models but also the set of worlds in which p is true. We will also use the double turnstile notation for worlds, and so we write $\omega \vDash p$ if $\mathfrak{M}_\omega \vDash p$, where $\mathfrak{M}_\omega$ is the model in $\mathcal{M}$ whose set of existent entities is ω.

For now, let us wait with discussing what interpretations in a necessitarian model might be. It will turn out that the details of these are fairly unimportant for us to be able to study the structure of necessitarian semantics. On the other hand, assumptions on the behaviour of the $[\![\ \cdot\ ]\!]$ operator do allow us to derive more about this structure.

For any claim p in A, we say that p is *positive monotonous* (or just *positive*) under the necessitarian semantics $\mathcal{S}$ iff $\omega \in [\![p]\!]$ and $\omega \subseteq \omega'$ imply that $\omega' \in [\![p]\!]$. We say that p is *negative monotonous* (or *negative*) under $\mathcal{S}$ iff $\omega \in [\![p]\!]$ and $\omega' \subseteq \omega$ imply $\omega' \in [\![p]\!]$. We call the semantics $\mathcal{S}$ itself positive iff every claim in its theory is positive, and negative iff every claim in its theory is negative.

A positive claim is one that, if true in a world, remains true in any world containing that world. An example is a claim of existence: if p holds the entity a to exist, and p is true in a world ω, that must be because a exists in ω. But then a must remain true in any world larger than ω, since these also must contain a. A negative claim true in ω remains true in any world *smaller* than ω, and examples of such claims are claims of non-existence.

Given that most theories allow us to say both that certain things exist, and that certain things do not exist, why would a semantics hold *all* claims to be positive, or to be negative? For the positivity case, the main motivation flows from the idea that truth is *grounded* in the world. Or, in the words of Dummett, which we already have quoted:

> If a statement is true, there must be something in virtue of which it is true. (Dummett, 1976, p. 52).

Ultimately, the principle can be traced back to Leibniz's *principle of sufficient reason*,

> [...] by virtue of which we observe that there can be found no fact that is true or existent, or any true proposition, without there being a sufficient reason for its being so and not otherwise, although we cannot know these reasons in most cases. (Leibniz, 1714, §32)

Positivity of the semantics follows from interpreting the *being* in any of these as the *being* of existence: it requires, for any claim to be true, that there exists something that makes it true. The criterion itself is however somewhat weaker. Call p a claim of *singular existence* iff $[\![p]\!]$ is the set of all worlds containing a given entity a. It is then evident, by the truth-conditions for necessitarian semantics given above, that $\omega \vDash p$ is true iff $a \in \omega$. Likewise, we call p a claim of *singular nonexistence* iff $[\![p]\!]$ is the set of all worlds *not* containing a given entity a.

As we mentioned, claims of singular existence are positive, and claims of singular nonexistence are negative, and as we also mentioned, not all positive claims are singular existence claims. The exact conditions that a positive or a negative claim lays on what exists are captured by the following theorems.

Theorem 5.7 : A claim p is positive iff there is a set $VP\ (p)$ of sets of sets of entities, such that p is true in ω iff $S \subseteq \omega$, for some $S \in VP\ (p)$.[10]

Proof. For the right-to-left direction, assume $VP\ (p)$ to be such a set of possible entities. Assume p to be true in ω. Then there is a set $S \in VP\ (p)$ of entities such that all of these are in ω. But any other world ω' such that $\omega \subseteq \omega'$ must then also contain all of S, and thus p is true in ω' as well. Thus p is positive.

Now, assume that p is positive. We are then free to take $VP\ (p) = [\![p]\!]$. By the truth-conditions for claims under necessitarian semantics,

[10] The reason for the notation "$VP\ (p)$" as well as the notation "$FP\ (p)$" of the next theorem will become clear in the next section.

$\omega \models p$ iff $\omega \in VP\,(p)$. Again, assume p to be true in ω and assume that $\omega \subseteq \omega'$. Then $\omega \in VP\,(p)$, and since $\omega \subseteq \omega$, there is some $S \in VP\,(p)$ such that $S \subseteq \omega$. Conversely, take there to be some set $S \in VP\,(p)$ such that $S \subseteq \omega$. Because $VP\,(p) = [\![p]\!]$, S must be a world, and because of p's positivity, if $S \in [\![p]\!]$, then $\omega \in [\![p]\!]$, so p is true in ω. □

Theorem 5.8 : A claim p is negative iff there is a set $FP\,(p)$ of sets of sets of entities, such that p is true in ω iff $S \cap \omega = \varnothing$, for some $S \in FP\,(p)$.

Proof. For the right-to-left direction, assume $FP\,(p)$ to be such a set of possible entities. Assume p to be true in ω. Then there is a set $S \in FP\,(p)$ of entities that do not overlap ω. But any other world ω', such that $\omega' \subseteq \omega$, cannot overlap S either, and thus p is true in ω' as well. Thus p is negative.

Assume that p is negative. We can then take $FP\,(p)$ to be the set of complements (relative to E) of the sets (i.e. the worlds) in $[\![p]\!]$. By the truth-conditions for claims under necessitarian semantics, p is true in ω iff $\omega \in [\![p]\!]$. Assume that $\omega \models p$. Then $\omega^C \in FP\,(p)$, and since $\omega \subseteq \omega$, there is some $S \in FP\,(p)$ such that $S \cap \omega = \varnothing$. Conversely, take there to be some set $S \in FP\,(p)$ such that $S \cap \omega = \varnothing$. Then S^C is a world in which p is true, and because $\omega \subseteq S^C$ and p is negative, p is true in ω as well. □

Thus, a positive claim is one that can be written as a (possibly infinite) disjunction of (possibly infinite) conjunctions of singular existence-claims, and a negative claim is one that can be written as a (possibly infinite) disjunction of (possibly infinite) conjunctions of singular claims of non-existence. The statement of positivity thus allows that there does not have to be any *unique* thing that makes p true, and it also allows that although no single thing may make p true, there can be several things that jointly do. A case in point would be the claim "there are at least three apples on that tree". Even if the tree contains, say, one hundred apples, any three of these suffice to make the claim true.

Negative semantics may seem harder to motivate than the positive kind, and indeed few of the semantics we shall investigate will be negative. One reason for adopting one could be obtained if we view claims as true by *default.* This means that if p is true, there does not have to be anything to make it true, but if p is false, that is because a *counterexample* or *falsifier* of p exists.

There is also the case of semantics that are both positive and negative. If the set E of all possible entities is a world (i.e. if the necessaritan metaphysics is only inessentially possibilist), such a semantics will give all claims the same truth-conditions through the inference

$$\omega \in [\![p]\!] \Rightarrow E \in [\![p]\!] \Rightarrow \omega' \in [\![p]\!]$$

for any worlds ω, ω', and this results in triviality. If not, there can still be useful semantics: a case in point is what we will refer to as *dichotomous* semantics later on.

Dummett's principle may be seen as central to all correspondence theories of truth, and these come in many varieties. Some (like Russell's version) depend on entities such as facts, while some (like Tarski's) do not. The following are some of the treatments of the concept that may be found in the literature:

Discrete world semantics. Assume that the operator $[\![\cdot]\!]$ is surjective on the set Ω of worlds (i.e. that there are no worlds in which *no* claim in the theory's language isn't true). We say that $\mathcal{S}$ is a *discrete world semantics* iff for any distinct worlds ω_1 and ω_2 in Ω, $\omega_1 \cap \omega_2 = \varnothing$. A special case is where all worlds in Ω are singletons; we may then call $\mathcal{S}$ an *atomic* world semantics.

Traditional possible world semantics, as it is used for modal logic, is atomic. We do not generally talk about what the worlds *are* in a relational model; sometimes they are just called *nodes* or *points.* This does not, however, mean that they do not have any internal structure, but just that any internal structure they may have is immaterial to relational semantics. Since the demise of monism at the beginning of the 20th century, few have denied that the actual world contains more than one thing.

Why would one want one's semantics to treat worlds as discrete? One reason, due to David Lewis (1986, ch. 4), seems to boil down to a wish for extensionality. If something, a, is in both worlds ω_1 and ω_2, any intrinsic property possessed by a in ω_1 must be possessed in ω_2 as well, since it is the same a, *with all its intrinsic properties*, that is in both worlds. This, however, means that a must have the same intrinsic properties in every world.

We have already bit that bullet: since what entities exist determines what world is actual, all of an entity's intrinsic properties *are* necessary. Indeed, Lewis bites it too, since on his counterpart theory, *a itself* cannot have different properties in different worlds. Both the present theory and Lewis's give versions of how we can talk about things whose intrinsics are non-essential: on our theory, by the use of a Carnapian individual concept or a trope reduction, and on Lewis's by use of the counterpart relation.

When we recognise that if we want to approach the matter fully extensionally, we *must* work with entities whose intrinsic properties are essential, we are returned to the question of why the very same (intrinsically-essential) entity cannot be in several possible worlds. Unfortunately, Lewis provides no answer to this specific question. It may be that he feels that since his counterpart theory can explain how something can be φ and also possibly not φ for non-essential properties, he may as well use that for the essential ones as well. But this is a reason to hold worlds never to overlap only when what we are after is the simplest solution to the problem of trans-world identity. It may also be that allowing worlds to overlap would wreak havoc with Lewis's modal realism, since he assumes worlds to be distinct just when they do not share the same space-time. But our aims here are different, and the notion of a possible entity that can be in several possible worlds is, as we will see, a very useful one.

Discrete world semantics are trivially both positive and negative, since both the antecedent in the clause that $\omega \in [\![p]\!]$ and $\omega \subseteq \omega'$ implies that $\omega' \in [\![p]\!]$, and the antecedent in its negative variant with the inclusions reversed, are true only when $\omega = \omega'$.

Straight correspondence. $\mathcal{S}$ is a *straight correspondence semantics* iff it makes every claim p a claim of singular existence of a unique entity $c(p)$ (the *correspondent* of p). In such a case, we commonly refer to the elements of E as *facts* or *states of affairs*. It follows that a claim is true iff its unique corresponding fact exists.

Straight correspondence mirrors the theory directly onto the world: every claim has its unique corresponding fact. This feature makes it instructive as an example, but very susceptible to criticism applicable to all versions of the so-called “picture theory” of language. Its largest problem may be its lack of independent motivation: *why* would the world, if it is not our free creation, have exactly the same structure as our theory, which is? Even if a theory happens to be true, such one-to-one correspondence seems too much to ask for.

Relaxing the uniqueness condition allows us some more interesting semantics. Generally, $\mathcal{S}$ is a *correspondence semantics* iff its truth-conditional function $[\![\cdot]\!]$ makes every claim p a claim of singular existence, although the entity $c(p)$ that p claims the existence of does not have to be unique to p . If $\mathcal{S}$ is complete, we must have that $c(p) = c(q)$ implies that p and q are equivalent in their theory, since if $c(p) = c(q)$, p and q by necessity must be true in exactly the same worlds. It may seem reasonable to take the converse of this to hold as well, i.e. that if p and q are equivalent in their theory, then $c(p) = c(q)$. In such a case, we individuate facts by theoretical (or if the theory in question is a logic, logical) equivalence.

Since correspondence semantics (both the strict and the non-unique kind) interpret every claim as a singular existence claim, and since singular existence claims are positive, all correspondence semantics are positive as well.

Logical atomism. This is defended in Wittgenstein’s *Tractatus*, and in Russell’s lectures on logical atomism from 1917–1918 (Russell, 1985). According to logical atomism, there is a subset of all claims called the *atomic* ones, such that the truth-values of all other claims is a function of the truth-values of these. The atomic claims are true iff their corresponding facts obtain, just as in correspondence semantics, but non-atomic claims do not have corresponding facts. Their truth-values are calculated purely logically.

For a theory of classical propositional logic, classes of logically equivalent sentences make up a free Boolean algebra, and a set of generators of this algebra can be taken to be the classes for atomic sentences, since the truth-functionality of the classical connectives allows us to infer the truth-value of any complex sentence from the truth-values of its atomic parts. Thus the sentences $\{p, q, r\}$ generate a countably large language with p, q and r as atomic.

Since the algebra used to generate a theory is specific to that theory, which claims are atomic is relative to a given algebraisation. When the theory's language has syntactic structure, that language's rules of syntax determines what is atomic and what is not. But this gives rise to a problem: if the things that truly exist are the facts corresponding to true atomic sentences, and what sentences are atomic is relative to the syntax that governs them, what exists becomes syntax-relative. A claim which is complex in one language may be atomic in another.

The solution to this problem may at first seem to be to go nonlinguistic, and use *propositions* or some other abstract entities, individuated by meaning or truth condition, instead of sentences. This only displaces the problem, however. Let $\mathfrak{A}$ be the $2^{2^3} = 256$-element propositional Boolean algebra freely generated by the propositions $\{p, q, r\}$.[11] Unlike what is the case in word algebras, such as those that make up a sentential language, this set of generators is not unique given $\mathfrak{A}$. The same algebra is equally well generated by the propositions $\{p, \neg q, \neg r\}$, $\{\neg p, \neg q, r\}$, $\{\neg p, \neg q, \neg r\}$ or any other such combination. The fundamental problem here is that given a free algebra, generally no unique set of generators of that algebra is determined.

An instance of this phenomenon is that, to logic, it does not matter what we call atomic, and what we call negated atomic. The inference-structure of the language, as well as the algebra, is fully symmetric. But how *do* we determine it? Russell held there to be no syntactic test we could use to find the sentences corresponding to negative facts. Given the question "does putting the 'not' into [the proposition] give it a formal character of negative and vice versa?" his terse reply was "no, I think you must go into the meaning of words" (Russell, 1985, p.78). But that is not much aid either, since we are not told what to look for in these meanings.

It is true that I have proposed a way to define positivity and neg-

[11] As usual, we have assumed that propositions are identical iff they are logically equivalent.

ativity of a claim in the preceding section, but that is of course just a convention: since we have taken the existence claim to be fundamental (which we do since our aim is ontology), existence claims are positive and claims of nonexistence are negative. We could just as well have reversed this, and regarded universal claims to be positive, and negated universal claims to be negative instead.

The same holds for suggestions to, for instance, take singular predications to be positive, and their negations to be negative. Which of the sentences “John is at least 40 years old” and “John is less than 40 years old” should we take to be positive, given that they are negations of each other? That a sentence is *true* is in a classical language the same as it being *not false*, and vice versa, but which of the sentences “p is true” and “p is false” is the negative one? To say that both are positive, since both are *written* as singular predications, is to take the grammatical form of our specific first-order language to determine reality directly.

The conventionality and language-relativity of what is atomic goes farther than being just about negation, however. The sentence “light-ray a is red” is equivalent to “light-ray a is scarlet or light-ray a is crimson or . . . ” for a disjunction of reddish colours, and each of these is in turn equivalent to disjunctions of sentences of the form “the wavelength of light-ray a is in the interval $\lambda_1 - \lambda_2$”, where λ_1 and λ_2 are numbers. “Jim is a bachelor” becomes syntactically atomic simply because English has a predicate that allows us to combine “Jim is male” and “Jim is unmarried”.

It was Russell’s belief that logical analysis would provide us with answers to the question of what the true logical forms of sentences were. About a hundred years later, that belief seems less and less well grounded. Even Russell acknowledged the theoretical possibility that there might not be any “fundamental” level in logic at all, so that we, at least not in a finite number of steps, ever would reach the truly atomic facts by means of analysis. But he should also have noted how, at each step in the analysis, we are making *choices* in how to proceed. We choose how to represent the logical features of a sentence by choosing a logical system (in our case a theory) to express that sentence in, and also a way of translating the sentence to our system. All these choices determine what gets counted as atomic and what gets counted as complex.

It is thus my belief that there are problems with the theory of logical atomism, since it seems that atomicity is a product of language alone, and not of any deeper features of reality or thought. Still, its promise of reducing the number of facts needed for correspondence theories and also its potential applicability to truthmaker theories, which we will study in the next section, make it an important theory to study in case one finds a way to solve these problems.[12]

Whether logical atomism is positive or not depends on whether we consider Wittgenstein's or Russell's version. In the *Tractatus*, all facts are positive. This means that negated atomic sentences are non-positive: $\neg p$ may be true in a world with the facts f_1 and f_2 and false in a world with the facts f_1, f_2 and f_3. Russell, however, argues for the existence of negative facts, and this makes his version of Logical atomism positive. We also, naturally, have that any theory of logical atomism is positive over the class of atomic sentences, since these are true or false by direct correspondence to fact.

5·5 Truthmaker Theories

Truthmaker theory was popularised through articles by Mulligan et al. (1984) and Fox (1987). The fundamental idea is the same as in Dummett's principle: whenever p is true, there is something that *makes* p true. As argued by Rodriguez-Pereyra, truth requires *grounding* in reality, grounding is a relation, and relations relate entities. So truth must be grounded in entities (Rodriguez-Pereyra, 2005). The very idea that truth could be ungrounded in the world seems to violate the re-

[12] The reductive potential should perhaps not be overestimated: logical atomism allows us to dispense with facts for sentences that depend truth-functionally on the sentences in the atomic class. It does nothing to help with other kinds of sentences that may follow logically from the atomic ones.

quirement that what is true is determined by what the world is like.

All contemporary kinds of truthmaker theory posit weaker correspondence principles than correspondence semantics does. Not only does a truthmaker not have to be unique to its truth, but truths may have more than one truthmaker as well, and a claim is true in all worlds where at least one such truthmaker exists. Thus all human beings are truthmakers for "there are humans". But apart from this characterisation, there seems to be little agreement on how truthmaking works. We will introduce the notion through a related one: that of *verifier.*

We say that a is a verifier of p (in symbols $a \Vdash p$) iff $\omega \vDash p$ for any world ω that contains a, and we denote the set of all verifiers of p by $V(p)$ (this set is, of course, relative to a semantics). The existence of a verifier for p is thus a sufficient but possibly unnecessary condition for the truth of p. As before, $[\![p]\!]$ is the set of worlds in which p is true, but we also use the notation $[\![p]\!]^C$ for the set $\Omega \backslash [\![p]\!]$, i.e. the set of worlds where p is false. The following theorem gives a method of finding the verifiers of a claim:

Theorem 5.9 : $V(p) = E \backslash \bigcup [\![p]\!]^C$

Proof. From the definition of $V(p)$, it follows that a is a verifier for p iff $\{\omega \in \Omega \mid a \in \omega\} \subseteq [\![p]\!]$, so $V(p) = \{a \in E \mid \{\omega \in \Omega \mid a \in \omega\} \subseteq [\![p]\!]\}$. This means that the *non*-verifiers of p are

$$\begin{aligned} E \backslash\ V(p) &= \{a \in E \mid \{\omega \in \Omega \mid a \in \omega\} \nsubseteq [\![p]\!]\} \\ &= \{a \in E \mid (\exists \omega \in \Omega)(a \in \omega \wedge a \notin [\![p]\!])\} \\ &= \left\{ a \in E \,\middle|\, (\exists \omega \in [\![p]\!]^C)(a \in \omega) \right\} \\ &= \bigcup [\![p]\!]^C \end{aligned}$$

Thus, $V(p) = E \backslash \bigcup [\![p]\!]^C$. □

Now, if p has a verifier in ω, p is true in ω, but nothing guarantees that the converse holds. Let us call a claim *substantial* if it has a verifier in every world in which it is true. Substantial claims are thus those that are true iff they have an actual verifier. Another characterisation is given by the following theorem:

Theorem 5.10 : p is substantial iff, for any world ω, $\omega \in [\![p]\!]$ iff $\omega \nsubseteq \bigcup [\![p]\!]^C$.

Proof. Obtained by placing the characterisation of $V(p)$ of theorem 5.9 into the condition $[\![p]\!] = \{\omega \in \Omega \mid V(p) \cap \omega \neq \varnothing\}$. □

The substantial claims of a theory, under a necessitarian semantics $\mathcal{S}$, are thus those that are true in all worlds that contain things over and above those things that make up the worlds where they are false. Now, say that a claim r is a *conjunction* of p and q iff $[\![r]\!] = [\![p]\!] \cap [\![q]\!]$, and a *disjunction* of p and q iff $[\![r]\!] = [\![p]\!] \cup [\![q]\!]$.[13] Then, we can prove that the substantial claims of a theory are closed under disjunctions, and if the metaphysics is mereological, they are closed under conjunctions as well.

Theorem 5.11 : If p and q are substantial, their disjunction is substantial as well.

Proof. We use theorem 5.10. Assume that r is a disjunction of p and q, and that $\omega \in [\![r]\!]$. Then $\omega \in [\![p]\!]$ or $\omega \in [\![q]\!]$. Assume that $\omega \in [\![p]\!]$ (the other case is symmetrical). Then, since p is substantial, there is an entity $a \in \omega$ such that $a \notin \bigcup [\![p]\!]^C$. But since, as is easily checked, $\bigcup [\![r]\!]^C \subseteq \bigcup [\![p]\!]^C$, we must have that $a \notin \bigcup [\![r]\!]^C$ as well. Thus $\omega \nsubseteq \bigcup [\![r]\!]^C$, and so one of the directions of the biconditional in theorem 5.10 is satisfied. The other direction follows directly from our definition of disjunction. □

[13] This is a model-theoretic characterisation of the connectives, and, as such, is relative to our semantics. Not all theories need to have conjunctions and disjunctions for arbitrary claims. In fact, only those that have the structure of a *distributive lattice* have them.

Theorem 5.12 : If p and q are substantial, and the necessitarian metaphysics is mereological, their conjunction is substantial as well.

Proof. Again, we use theorem 5.10. Assume that r is a conjunction of p and q, and that $\omega \in [\![r]\!]$. Then $\omega \in [\![p]\!]$ and $\omega \in [\![q]\!]$. Since p is substantial, there is an entity $a \in \omega$ such that $a \notin \bigcup [\![p]\!]^C$ and an entity $b \in \omega$ such that $b \notin \bigcup [\![q]\!]^C$. By the assumption that the model space is mereological, there is then a further entity $a + b \in \omega$, and since $a + b$ is in exactly those worlds where both a and b are, $a + b \notin \bigcup [\![r]\!]^C$. Again, the other direction of the biconditional follows by the definition of conjunction we have used. □

The requirement that the model space has to be mereological is necessary here. Let $[\![p]\!] = \{\omega_1, \omega_2\}$ and $[\![q]\!] = \{\omega_2, \omega_3\}$. Furthermore, let $\omega_1 = \{a\}, \omega_2 = \{a, b\}$ and $\omega_3 = \{b\}$, and assume that these are the only possible worlds there are. Both p and q are then substantial: p has a as verifier, and q has b. There is however no verifier for $a \wedge b$, since both a and b need to exist for that.

It is clear that any substantial claim also is positive. Under the assumption that the model space is mereological, the reverse holds as well: a substantial claim is then true iff at least one of the mereological sums in a certain set exists, and these sums in turn exist iff all their atomic parts do. This means that the substantial claim is true iff all the entities in some set of a certain set exist, which is the condition of positivity of theorem 5.7.

Parallel to the verifier notion, there is that of a *falsifier* of p: an entity a such that in every world where a exists, p is false. We call a claim p which is true iff it *lacks* an actual falsifier *antisubstantial*. Reasoning symmetric to that regarding substantiality shows that antisubstantiality entails negativity, and that it is equivalent to negativity if the model space is mereological.

Now, how is the *verifier* notion connected to that of *truthmaker*? The weakest form of truthmaker principle is one summarised by Bigelow in the slogan “truth is supervenient on being” or, as he also frames it, “If something is true, then it would not be possible for it to be false unless either certain things were to exist which don’t, or else certain things had not existed which do.” (Bigelow, 1988, p. 133)

Seen in terms of possible worlds, this means that if ω_1 and ω_2 are worlds and $\omega_1 \vDash p$ but $\omega_2 \nvDash p$, then there must be some entity in ω_1 which is not in ω_2, or some entity in ω_2 which is not in ω_1. But this is already implicit in $\mathcal{N}$'s characterisation of worlds as determined by what exists in them, so Bigelow's weak truhmaker principle is satisfied by all necessitarian semantics.

While on this weak truthmaker principle the *lack* of something may make a claim true, most truthmaker theoreticians take truthmaking to require the existence of a thing – that in virtue of which the claim is true. Bigelow's characterisation of this position is

> Whenever something is true, there must be something whose existence entails *in an appropriate way* that it is true. (Bigelow, 1988, p. 126, emphasis in original).

The simplest way to interpret this is to let any way be appropriate. Thus we arrive at John Fox's interpretation of the truthmaker principle: "[...] by a truthmaker for A, I mean something whose very existence entails A" (Fox, 1987, p. 189). But, as p entails q iff q is true in all models p are true in, and models correspond to worlds, this is exactly our concept of a verifier.

The principle that a truthmaker is a verifier remains valid even when we do not take every way in which the verifier's existence entails the truth to be appropriate, although not every verifier is a truthmaker then. So the truthmakers of p are some (in the reading where "some" does not exclude "all") of the verifiers of p.

Why would some verifiers fail to be truthmakers? The reason lies in the connotations (or perhaps even meaning) of "making": to *make* p true seems to involve taking active part in bringing about its truth. By contrast, a verifier is just something whose existence *guarantees* the truth of p. This means that for verifiers, the *entailment principle* holds:

$$\text{If } a \Vdash p \text{ and } p \text{ entails } q\text{, then } a \Vdash q.$$

The alleged problem with this principle comes out clearest with necessary statements (i.e. those that are in $\top_A$, where A is the theory we are working with). Since these claims are true in all worlds, any possible entity is sufficient for their truth. Yet, it feels strange to say that a purportedly necessary truth such as $\varnothing \subseteq \varnothing$ is *made* true by, say, my pencil. Similar counter-intuitive consequences also follow from the related (but distinct) *containment principle*, imposed by Mulligan et al. (1984, p. 315):

$$\text{If } a \Vdash p \text{ and } b \text{ contains } a \text{ as a part, then } b \Vdash q.$$

If worlds have mereological sums, so that the actual world as a whole makes up a possible entity, then the containment principle entails that this world-sum is a truthmaker for all actual truths. Yet it is hardly an *interesting* truthmaker, since it gives no information about which *specific* things in the world make which sentences true.

The difference between the problems stemming from the entailment principle, and those stemming from the containment principle, is that it seems like in the former class, the thing made true is unnecessarily weak, while in the latter, the truthmaker itself is unnecessarily strong. One perspective from which the two principles could seem questionable is if one takes the truthmakers of p to *explain* why p is true. On a certain reading, something's explaining why p is the case does not explain why q is, even if q follows from p.

To borrow an example of Kyburg's, all salt which has had a dissolving spell cast on it dissolves in water, but it still feels wrong to say that the sentence "substance s is salt which has had a dissolving spell cast on it" explains the truth of the sentence "substance s dissolves in water" (Kyburg, 1965). But "substance s is salt which has had a dissolving spell cast on it" entails "substance s is salt", so if the latter explains the fact that s dissolves in water, then so should the former. However, as an *explanans*, it seems to be too strong and include irrelevancies, and this makes us doubt whether an explanation is given.

For an example of an explanation where the explanandum seems too weak, consider the explanation "the window broke because I threw a stone at it". From "the window broke", it follows that either the window broke or turned into a platypus. But again, it is counter-intuitive to say that my throwing a stone at the window explains why it broke or turned into a platypus.

We will call the property supposedly lacking in those verifiers of p that fail to be truthmakers *effectiveness*. There are several ways to try to substantiate this notion. One, which involves a *prima face* fairly small change, accepts the entailment principle but locates the problem in the specific entailment relation. Thus both Mulligan et al. (1984) and Restall (1996, 2003) advocate using some kind of relevant entailment relation instead of the classical variant. This may at first seem like nothing which is excluded by our method, since the principles of relevant entailment very well can be expressed in a consequence operator, and systems of relevant entailment framed as theories. But the point that these philosophers make is that even if we otherwise accept classical logic, truthmaking is not preserved across entailment. Thus, what they count as truthmakers will generally only be some of the things that we count as verifiers here. However, this attempt runs into serious difficulties, as we will see in chapter 7.1.

Another attempt to capture the effectiveness of a truthmaker might proceed via the notion of a *minimal* verifier: an entity a that verifies p, such that no proper part of a verifies p.[14] This takes care of the perceived problem that the world verifies every truth: most truths will have some smaller part of the world that verifies them as well. But it is hard to use as a universal solution, since we have no guarantee that every truth has a minimal verifier. Take, for instance, a sentence such at "this pole is over one metre long", and assume that the pole in question is, say, one and a half metres long. Any part of the pole longer than one metre is then a verifier for the sentence, but because of the continuousness of space, there is no *least* length over one metre that such a part can be.

Now, there may of course be no contradiction inherent in accepting entities such as *the fact that the pole is over one metre long*, which would make the sentence true. The problem is that they are quite

[14] There another notion of minimality floating around, according to which a minimal truthmaker for p is a truthmaker for p that is part of any truthmaker for p (cf. Restall, 1996). This does however have very few applications: any truth made true by more than one thing may fail to have a minimal truthmaker in this sense.

strange entities – we can see them as a kind of infinitely disjunctive facts of the form *the pole is x_1 metres long or the pole is x_2 metres long or …* where x_1, x_2 etc. are all real numbers larger than one, and we may have reasons not to accept disjunctive facts in our ontology. But even if we do, requiring truthmakers to be minimal verifiers does not solve all the perceived problems: assuming that the world contains mereological atoms, any such atom will still be a minimal verifier for every necessary truth, since the atoms do not have any parts at all.

For a third way to characterise effectiveness, we may look more to the logical side than the mereological. Say that a verifier a of p is a *weakest* verifier of p iff for any claim q such that q entails p, $a \Vdash q$ implies that p and q are equivalent. Thus, while $a \Vdash p \wedge q$ implies that both $a \Vdash p$ and $a \Vdash q$, a can be a weakest verifier neither for p nor for q, unless one of these follows from the other. This accords with an argument of Rodriguez-Pereyra's (Rodriguez-Pereyra, 2006) that the truthmaker of a conjunction generally is not what makes true its conjuncts. But, *prima facie*, not all truths need to have weakest verifiers either. If a verifies p and b verifies q, both a and b thereby verify $p \vee q$. Neither a nor b can however be a weakest verifier for $p \vee q$, unless one of p or q entails the other. To obtain a weakest verifier for $p \vee q$, we need to assume that this claim has its own truthmaker, and this will again be a kind of "disjunctive" entity. Another problem with this characterisation of effectiveness is that it does not allow individual X's to be truthmakers for the claim "there are X's", and this is one of the possibilities motivating many philosophers' adoption of truthmaker theories, rather than more traditional correspondence semantics.

It might also naturally be the case that there is no systematic way to characterise what verifiers are the effective ones, and that we will be forced to rely on intuitions about which entity is the "active agent" in bringing about the truth of a claim. The reason for this may be that effectiveness simply is not a structural property. Again we could make analogies with the theory of explanation. All explanation is critically context-sensitive, and which deductions are explanations depends on what we know, and this is one of the lessons Kyburg draws from his example. Salmon's famous example of the length of a flagpole's shadow not explaining the length of a flagpole (Salmon, 1989, p. 47) does not apply in a case where the flagpole has been raised just high enough so that its shadow at noon will reach a certain point, and we ask why it has just this height.

But invoking knowledge, context or pragmatics in truthmaking is clearly inappropriate; truthmaking should not be relative in that way. So if effectiveness is to be applicable to truthmaking, we *need* some criterion to decide whether it holds or not, and we need this criterion to be context- and knowledge-independent. Unfortunately, no such criterion seems to be available.

In any case, even for a philosopher who holds all real truthmakers to be effective, the notion of a verifier is interesting as a way to delimit the range of *potential* truthmakers for a claim. For an effectivist, the "real" truthmakers will be a proper subset of these, but as the way to pick out this subset is far from clear, all we will take a truthmaker for p to be is *some kind of verifier*. Letting $TM(p)$ be the set of truthmakers of p, we write this condition as $TM(p) \subseteq V(p)$. Consequently, we will also take a *falsemaker* (i.e. that in virtue of which a claim is *not* true) to be some kind of falsifier.[15]

The fundamental rule of truthmaking, which we accept, is that if a makes p true and a exists, then p is true. Apart from this, however, opinions on how to substantiate the theory diverge. *Truthmaker maximalism* (Armstrong, 2004) holds that truthmaking is required for truth, so that for *any* claim p, if p is true in a world, then there exists some truthmaker for p in that world.

[15] An interesting analogue may be made with a theory that has roughly the same structure as truthmaker theory: the intuitionistic characterisation of mathematical truth. According to such an interpretation of truth, what we mean by "p is true" is that we are in possession of (or have means of obtaining) a proof with certain characteristics (i.e. those that make it a proof of p, rather than of something else). Not every such proof can however be constitutive of the meaning of "p is true", so a special class of *canonical* proofs is often identified (see Dummett, 2000, pp 68–98).

The similarity should be clear. Intuitionistic mathematics rests on the truthmaker principle, and takes proofs to be the truthmakers. The canonical proofs correspond to our effective truthmakers. There are some differences, however: truthmaker theorists generally believe that several truths may have the same truthmaker, but if the identity of a proof determines the identity of what it is a proof of, then no proof can make true more than one statement. We will return to the relationship between truthmaker theory and intuitionism in ch. 7.1.

Given the assumption that truth requires truthmakers, why would a truthmaker theorist not want to be a maximalist? One reason is if you believe that some truths are fundamental, and others are derivative. Thus, truthmaker theory is combinable with logical atomism, or variants of it. We may hold that conjunctions have their own truthmakers (for instance, the mereological sum of the truthmakers of the conjuncts), but that disjunctions are true only because one of the disjuncts is made true by something. A more common standpoint is to hold that some claims have truthmakers, but that their negations are true simply in virtue of their lack of falsemakers.

The same problems with identifying the fundamental claims that we found when discussing logical atomism apply to logically atomistic truthmaker theory as well. How do we determine which claims in a theory are the ones that have truthmakers? We will not attempt to do that here, as we can still apply truthmaker theory to the fragment of a theory consisting of claims that *do* have truthmakers. For every theory A and truthmaker semantics $\mathcal{S}$ from A, there must be some theory A' which is part of A, and whose language consists of all claims in L_A that are true iff they have actual truthmakers according to $\mathcal{S}$.

Another way of weakening the maximal truthmaker principle is to allow that although some truths may have no single truthmaker, several things jointly can make them true. This is the version advocated by Mulligan, Simons and Smith, and it has the advantage that we do not have to postulate the existence of a single thing such as *the three apples in the bowl*. Such an entity's existence does follow from accepting a mereological metaphysics, so if we have assumed that anyway, the singular truthmaker for "there are three apples in the bowl" will be no further commitment. We may still want the plurality of the three apples as well, however, since it allows us to hold on to the principle that sentences of the form "there are n x's" always are made true by n things jointly. Unlike the mereological sum of the three apples, their plurality retains its *threeness*.

If $\mathcal{S}$ allows truthmaking by pluralities, and we interpret truthmaking in the way that the truthmakers of p are the verifiers of p, then p is true iff all the things in at least one of the pluralities that make p true exist (we can see here that it is not the pluralities themselves that need to exist, but only the things in them). But this is exactly the same condition that theorem 5.7 shows charactersises positivity. Thus, p is made true by some plurality iff p is positive. Likewise, if we accept plural falsemakers, p is made false by a plurality iff p is negative.

While a plurality should not be taken to be an entity in its own right, but rather as a vehicle for plural reference, nothing hinders us from *representing* a plurality as a set. *Plural truthmaking semantics* then involves the condition that for any claim p, there is a set $\mathit{VP}(p)$ (the verifying plurality) of sets of entities such that p is true in the world ω iff $X \subseteq \omega$, for some $X \in \mathit{VP}(p)$. Since a positive semantics fulfils the same condition, positive semantics also allows us to identify the truthmaking pluralities for any claim p. We can also define falsifying pluralities the same way, so that a negative semantics gives rise to a set of falsifying pluralities $\mathit{FP}(p)$ for every claim p.

One thing worth noting about plural truthmaker semantics is that it also automatically gives us pluralities for making true sets of claims: it is quickly proved that if the sets $X_1, \ldots X_n$ of entities make true the claims $p_1, \ldots, p_n$, respectively, then the union of $X_1, \ldots X_n$ makes true all of p_1 to p_n. While this holds for non-effectivist regular truthmaker semantics with the assumption that the metaphysics is mereological as well, we may need further assumptions to prove the same thing without a mereological metaphysics or with an effectivist notion of truthmaking.

5.6 Necessitarian Interpretations

The preceding sections have discussed necessitarian semantics from a top-down, structural perspective, and we have thus not said anything about what the interpretations in these semantics are. As we have stressed, the role of an interpretation of A is to provide information that together with knowledge of the interpretation's model will allow the semantics to assign a semantic value to the claims in A's language L_A. This principle makes it fairly easy for us to find reasonable interpretation functions for different kinds of necessitarian semantics.

The most direct forms of necessitarian semantics that we have encountered are the correspondence semantics. Here, we have that the truth-value of p in the model $\mathfrak{M}$ is determined by whether p's correspondent exists in $\mathfrak{M}$ or not. It is therefore natural to take an interpretation of A in $\mathfrak{M}$ to be a partial function $h_{\mathfrak{M}}$ from L_A to $\mathfrak{M}$, such that $h_{\mathfrak{M}}(p) = c(p)$ iff $c(p) \in E_{\mathfrak{M}}$. We then let $\mathcal{S}(h_{\mathfrak{M}}, p) = \mathit{true}$ iff $h_{\mathfrak{M}}(p)$ is defined.

One property of this semantics is that *when* a correspondent of p exists, it is always the *same* correspondent. This is attractive because it means that what A corresponds to does not depend on what the world is like. If A is a language, it can be taken as an indication that we can learn that language's reference rules separately from learning about the rest of the world.

As we mentioned, correspondence semantics are not very plausible. Far more popular these days are be their generalisations in various forms of truthmaker semantics. Here, the lack of a unique correspondent means that an interpretation cannot in general associate a single entity with each claim, or even with each true claim. Instead, let us take $h_{\mathfrak{M}}(p)$, for any model $\mathfrak{M}$, to be the intersection of the set $\mathit{TM}\ (p)$ of truthmakers of p with the set $E_{\mathfrak{M}}$ of entities in $\mathfrak{M}$. This way, a claim gets interpreted as the set of its existent truthmakers. Naturally, p is then true iff this set is not empty.

Moving upwards in generality, we come to the case of positive semantics. But we have already seen that this is equivalent to truthmaking via pluralities, so the natural generalisation is to let $h_{\mathfrak{M}}(p)$ be the set of those pluralities that verify (or make true, depending on whether we consider the effectivist version or not) p, and that wholly exist in A. Formally, since we represent pluralities using sets, we let $h_{\mathfrak{M}}(p) = \{X \in \mathit{VP}\ (p) \mid X \subseteq E_{\mathfrak{M}}\}$. As in the case of non-plural truthmaking, the semantics must then take p to be true in $\mathfrak{M}$ iff $h_{\mathfrak{M}}(p) \neq \varnothing$.

Finally, how should we characterise interpretations in the fully general case, where we have made no further assumptions on the function $[\![\cdot]\!]$? One idea is that we could let $h_{\mathfrak{M}}(p)$ be the whole of $\mathfrak{M}$, since it appears that all of $\mathfrak{M}$ is relevant to whether certain entities *do not* exist. But we can also reformulate the metaphysics slightly, in order to accommodate this case, and it will turn out in the next chapter that this gives the semantics nicer properties.

Given a necessitarian metaphysics $\mathcal{M} = \langle E, \succ\!\!\!\in \rangle$, let a *circumstance* $(X|Y)$ be a pair in which X and Y are disjoint subsets of E, and let $CRC(E)$ be the set of all circumstances constructible from the set E. Where $\mathfrak{M}$ any model in $\mathcal{M}$, we say that $(X|Y)$ *holds* in $\mathfrak{M}$ iff $X \subseteq \omega$ and $Y \cap \omega = \varnothing$ (i.e. iff $\mathfrak{M}$ contains all of X and nothing from Y), and we write the set of all circumstances that hold in $\mathfrak{M}$ as $CRC(\mathfrak{M})$. Intuitively, a circumstance can thus be interpreted as an *occurrence* of certain entities, together with an *exclusion* of certain others.

The circumstances can do work analogous to that done by truthmakers in a truthmaker semantics. Let the set of *verifying circumstances* $VC\,(p)$ be defined as

$$VC\,(p) = \{(X|Y) \mid (\forall \omega \in \Omega)((X \subseteq \omega \wedge Y \cap \omega = \varnothing) \rightarrow \omega \models p)\}$$

i.e. the set of all circumstances such that if any of them holds in the world ω, then p is true in ω. These sets, for various claims p, can be used to define values of the interpretation function for a model $\mathfrak{M}$, by letting $h_{\mathfrak{M}}(p) = VC\,(p) \cap CRC(\mathfrak{M})$. The resulting semantics is, as with other truthmaker theories, defined to give the value *true* for the claim p in the model $\mathfrak{M}$ iff $h_{\mathfrak{M}}(p) \neq \varnothing$.

Do circumstances exist? In one sense they can be held to do: we are free to say that the circumstance $(X|Y)$ exists when the elements in X exist, and none of those in Y do. However, no entities are involved other than those of the model space we are working with, which follows from the fact that we can translate talk of circumstances into talk of possible entities without loss. Circumstances avail us of another vantage point from which to view necessitarian metaphysics, and thus a translation into circumstance-talk functions as a sort of *transformation* of our area of discourse, after which certain problems may be easier to solve. In this it works much as the Fourier transform or the Taylor series expansion do in mathematics, and just as an analytic function's being an infinite sum of sines does not rule out its being an infinite sum of polynomials as well, we can hold that a possible world is *both* a collection of possible entities, and a collection of circumstances, depending on how we see it.

This means that, at least with a little ingenuity, all forms of necessitarian semantics can be modelled on the truthmaker paradigm.[16] As a summary, table 5.1 collects the four classes of semantics for necessitarian metaphysics that we have discussed, in increasing order of generality. For the three last, there are furthermore two varieties: the non-effective one, and the effective. According to the effective versions, the truthmakers (or truthmaking pluralities, or truthmaking circumstances) of p are taken to be only a subset of those that are sufficient for the truth of p. Since correspondence semantics matches claims to unique features of models, these are trivially effective, or the claims in question could not have been true at all.

Semantics	***Effective***	$h_{\mathfrak{M}}(p)$
Correspondence	—	$c(p)$ if $c(p) \in E_{\mathfrak{M}}$ undefined otherwise.
Truthmaking	*No*	$VC\,(p) \cap E_{\mathfrak{M}}$
	Yes	$TM\,(p) \cap E_{\mathfrak{M}}$
Positive	*No*	$VP\,(p) \cap \wp(E_{\mathfrak{M}})$
	Yes	$TMP\,(p) \cap \wp(E_{\mathfrak{M}})$
General	*No*	$VC\,(p) \cap CRC(\mathfrak{M})$
	Yes	$TMC\,(p) \cap CRC(\mathfrak{M})$

Table 5.1: ***Types of necessitarian semantics***

For the first of these semantics, we have that p is *true* iff $h_{\mathfrak{M}}(p)$ is defined. For the others, the truth condition for an arbitrary claim p is

[16] This might not be that surprising, since necessitarian semantics already has been noted to coincide with the weak "truth supervenes on being" interpretation of truthmaking.

$$\mathfrak{M} \models p \text{ iff } h_{\mathfrak{M}}(p) \neq \varnothing$$

Common to all of the semantics described here is the fact that $h_{\mathfrak{M}}(p)$ is *constant* wherever it is defined, or at least constant on the overlapping parts of models. This is typical of a Leibnizian semantics, and allows us to "paste together" the interpretations in each model in order to define a *global* interpretation function h, common to the whole semantics. Such an interpretation function is given by c for correspondence semantics, by *TM* (or *V*) for truthmaker semantics, by *TMP* (or *VP*) for plural truthmaker semantics, and by *TMC* (or *VC*) for circumstance semantics.

The seven semantics we have defined here are, of course, only a few of those that could be defined, and necessitarian metaphysics only make up a small part of the conceivable model spaces. We are thus faced with an infinite multitude of choices whenever we are to interpret a theory A.

Seen from a certain viewpoint, this freedom may appear almost contradictory. Does not our use of the claims in a theory determine their meaning, and should not that meaning determine which is the correct semantics to use? This can look even more perplexing when we consider the fact that our theories are not just sets of uninterpreted sentences, but sets of truth-bearers together with consequence operators. The possibility of meaningfully assigning truth or falsity to a claim seems to require us to have some interpretation in mind of that claim, or we wouldn't know what it was that we called true or false. Likewise, the existence of consequence relations among claims may seem to require these to be interpreted, or we would not have any reason to believe these consequence relations to hold. According to this line of thought, a semantics is necessary as a precondition both for judging claims true or false, and for being justified in believing one claim to follow from another.

There is of course no reason to deny that when a scientist deems it true that gold melts at a temperature of 1064°C, she has some kind of interpretation of her words in mind, and does not treat them as meaningless symbols. But this interpretation does not have to involve any kind of full referential semantics. The scientist's acceptance of "gold melts at 1064°C" is generally based on observations and experiments carried out, and knowing which observations are relevant to the truth or falsity of the sentence is all she needs to know about its meaning. In short, she does not need the truth conditions, but only verification conditions.[17] For more deeply theoretical claims, such as “the neutrino is an uncharged particle”, the verification conditions may take a back seat to the more general idea of conceptual role (Harman, 1974; Field, 1977). Still, no knowledge of a semantics in the sense we have been using the word is necessary for the working scientist.

Likewise, the consequence relation (or operator) does not have to arisen from a referential semantics, but can very well be the product of trial and error: if we have observed that occurrences of p are correlated with occurrences of q, we can try allowing inferences from p to q. So long as these do not lead from a claim we have good reason to believe to be true, to one that we have good reason to believe to be false, such an inference rule can be seen as empirically justified.

Of course, it is not my intention to argue against referential semantics for natural language in general here. In fact, considering different semantics makes sense even if the speaker already should have a definite semantics in mind, at least as long as the model spaces that the semantics take claims into differ. Consider, for instance, different Tarskian model spaces, and a claim such as “there is a hand”, translated into predicate logic as $(\exists x)\mathit{Hand}(x)$. In a “common sense” model space $\mathcal{M}_1$ whose domains include hands and other body parts, the extension of “*Hand*” naturally must include the hands in the domain, and nothing else. But what of another model space, whose models have different domains?

Let $\mathcal{M}_2$ be a model space whose domains only include elementary particles, spacetime points and mereological sums of these. Now, in some of these models, it may still be true that there is a hand. Let a mereological sum of particles satisfy $\mathit{Hand}(x)$ iff their spacetime positioning makes them sufficiently hand-like. Then some models contain hands, and some do not. But how is “sufficiently hand-like” to be interpreted? While the meaning of the everyday word “hand” excludes some

[17] By “verification conditions” we do not only mean those conditions under which the sentence would be conclusively verified, but also the ways in which observations count as evidence for or against that sentence.

shapes (for instance a completely spherical one), it is simply not exact enough to determine a unique semantics into $\mathcal{M}_2$. Several semantics may thus be equally right.

Van Fraassen, borrowing a term from Eco, characterises science as an *open* text—one that does not come with a full, detailed interpretation (van Fraassen, 1991, 8–12). But the same, in varying degrees, holds for every area of discourse. There can be no such thing as a *completely* closed text, since whenever we are to specify how to interpret a certain statement, the statements that we use in such a specification need interpretation as well.

Similar considerations apply to the role of truth conditions. To a certain brand of philosopher of language, it may seem like a truism that understanding a sentence requires knowing its truth conditions, and there is indeed a sense in which it is. If one understands "snow is white", and understands what it is for a sentence to be true, one knows that "snow is white" is true iff snow is white. This, however, has more to do with knowing about truth than about "snow is white", since the concepts used on the right-hand side of the biconditional are the same as those on the left-hand side. In the simplest cases, the metalanguage contains the object language, so the translation of p into this language will always be homonymous, yielding no further information.

Taking understanding p to involve knowing p's truth conditions in *some* language thus imposes close to no limitation at all. But requiring an understander to know the truth conditions expressed in *all* languages she knows seems too strong. I may be able to understand the language of quantum mechanics quite well, and also to understand English well enough, but still have no idea of what the truth conditions of "there are three apples in the bowl on my table" are, expressed in the language of quantum mechanics. I simply do not know enough about the constitution of apples to do that.[18] It is also worth pointing out that much

[18] A very similar observation is made by Feynman in his *Lectures on Physics*: "In order for physics to be useful to other sciences in a *theoretical* way, other than the invention of instruments, the science in question must supply to the physicist a description of the object in a physicist's language. They can say 'why does a frog jump?,' and the physicist cannot answer. If they tell him what a frog is, that there are so many molecules, there is a nerve here, etc., that is different." (Feynman, 1963, p. 3-9).

of the information lacking seems to be of an empirical rather than a linguistic kind.

A model space, as we have explained, *is* a kind of language, since it through its canonical theory provides the means to say that the actual world is in some subset of the models in the space. We have just made the point that knowing a model space $\mathcal{M}$ and knowing a theory A is not sufficient for us to be able to infer how A is to be mapped to $\mathcal{M}$. But how do we then choose our semantics? How do we determine if truthmaker maximalism is correct or not, for instance?

There are *some* conditions that exclude certain semantics. An unsound semantics, for example, cannot do the work we need it to do, and ideally the semantics should be complete as well. There are trade-offs to take into account here, however: many logicians prefer to use the intended semantics for second-order logic, despite this semantics being incomplete, rather than the Henkin semantics, which is complete. In some way these logicians may be said to hold that the intended semantics better captures what they *mean* by claims such as $(\exists P)P(c)$ than the Henkin semantics does. So questions of meaning may deliver *some* guidance in the choice of semantics.

We can also focus on the theoretical side of the question. As theorem 5.3 shows, we can regard a semantics from A into $\mathcal{M}$ as a translation of A into $Th(\mathcal{M})$. But a translation between two theories is itself a theory: one that says what claims of the two theories are equivalent. Since the notion of equivalence we are interested in here is truth-conditional equivalence, adopting a translation involves possibly *substantial* claims of the type "$p \in L_A$ is true iff $q \in L_B$ is true", where A and B are the theories in question. If we hold all theories to be true or false, these conditions may also place further restrictions on which semantics are appropriate.

Chapter 6
The Theory–World Connection

Here we show how to use the concepts introduced in chapters 2 to 5 to draw conclusions about metaphysics from the truth of theories. We call a semantics *Hertzian* if it induces a specific type of connection between the logical structure of its theory, and the necessitation-structure of its metaphysics. It is shown that all the kinds of necessitarian semantics we have discussed fulfil this condition, and it is this that make them useful for metaphysical investigation.

The second half of the chapter concerns questions about ontological commitment. First, we give a general theory of ontological commitment, applicable to all types of model space. We distinguish between *specific* commitments, which are those things that are in all models a claim is true in, and *general* commitments, which are the *types* of things that a claim commits one to. Of these, the second is in general the more useful.

In the final section, these concepts are applied to necessitarian metaphysics. The most important result here is that from the point of view of ontology, it does not matter if we take truthmaking to require effectiveness or not—the ontological commitments of all claims are the same anyway. Since non-effective truthmaking is a much clearer concept than effective, we are therefore justified in concentrating primarily on this type of semantics.

6.1 Hertz's Principle

We have highlighted the amount of choice and conventionality involved both in choosing what model space to represent a theory in, and what semantics to use for mapping the theory to that space. But this choice is certainly not *arbitrary*: many requirements on theories can be turned around and viewed as requirements on model spaces or their semantics instead. We have treated soundness and completeness as properties of semantics, but this works only because we have taken a semantics to determine the theory and the model space that it involves. Thus we can envisage holding the semantics and the model space fixed, and see which inferences preserve truth in all interpretations, as is what is done when we try to axiomatise a theory for which we already have a semantics. Alternatively, we can hold the theory and the semantics fixed, and see how different types of model space fit in. This is the task of metaphysics: to design and study model spaces for a given theory.

But model spaces can not be studied on their own, when we are looking to use them for a given theory. We have to look at model spaces together with their semantics. Thus, we should look at ways of evaluating combinations of semantics and model spaces, given a theory A. At first, it may seem like *consistency* ought to be one of the requirements we can place on a model space or a semantics. But since we have defined both model spaces and semantics set- and category-theoretically, and not in terms of descriptions of these sets (or functions) and categories, consistency is not applicable. If the description we have given of what a model in $\mathcal{M}$ is happens to be inconsistent, then $\mathcal{M}$ is the empty category, and thus unusable for any semantics (since we cannot define interpretations to it). But this depends on the relation between model space and theory, rather than on the model space itself.

Some properties on this level that are relevant to us are Leibnizianness and Bolzanianness. Since we are interested in drawing metaphysical conclusions from theories, we should concentrate on the use of Leibnizian semantics. A non-Leibnizian semantics mixes the metaphysical with the linguistic, and this makes it much harder for us to find out what the theory says about reality, rather than about the language the theory is expressed in. So, unless we specify otherwise, we will take the semantics we work with to be Leibnizian.

Ideally, we should also want the semantics to be appropriate in the sense that the actual model $\mathfrak{A}$ (i.e. the theory's subject matter) is an element of the model space. As we explained in section 5.2, this is unfortunately not a rule that can be enforced: we cannot decide whether a model is in a model space except through the use of theories and semantics. This moves the proper focus from appropriateness to *soundness* and *completeness*.

Soundness gives us some kind of safety against the theory contradicting the semantics. That A is true means that if $X \vdash_A p$ and X is true, then p is true as well. But if $\mathcal{S}$ is unsound, then it may be that the actual model $\mathfrak{A}$ is such that $\mathfrak{A} \vDash_{\mathcal{S}} X$ but $\mathfrak{A} \nvDash_{\mathcal{S}} p$, and if $\mathcal{S}$ then is appropriate, this would mean that X is true, but p is not, so this contradicts what the theory says. Using a non-sound semantics for a theory runs the risk of being incompatible with the theory itself.

Completeness may at first be thought to be slightly less crucial. A theory usually does not come with any guarantee that the inferences it licenses are *all* the inferences in its language that happen to be truth-preserving. While there are exceptions, such as second-order Peano Arithmetic, most theories purport to tell us only part of the truth about the things they concern. But complete semantics still hold special interest for our purposes: if we want to investigate the metaphysics involved *in a specific theory* A, then using a complete semantics makes sense. Since an incomplete semantics leaves out some things about the ways the world can be according to A, it does not give us full information about A's metaphysics. Furthermore, if the semantics in question is bivalent, completeness also guarantees that there is *some* model in which the same claims are true and false as in the actual one.

Soundness and completeness thus give us a more *objective* way of evaluating semantics than appropriateness. But they do not take us very far into metaphysics: as the following theorem shows, given *any* model space $\mathcal{M}$ of sufficient cardinality and any theory A, we can design a semantics $\mathcal{S}(A \mapsto \mathcal{M})$ which is sound and complete.

Theorem 6.1 : Let A be a theory, and $\mathcal{M}$ a model space of cardinality at least $2^{|L_A|}$. Then there is a Leibnizian semantics $\mathcal{S}(A \mapsto \mathcal{M})$ from A to $\mathcal{M}$ which is sound and complete.

Proof. One way to construct such a semantics is as follows. Let φ be a surjective function from the models in $\mathcal{M}$ to the theories in A (such a function exists by the axiom of choice, and it can be surjective because of the cardinality requirement). For each model $\mathfrak{M} \in \mathcal{M}$, define an interpretation to be a partial function $h_{\mathfrak{M}} : L_A \to \mathcal{M}$ such that $h_{\mathfrak{M}}$ is defined and $h_{\mathfrak{M}}(p) = \mathfrak{M}$ iff $p \in \varphi(\mathfrak{M})$. Let the semantics $\mathcal{S}$ map $h_{\mathfrak{M}}, p$ to *truth* iff $h_{\mathfrak{M}}$ is defined at p.

The function φ effects a translation for $\mathcal{M}$ to the model space $\mathcal{T}\!h_A$ of theoretical models of A. The resulting semantics is sound and complete because this semantics is. □

To be useful for metaphysics, we need the semantics we have used to incorporate deeper connections. We still, however, want to stay on a structural level: purported conditions such as the semantics having to capture "what the theory truly means" are not what we are after here. Instead, our main idea will be a principle that we attribute to Hertz, on basis of his position in the philosophical introduction to *The Principles of Mechanics Presented in a New Form* (a book that, incidentally, is said to have had a great effect on Wittgenstein). Hertz defends the so-called *picture theory of science* (not to be confused with what Heil called the "picture theory" of language; cf. sct. 1.2) according to which the creation of a scientific theory is much like the painting of a picture.

How to properly paint such a picture is, however, constrained by certain principles. The most important of these is described by him as follows.

> We form for ourselves images or symbols of external objects; and the form which we give them is such that *the necessary consequents of the images in thought are always the images of the necessary consequents in nature of the things pictured.* In order that this requirement may be satisfied, there must be a certain conformity between nature and our thought. Experience teaches us that the requirement can be satisfied, and hence that such a conformity does in fact exist. (Hertz, 1899, p. 1, emphasis added)

We will refer to the principle empasised in the quotation as *Hertz's principle.* It requires the consequences we can draw from the theory to match the those that follow by necessity in nature. Graphically, expressed in our terminology, Hertz's principle says that the following diagram must commute, for any set X of claims (i.e. "images"):

$$\begin{tikzcd}
X \arrow[r, "inference"] \arrow[d, "h"'] & C(X) \arrow[d, "h"] \\
\Phi \arrow[r, "necessitation"'] & \Phi'
\end{tikzcd}$$

Here, $\Phi = \{\phi_1, \ldots, \phi_n\} = h[X]$ contains the features of reality that the claims in X are images of (to use Hertz's terminology), and likewise for Φ'. The interpretation h maps each claim into the feature of reality that it is an image of, and Hertz's principle says that the claims we can infer from X must be those that are images of the features of reality that are necessitated by those imaged by X. Although Hertz was interested in motivating the inferences of a theory (the "necessary consequents of the images in thought") from observed necessities, nothing hinders us from reversing this process when we are given a theory, and using the principle to evaluate semantics and metaphysics as well.

Looking closer, we can see that Hertz's principle is the conjunction of two subprinciples. Where $\mathcal{S}(A \mapsto \mathcal{M})$ is a semantics, we define the following properties:

> $\mathcal{S}(A \mapsto \mathcal{M})$ is *Hertz-sound* iff, for all $X \subseteq L_A$ and $p \in L_A$ such that $X \vdash_A p$, and any model $\mathfrak{M} \in \mathcal{M}$, any features of $\mathfrak{M}$ in virtue of which X is true necessitate some feature in virtue of which p is true.
>
> $\mathcal{S}(A \mapsto \mathcal{M})$ is *Hertz-complete* iff, for all $X \subseteq L_A$ and $p \in L_A$, the necessitation of some feature of any model $\mathfrak{M}$ in virtue of which p is true by any features of $\mathfrak{M}$ in virtue of which X is true, entails that $X \vdash_A p$.

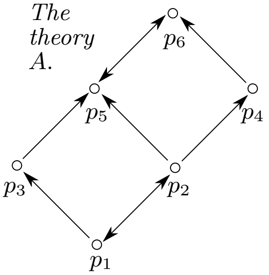
A semantics that is both Hertz-sound and Hertz-complete will be called *Hertzian*. We have used the word "necessitate" in a general, vague sense here, to enable us to specify what this means more closely depending on which semantics or metaphysics we use. The "features" of a model $\mathfrak{M}$ are what an interpretation $h_{\mathfrak{M}} : L_A \to \mathfrak{M}$ maps claims to, and that p is true in virtue of such a feature ϕ simply means that p's truth in $\mathfrak{M}$ can be inferred from knowing that $\mathfrak{M}$ has the feature ϕ, and that p expresses possession of this feature (i.e. that $h_{\mathfrak{M}}(p) = \phi$).

The specific features of models involved are thus determined by the interpretation functions available in a semantics. We should remark right away, though, that they do not have to be taken as ontologically primitive, in the sense that we do not have to say that this or that feature of reality *exists*. It is a convenient language in which to express connections between theories and metaphysics, and it can be translated on a case-by-case basis to language that does not use these concepts. We will show how to do so for varieties of necessitarian semantics in the next section.

Hertzianness gives us a way to evaluate semantics and metaphysics which is slightly stronger than using only soundness or completeness. For one thing, theorem 6.1 does not give us Hertzianness of the induced semantics, since the procedure outlined in the proof makes the interpretation function $h_{\mathfrak{M}}$ map all true claims to the same feature (viz. the model $\mathfrak{M}$ itself). If we then say that $\mathfrak{M}$ necessitates itself, Hertzianness would require the theory to allow all true statements to be interderivable, and if we deny that $\mathfrak{M}$ necessitates $\mathfrak{M}$, the condition entails that no inferences among true statements are allowed at all.

The important difference between Hertz-soundness and regular soundness, and Hertz-completeness and regular completeness, is that the former in each pair relates a theory to properties of *single* models. By contrast, regular soundness and completeness concern only which models are taken to satisfy which claims. This is why we can say that Hertz-soundness and Hertz-completeness go deeper: they can provide *reasons* for soundness and completeness to hold.

These properties thus guarantee intimate connections between the structure of a true theory and the structure of the metaphysics (i.e. the model space). Take, for instance, the theory A depicted on the right, with a language consisting of the claims $p_1, ..., p_6$, and an inference relation for which $p_i \vdash_A p_j$ iff there is a way to go from p_i to p_j by following the arrows. Inferences from sets of claims can be defined by letting $X \vdash_A p_i$ iff the greatest lower bound (i.e. the meet) of all claims in X has p_i as a consequence.


The theory A.

Interpreting this theory through a Hertzian semantics gives rise to the kind of correlation shown in fig. 6.1. Here, we have A with its inference structure on the left, and we have a fragment of the metaphysics, with its necessitation structure, on the right. The dashed lines represent the *true in virtue of* relation.

$\phi_1, \ldots, \phi_4$ are features $\mathfrak{M}$ in virtue of which claims in A can be true. It is quickly checked that the interpretation function in fig. 6.1 gives a Hertzian semantics: whenever q is inferrable from p in A, the feature of reality in virtue of which p is true necessitates the one in virtue of which q is true, and vice versa.

If a semantics is Hertzian, this means that the structure of the theory and the structure of the part of the metaphysics described by the theory are equivalent (in the category-theoretic sense which we described in ch. 3), although they do not in general have to be isomorphic, since

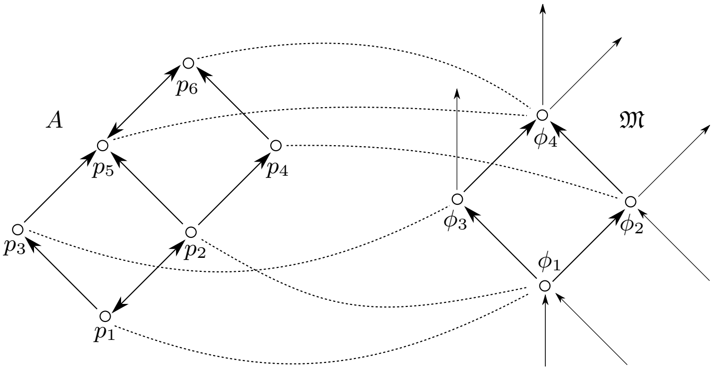


Figure 6.1: ***Hertzian semantics for*** A**.**

neither Hertz-soundness nor Hertz-completeness is sufficient to make $h_{\mathfrak{M}}$ invertible or surjective.

6.2 Necessitarian Semantics are Hertzian

The last section of the previous chapter introduced four types of necessitarian interpretation (seven if you count effectivist versions as separate), and we will now show that these *all* give rise to Hertzian semantics, so long as they are sound and complete. The structural relationship between theory and reality that Hertz required, and which he held we had empirical support for, thus falls out of our methodology without us having to assume it explicitly. What we obtain is a mathematical method for reading off the necessity-structure of the world from the logical (inferential) structure of our best theories.

The easiest variant of necessitarian semantics for which we can prove Hertzianness is correspondence semantics. As its interpretations take every claim p to a uniquely determined single entity $c(p)$, the type of necessitation involved can be taken to be the deterministic kind, according to which $Z \rightarrowtail e$ iff every world in which all entities in the set Z exist also contains the entity e. The theorem is as follows.

Theorem 6.2 : Let $\mathcal{S}(A \mapsto \mathcal{M})$ be a correspondence semantics, and interpret the necessitation of the entity e by the entities in Z as $Z \rightarrowtail b$. Then $\mathcal{S}$ is Hertz-sound iff it is sound, and Hertz-complete iff it is complete.

Proof. We show that $X \vDash_{\mathcal{S}} p$ iff $c[X] \rightarrowtail c(p)$, from which the theorem follows directly, since semantic consequence then coincides with necessitation. But, by definition, $\omega \vDash X$ iff $c[X] \subseteq \omega$, and $\omega \vDash p$ iff $c(p) \in \omega$, so what we need to show is that all worlds that contain all of $c[X]$ also contain $c(p)$ iff $c[X] \rightarrowtail c(p)$. This, in turn, follows directly from the representation theorem for necessitation relations. □

Corollary 6.3 : If a correspondence semantics is sound and complete, we have that

$$p \vdash q \Leftrightarrow c(p) \circ\!\!\rightarrow c(q)$$

We have already remarked that correspondence semantics tends to mirror the theoretical structure directly on the metaphysical, and theorem 6.2 shows how: in a sound and complete correspondence semantics, $X \vdash p$ holds iff the entities that X correspond to together necessitate the occurence of the entity p corresponds to. This means that we can go freely between the logical relation of consequence, and the metaphysical relation of necessitation, since they are equivalent.

These kinds of correspondence semantics are, as we remarked, not that popular anymore, and have in many cases been replaced by versions based on truthmaking. These also display a connection between theory and metaphysics, although on a slightly different structural level than the correspondence semantics. The feature in virtue of which p is true in the world (i.e. model) ω is naturally the existence in ω of any of its truthmakers $\mathit{TM}\,(p)$, and the type of necessitation involved is that which is captured by the distributive necessitation relation $\supset\!\!\!\Rrightarrow$, according to which $Z \supset\!\!\!\Rrightarrow Z'$ iff the existence of any entity in Z guarantees the existence of some entity in Z'.

Hertzianness, however, is about *collective* necessitation. It is easier for us to represent this if we take the metaphysics to be mereological, in which case we can use the cross-sum operator $\otimes$ for this purpose.

Theorem 6.4 : Let $\mathcal{S}(A \mapsto \mathcal{M})$ be a truthmaker semantics, let $\mathcal{M}$ be mereological, and interpret necessitation of what p is true in virtue of by what X is true in virtue of as $\otimes\ \mathit{TM}\,[X] \supset\!\!\!\Rrightarrow \mathit{TM}\,(p)$. Then $\mathcal{S}$ is Hertz-sound iff it is sound, and Hertz-complete iff it is complete.

Proof. It is sufficient to show that $X \models_{\mathcal{S}} p$ iff $\otimes\ \mathit{TM}\,[X] \succ\!\!\!\Rrightarrow \mathit{TM}\,(p)$. But every world in which every claim in X is true must contain a truthmaker for each of these, and because $\mathcal{M}$ is mereological, furthermore a sum of all these truthmakers. Such a sum is always a member of the cross-sum $\otimes\ \mathit{TM}\,[X]$. Conversely, every world that contains some element of the cross-sum $\otimes\ \mathit{TM}\,[X]$ must be one in which all claims in X are true, since the cross-sum necessitates a truthmaker for each X. Thus $\omega \models X$ iff $\otimes\ \mathit{TM}\,[X] \cap \omega \neq \varnothing$, and since $\omega \models p$ iff $\mathit{TM}\,(p) \cap \omega \neq \varnothing$ as well, $X \models_{\mathcal{S}} p$ iff $\otimes\ \mathit{TM}\,[X] \succ\!\!\!\Rrightarrow \mathit{TM}\,(p)$ □

Corollary 6.5 (*Fundamental theorem of truthmaking*) : If a truthmaker semantics is sound and complete, we have that

$$p \vdash q \Leftrightarrow \mathit{TM}\,(p) \supset\!\!\!\Rrightarrow \mathit{TM}\,(q)$$

We have called this corollary the *fundamental theorem of truthmaking* since it is extremely useful for metaphysical investigation whenever we have a truthmaking semantics. Since it only concerns single claims, it does *not* require the metaphysics to be mereological. It is also worth pointing out that all these theorems hold whether the semantics in question is effectivist or not, i.e. whether we take the truthmakers of p to be all those entities whose existence entail p, or only some of them. Since *any* useful semantics needs to be sound, a truthmaker semantics thus at the very least always allows us to read off some of the necessitarian structure of reality from the structure of our true theories. To allow the reading off of *all* such structure, we need completeness as well.

Attempting to weaken our conditions even further, we may use the observation that a mereological truthmaker semantics is equivalent to a plural truthmaker semantics. Let $\underset{pl}{\rightrightarrows}$ be a relation between sets of sets of entities such that $\mathcal{X} \underset{pl}{\rightrightarrows} \mathcal{Y}$ iff any world ω that contains all of some set in $\mathcal{X}$ also contains all of some set in $\mathcal{Y}$. If $\mathcal{X} \underset{pl}{\rightrightarrows} \mathcal{Y}$, the pluralities in $\mathcal{X}$ distributively necessitate those in $\mathcal{Y}$. Because we have plural truthmaking, we do not need to assume the existence of mereological sums. We still, however, need an operation to combine truthmaker sets for different claims, analogous to the cross-sum. We define

$$\bigcup\!\!\!\!\!\!\times\;\mathbf{X} \underset{def}{=} \left\{ \bigcup \mathcal{Y} \;\middle|\; (\forall \mathcal{X} \in \mathbf{X})(\mathcal{Y} \cap \mathcal{X} \neq \varnothing) \right\}$$

Since a plural truthmaker semantics is equivalent to a positive semantics, we get the following generalisation of the Hertzianness theorem:

Theorem 6.6 : Let $\mathcal{S}(A \mapsto \mathcal{M})$ be a positive semantics, and interpret necessitation of what p is true in virtue of by what X is true in virtue of as $\bigcup\!\!\!\!\!\!\times\; \mathit{TMP}\ [X] \underset{pl}{\rightrightarrows} \mathit{TMP}\ (p)$. Then $\mathcal{S}$ is Hertz-sound iff it is sound, and Hertz-complete iff it is complete.

Proof. It is sufficient to show that $X \vDash_{\mathcal{S}} p$ iff $\bigcup\!\!\!\!\!\!\times\; \mathit{TMP}\ [X] \underset{pl}{\rightrightarrows} \mathit{TMP}\ (p)$. The only non-trivial part is to prove that $\omega \vDash_{\mathcal{S}} X$ iff there is a plurality $Z \in \bigcup\!\!\!\!\!\!\times\; \mathit{TMP}\ [X]$ such that $Z \subseteq \mathfrak{M}$, since the theorem then follows from the assumed truth-conditions of plural truthmaker semantics. So assume that $\omega \vDash_{\mathcal{S}} X$. ω must then contain some pluralities $Z_1, \ldots, Z_n$ (strictly, there may be uncountably many of these) such that all of $Z_1, \ldots Z_n$ are included in ω, and such that $Z_1 \in \mathit{TMP}\ (q_1), \ldots, Z_n \in \mathit{TMP}\ (q_n)$ for all $q_i \in X$. The second of these conditions entails that $Z_1 \cup \ldots \cup$

$Z_n \in \mathop{\bigcup\!\!\!\!\!\times} \mathit{TMP}\,[X]$, and the first that $Z_1 \cup \ldots \cup Z_n \subseteq \omega$, so ω contains some plurality in $\mathop{\bigcup\!\!\!\!\!\times} \mathit{TMP}\,[X]$. Conversely, any plurality in $\mathop{\bigcup\!\!\!\!\!\times} \mathit{TMP}\,[X]$ contains the whole of some plurality in each set $\mathit{TMP}\,(q_i)$, for $q_i \in X$. This means that any such plurality is sufficient for the truth of X in any model it is included in. □

Corollary 6.7 : If a positive semantics is sound and complete, we have that

$$p \vdash q \Leftrightarrow \mathit{TMP}\,(p) \underset{pl}{\succ\!\!\!\Rrightarrow} \mathit{TMP}\,(q)$$

Thus, as soon as we have a sound and complete positive semantics $\mathcal{S}(A \mapsto \mathcal{M})$, we can interpret $\mathcal{S}$ as a plural truthmaker semantics, and A's consequence relation as a type of necessitation relation. While we have written this relation in terms of possible worlds, the representation theorem of necessitarian metaphysics guarantees that it can be defined in terms of the relation $\succ\!\!\!\in$ as well.

So long as the semantics is positive, we can thus always find entities to base the truth of claims on, and it then follows that consequence is based on some form of necessitation. It might seem at first that this cannot be done with nonpositive semantics, since the necessitation relations we have used hold between entities, and not between nonentities (whatever that may be). However, using the trick of section 5.6 of transforming our talk of entities into talk about circumstances allows us to go all the way, and show in what way all necessitarian semantics can be said to paint the structure of their theories onto the world.

We said that a circumstance $(X|Y)$ holds in a world ω iff $X \subseteq \omega$ and $Y \cap \omega = \varnothing$. Where Γ and Δ are sets of circumstances, we write $\Gamma \underset{c}{\succ\!\!\!\Rrightarrow} \Delta$ iff, for every world in which some circumstance in Γ holds, some circumstance in Δ holds. Since this is a condition on possible worlds, it can theoretically, through the representation theorem for necessitation relations, be written entirely in terms of the regular necessitation relation $\succ\!\!\!\in$.

A feature in virtue of which p is true is, on this reading, a circumstance that makes p true. For features that make a set of claims true, we introduce the notation

$$\overset{\times}{\bigvee} \mathbf{\Gamma} \underset{def}{=} \{(X|Y) \mid (\forall \Gamma \in \mathbf{\Gamma})(\exists\, (X'|Y') \in \Gamma)(X' \subseteq X \wedge Y' \subseteq Y)\}$$

This means that $\overset{\times}{\bigvee}\mathbf{\Gamma}$, where $\mathbf{\Gamma}$ is a set of circumstances, is the set of circumstances that hold when some circumstance in each set in $\mathbf{\Gamma}$ holds. Taking the necessitation involved in Hertzianness to be our relation $\underset{c}{\Rrightarrow}$, we finally arrive at

Theorem 6.8 : Let $\mathcal{S}(A \mapsto \mathcal{M})$ be any necessitarian semantics, and interpret necessitation of what p is true in virtue of by what X is true in virtue of as $\overset{\times}{\bigvee}$ *TMC* $[X] \underset{c}{\Rrightarrow}$*TMC* (p). Then $\mathcal{S}$ is Hertz-sound iff it is sound, and Hertz-complete iff it is complete.

Proof. As before, we show that $X \vDash_{\mathcal{S}} p$ iff $\overset{\times}{\bigvee}$ *TMC* $[X] \underset{c}{\Rrightarrow}$*TMC* (p), and we show this by proving that $\omega \vDash_{\mathcal{S}} X$ iff there is a circumstance $(Y|Z) \in \overset{\times}{\bigvee}$ *TMC* $[X]$ such that $Y \subseteq \omega$ and $Z \cap \omega = \varnothing$. So assume that there is such a circumstance $(Y|Z)$. By the definition of $\overset{\times}{\bigvee}$, $Y' \subseteq Y$ and $Z' \subseteq Z$ for some circumstance $(Y'|Z')$ in *TMC* (q), for all $q \in X$. Since $Y' \subset Y$ and $Y \subseteq \omega$ entail $Y' \subset \omega$, and $Z' \subset Z$ and $Z \cap \omega = \varnothing$ entail $Z' \cap \omega = \varnothing$, X is true in any model in which $(Y|Z)$ holds.

In the other direction, assume that $\omega \vDash_{\mathcal{S}} X$. Then there are circumstances $(Y_1|Z_1), \ldots, (Y_n|Z_n)$ that make $q_1, \ldots, q_n$ true, where $X = \{q_1, \ldots, q_n\}$, such that $(Y_1|Z_1), \ldots, (Y_n|Z_n)$ hold in ω. Suppose, for contradiction, that there is *no* circumstance $(Y|Z)$ which holds in ω, such that $(Y|Z) \in \overset{\times}{\bigvee}$ *TMC* $[X]$. Then there has to be some q_i in X such that no circumstance in *TMC* (q) holds in ω. But then all of X couldn't have been true, by the truth-condition of general necessitarian semantics. □

Corollary 6.9 : If a necessitarian semantics is sound and complete, we have that

$$p \vdash q \Leftrightarrow \mathit{TMC}\,(p) \underset{c}{\Rrightarrow} \mathit{TMC}\,(q)$$

The theorem builds on the fact that truthmaking by circumstances is equivalent to the "truth supervenes on being" formulation of truthmaking, which is another corollary of this theorem. Any necessitarian semantics which is sound thus allows us to determine the necessity-structure of metaphysics from the structure of a true theory, in a certain sense. The general theme here—that the structure of theory (or language) matches reality—may seem Tractarian. This should perhaps not come as a surprise, considering the common inspiration taken from Hertz's *Principles of Mechanics*. But in the Tractarian form, the principle works on the level of individual thoughts, propositions, and pictures:

> 2.14 That the elements of the picture are combined with one another in a definite way, represents that the things are so combined with one another. This connexion of the elements of the picture is called its structure, and the possibility of this structure is called the form of representation of the picture.
>
> ⋮
>
> 2.16 In order to be a picture a fact must have something in common with what it pictures.
>
> 2.161 In the picture and the pictured there must be something identical in order that the one can picture the other at all.
>
> 2.17 What the picture must have in common with reality in order to be able to represent it after its manner—rightly or falsely—is its form of representation.
>
> ⋮
>
> 4.04 In the proposition there must be exactly as many things distinguishable as there are in the state of affairs, which it represents.
>
> They must both possess the same logical (mathematical) multiplicity (cf. Hertz's Mechanics, on Dynamic Models).

(Wittgenstein, 1922)

Hertz, in the section Wittgenstein refers to here, gives a theory of what it takes for one system to be a *model* of another, and the condition mentioned by Wittgenstein is that both the modelling and the modelled system have to have the same number of coordinates. Hertz furthermore requires that the *equations* for the systems should be identical, and that the magnitude of displacements agree in them (Hertz, 1899, §418). These are clearly conditions necessary for one system to be able to give *information* about the other. In particular, a system with less degrees of freedom can never be used to describe one with more. This is the basis for Wittgenstein's idea that the proposition, if it is to be able to describe the world, must have *some* structural similarity with reality, even if this similarity does not have to be immediately visible.

But reality itself is not determined very strongly by such similarities, especially if we allow that it may have non-empirical aspects. Hertz puts it as follows, in a passage that seems very close to our own characterisation of models as constrained, but also underdetermined, by theory:

> We can then, in fact, have no knowledge as to whether the systems which we consider in mechanics agree in any other aspect with the actual systems of nature which we intend to consider, than in this alone,—that the one set of systems are models of the other. [...]
>
> The relation of a dynamical model to the system of which it is regarded as the model, is precisely the same as the relation of the images which our mind forms of things to the things themselves. For if we regard the condition of the model as the representation of the condition of the system, then the consequents of this representation, which according to the laws of this representation must appear, are also the representation of the consequents which must proceed from the original object according to the laws of this original object. The agreement between mind and nature may therefore be likened to the agreement between two systems which are models of one another, and we can even account for this agreement by assuming that the mind is capable of making actual dynamical models of things, and working with them. (Hertz, 1899, §§427–428).

Wittgenstein interpreted this to mean that the proposition has to be a picture of the fact, and since the picture is *structural*, both the proposition and the fact have to be complex. But this is not the only level on which we can impose the requirement. We have not assumed claims to have a structure at all. *If* they have an algebraic structure, we can infer that the metaphysics has such a structure as well, as we will see in the next section. But even without an internal structure in the claim, the inferential structure of the entire theory mirrors itself onto the metaphysics. As we have shown, it *has to*, if it is to have the ability to describe reality truthfully.

6.3 Algebraic and Probabilistic Theories

In the last section, we proved that the inferential structure of a theory matches the necessitation-structure of its metaphysics. But this also extends to structure that does not explicitly concern consequence relations or necessity. Algebraic structure is preserved by necessitarian semantics as well. First we prove a general result, from which we then can find the exact algebraic operations in the metaphysics that operations in a theory correspond to.

Theorem 6.10 : Assume that the theory A is formalised self-extensionally by the algebra $\mathfrak{A} = \langle L_A, f_1, \ldots, f_n \rangle$. Let $\mathcal{S}(A \mapsto \mathcal{M})$ be a sound and complete necessitarian semantics with global interpretation function h. Then there is an algebra $\mathfrak{B} = \langle h[L_A], g_1, \ldots, g_n \rangle$ on the image of L_A under h, of the same signature as $\mathfrak{A}$, such that h is a homomorphism onto $\mathfrak{B}$.

Proof. That h is a function onto $\mathfrak{B}$ is trivial from its definition. Define the operations $g_1, \ldots, g_n$ through the identities

$$g_i(y_1, \ldots, y_m) \underset{def}{=} h(f_i(x_1, \ldots, x_n))$$

where $x_k \in h^{-1}(y_k)$. For this to work, we must have that $\ker h$ (the equivalence relation that holds between x and y iff $h(x) = h(y)$) is a congruence on $\mathfrak{A}$. So assume that f is an operation on $\mathfrak{A}$, and that there are elements $p_1, \ldots p_k, \ldots, p_n$ and p'_k such that $h(p_k) = h(p'_k)$. Since $\mathcal{S}$ is Hertzian, we have that $h(p_k) = h(p'_k)$ entails that $p_k \dashv\vdash p'_k$. But the self-extensionality condition then gives us that p_k is congruent with p'_k, so the homomorphism is well-defined. □

Thus, even the algebraic structure of a true theory is interpretable as the algebraic structure of the metaphysics, if we use necessitarian semantics. For example, we can use this fact to prove that in any truthmaker semantics for a theory based on classical logic, there must be a one-to-one correspondence between sets of truthmakers that corresponds to *negation* in the theory. Which specific correspondence this is, is given in the following theorem.

Theorem 6.11 : Let A be a theory that contains classical logic, and $\mathcal{S}(A \mapsto \mathcal{M})$ a sound and complete truthmaker semantics for A. Let $\perp\!\!\!\perp$ be a binary relation on $\wp(E_{\mathcal{M}})$ such that, for every $X, Y \subseteq E_{\mathcal{M}}$, $X \perp\!\!\!\perp Y$ iff

- (*i*) $\{x, y\} \mathrel{\succ\!\!\Lleftarrow} \varnothing$ for every $x \in X$ and $y \in Y$, and
- (*ii*) $\varnothing \mathrel{\succ\!\!\Lleftarrow} X \cup Y$.

Assuming that not every sentence in L is true, it then follows that $\mathit{TM}\,(p) \perp\!\!\!\perp \mathit{TM}\,(q)$ iff $q \dashv\vdash_A \neg p$.

Proof. An operation $\neg$ on a distributive lattice $\langle L_A, \wedge, \vee \rangle$ is a classical negation iff it satisfies $C_A(\{p \wedge \neg p\}) = L_A$ and $p \vee \neg p \in \top_A$. By Hertzianness, every world must then contain something in $\mathit{TM}\,(p)$ or $\mathit{TM}\,(\neg p)$, so (*ii*) follows. Furthermore, since not every sentence is true, there is no world in which both p and $\neg p$ is true, so (*i*) is fulfilled as well. The converse is trivial. □

The relation $\perp\!\!\!\perp$ is called *strong orthogonality*, in contrast to the *weak* orthogonality of section 4.4. While weak orthogonality captures a relation that holds between the truthmakers of incompatible claims, strong orthogonality captures that which holds when they are *complements* of one another.[1]

The presence of a metaphysical correlate of classical negation such as the relation $\perp\!\!\!\perp$ has interesting consequences. Let a *dichotomy* of a necessitarian metaphysic $\langle E, \gtrless \rangle$ be a pair of functions $S, \overline{S}$ from an index set I to subsets of E, such that

$$S(i) \perp\!\!\!\perp \overline{S}(i)$$

for all $i \in I$. A dichotomy splits the possible entities E into $2^{|I|}$ non-overlapping sets, and no world can contain entities from more than one of each pair $S(i)$, $\overline{S}(i)$. Thus the dichotomy gives rise to an equivalence relation $\equiv$ on worlds such that $\omega_1 \equiv \omega_2$ iff

$$\omega_1 \cap S(i) = \varnothing \Leftrightarrow \omega_2 \cap S(i) = \varnothing$$

holds, for all $i \in I$. Call a necessitarian semantics $\mathcal{S}(A \mapsto \mathcal{M})$ *dichotomous* iff there is a dichotomy on $\mathcal{M}$ such that $\omega_1 \equiv \omega_2$ and $\omega_1 \models p$ together entail that $\omega_2 \models p$, for all $p \in L_A$. In a dichotomous semantics, the dichotomy can thus be used to specify the identity of any world up to elementary equivalence.

Classical logic with truthmaker semantics is dichotomous: let L'_A be the sublanguage of L_A that contains the sentences with an even number of negations first. Then the functions $S : L'_A \to \wp(E)$ and $\overline{S} : L'_A \to \wp(E)$ defined as

$$S(p) = TM\,(p)$$

$$\overline{S}(p) = TM\,(\neg p)$$

form a dichotomy of $\mathcal{M}$.

[1] A complement of an element c in an arbitrary lattice with top 1 and bottom 0 is some element c' such that $c \wedge c' = 0$ and $c \vee c' = 1$. In a Boolean lattice, complements are unique, and correspond exactly to the logical notion of negation.

Theorem 6.12 : A dichotomous semantics is both positive and negative.

Proof. Let $S : I \to \wp(E), \overline{S} : I \to \wp(E)$ be a dichotomy of $\mathcal{M}$ such that $\omega_1 \equiv \omega_2$ and $\omega_1 \vDash p$ entail that $\omega_2 \vDash p$. We show that if $\omega_1 \subseteq \omega_2$, then $\omega_1 \vDash p$ iff $\omega_2 \vDash p$, for any claim p in the language. But for any $i \in I$, we have that if $\omega_1 \cap S(i) \neq \varnothing$ then $\omega_2 \cap S(i) \neq \varnothing$, and if $\omega_1 \cap S(i) = \varnothing$, then $\omega_1 \cap \overline{S}(i) \neq \varnothing$, so $\omega_2 \cap \overline{S}(i) \neq \varnothing$. Thus $\omega_1 \equiv \omega_2$, and since the semantics is assumed to be dichotomous, the same claims are true in ω_1 and ω_2. □

For probabilistic theories, we need to make a slight generalisation of our concepts of Hertz-soundness and Hertz-completeness if these are to be applicable. A probabilistic semantics interprets $X \vdash^{\pi} p$ as "the proportion of p-models among the X-models is π". But just as with regular consequence, this is something that concerns *all* models, and not only the actual one. Probabilistically necessitarian metaphysics allow us to descend from the inter-model perspective to an intra-model one.

Since generalisation to positive and general necessitarian semantics proceeds much as in the non-probabilistic case, we focus on truthmaker semantics. The proper characterisation of Hertzianness in this case would be that

$$X \vdash^{\pi} p \Leftrightarrow \otimes \; TM\,[X] \overset{\pi}{\rightrightarrows} TM\,(p)$$

i.e. that any truthmaker of the whole of X necessitates to a degree π that some truthmaker for p exists. By using the representation theorem for probabilistic necessitation, we can prove that this indeed holds iff the semantics is sound and complete: probabilistic necessitation is intertranslatable with a probability measure on the set of possible worlds, and since possible worlds are models, this is equivalent to a probability measure on the model space.

An interesting point is that this gives us two different viewpoints from which to look at the same facts. As we mentioned in chapter 5, probabilistic semantics gives us a kind of frequency interpretation of the probability concept, according to which $P(Y|X)$ is the relative frequency of Y-models among the X-models. It is not a *purely* frequentistic account, however, since necessitarian metaphysics generally do not come with an order. This means that a *limiting* frequency cannot be defined, unless we impose such an order explicitly.

A probabilistic necessitation relation, however, is rather a kind of *propensity*, to use Popper's term (Popper, 1959). Or, at least, it can be interpreted that way: if $X \overset{\pi}{\succ\!\!\Rrightarrow} Y$, then the X's collectively have the propensity π to produce some Y. This can be seen as a property of the X's, even if a relational one. Unless it is actually manifested, it does not depend on anything outside the X's themselves.

But probabilistic necessitation does not fit *all* forms of propensity theory. Since a probabilistic necessitation relation is intertranslatable with a probability measure, it is tied to the standard probability calculus. But there are arguments that propensity should *not* conform to these axioms if it is to describe a chancy disposition. As Humpreys writes,

> Consider first a traditional deterministic disposition, such as the disposition for a glass window to shatter when struck by a heavy object. Given slightly idealized circumstances, the window is certain to break when hit by a rock, and this manifestation of the disposition is displayed whenever the appropriate conditions are present. Such deterministic dispositions are, however, often asymmetric. The window has no disposition to be hit by a rock when broken, and similarly, whatever disposition there is for the air temperature to go above 80°F is unaffected by whether my neighbor loses his temper, even though the converse influence is certainly there. (Humphreys, 1985, p. 558).

Unless they involve events with zero probability, probabilistic relationships are always "invertible" using Bayes's theorem, but perhaps we should not expect to be able to invert propensities in the same way. However, it is easy to see that if $X \overset{\pi}{\succ\!\!\Rrightarrow} Y$, then generally $Y \overset{\pi'}{\succ\!\!\Rrightarrow} X$ for some value π'. Thus, at least not all instances of probabilistic necessitation are manifestations of a chancy disposition.

This problem can be solved by using the notion of *basis* introduced in section 4.3. The probabilistic necessitation relation, just as the nonprobabilistic variant, mixes together all forms of necessity relationships. It may be very difficult to separate out the causal aspect alone. Unlike in the nonprobabilistic case, we cannot always take part of a probabilistic necessitation relation and extend it in a unique “minimal” way to satisfy the axioms.

This can be illustrated by ordering the possible entities in a time series. Let $t : E \to \wp(\mathbb{R})$ be a function from the entities to the set of points in time when they exist. Taking $\overset{\pi}{\succ\!\!\Rrightarrow}$ to be probabilistic causality, we assume that $X \overset{\pi}{\succ\!\!\Rrightarrow} Y$ is defined for non-empty pairs of intervals X, Y such that

$$\sup_{x \in X} t(x) = \inf_{y \in Y} t(y)$$

This may indeed be the most that one could ask for in a probabilistic theory: given how the world has been so far, what are the probabilities that it will be a certain way hereafter? But the so-called *initial conditions* are not included here, which mirrors itself in the fact that we have not defined $\varnothing \overset{\pi}{\succ\!\!\Rrightarrow} Y$. This concept does not, in itself, require that there has to be a “first moment” in time, as Hume pointed out in *Dialogues concerning natural religion* (Hume, 1779, part IX). As Demea put it in the dialogue, even if time went infinitely far back so that any event had a sufficient cause, we would still want to know why the entire series occurred, rather than some other series. Translated to our case, we note that even if every entity’s probability is determined by the entities before it, we cannot assign these probabilities without knowing how to do so to various *initial segments* of the world.

In fact, the very concept of *time* is something of a red herring here. We can envisage things happening later in the time series as well, for which we cannot give a probability. An example of this, which we will discuss in the next chapter, appears in quantum mechanics. In its classical form, QM does not allow one to calculate the probabilities that certain measurements are made, but only probabilities of various results of these measurements. The sequence of measurements can thus itself be seen as part of the “initial conditions”, even if they may occur now and then during the entire lifetime of the universe.

Due to the strength of probabilistic necessitarian semantics, completeness (and thus also Hertz-completeness) may be too much to ask for in general. Most probabilistic theories do not specify probabilities for every inference in their language. Therefore it would be interesting to find some weaker form of completeness, which still captures the fundamental idea that what follows semantically should follow syntactically (or theoretically) as well. We will not attempt to do so here, however.

6.4 Ontological Commitments

So far, we have mainly studied the relationship between a theory and a metaphysics $\mathcal{M}$, which can be taken as a selection of ways the world could be. But it should be obvious that if the world actually is one of the ways it could be (which is guaranteed if the theory we have used is true, and the semantics is appropriate), then one of the models of the metaphysics will not only be possible, but actual. This means that a true theory's structure not only imposes itself on its metaphysics, but also on the actual world.

One of the most fundamental questions we can ask about the world's structure concerns what exists, given a certain theory's truth, and the cluster of issues around this is known as the problem of *ontological commitment*. There are numerous aspects of it, and we will try to separate them somewhat in order to be able to give a more systematic treatment.

First of all, we have the question of what ontological commitment is a property of. It is usual to take it to pertain to some sort of claims (i.e. theories, sentences, beliefs etc.), but these do not, on their own, determine what the world is like. Only when they are given a semantics do they have metaphysical import, and this import is captured by the models that they are true in. The more fundamental ontological properties thus pertain to models, and only derivatively to claims.

We have already stressed the choices involved in selecting a semantics, as well as those we make when we decide on a model space to use – no claim ever interprets itself, and no theory determines its own semantics. This means that ontological commitments *always* are relative to a semantics, and thus also a model space. But a model space is a kind of theory, so ontology is theory-relative. This is, of course, nothing other than Quine's position in *Ontological Relativity*, which we mentioned in chapter 1:

> What makes sense is to say not what the objects of a theory are, absolutely speaking, but how one theory of objects is interpretable or reinterpretable in another. (Quine, 1969, p. 50)

For us, however, this is not the end of ontology, but rather the start of it. Ontological relativity may be true, but it does not make ontology any less interesting or important, just as the relativity of most geometrical concepts does not make geometry any less profound or powerful.

We have already noted that a semantics can be taken to be a kind of translation between a theory's logical and metaphysical points of view. The metaphysical point of view, in turn, corresponds to Quine's "theory of objects". Since the purpose of determining a theory's ontological commitments is to obtain an inventory of what objects exists according to that theory, one way to see the problem of ontological commitment is as being about the translation of model spaces into $\mathcal{V}$.

The advantage of $\mathcal{V}$ is that each of its models has an explicit ontology of well-behaved, well-individuated objects. This, however, means that not all model spaces may be usefully interpretable in such a way. Since a category which is concrete over $\mathcal{V}$ is called a construct, we say that $\mathcal{M}$ is *constructible* if there is a faithful functor $F : \mathcal{M} \to \mathcal{V}$.

An example of a model space (or rather, a category) that is not constructible is hTop, whose objects are topological spaces, and whose morphisms are homotopy classes of continuous functions between these.[2] But the largest problem with using constructibility in order to decide questions of ontological commitment is not the existence of the occasional inconstructible model space, but the arbitrariness involved in imposing a forgetful functor F. Usually, several such functors are identifiable, and the question of which one gives the "true" ontology of the space's models therefore becomes acute.

[2] Two functions are *homotopic* if they can be continuously deformed into one another. The space hTop thus consists of topologies, but disregards certain differences between transformations between these.

One may, of course, hold with Quine that "...the question of the ontological commitment of a theory does not properly arise except as that theory is expressed in classical quantificational form, or insofar as one has in mind how to translate it to that form" (Quine, 1969, p. 106), but we have already, at the outset, made clear our intention to break free of his reliance on classical first-order logic as the "one true logic". One way to do this is to resist the temptation to reinterpret the model space we started with, and instead use the category-theoretic concepts developed in chapter 3 to approach the problem.

Even when we limit ourselves to models in the same model space, there are absolute and relative notions of ontological commitment. In the absolute sense, we can ask "does $\mathfrak{M}$ contain X's"? Alternatively, we may wish to know if one model contains anything *more* than what is contained in another. For instance, we may be interested in the question of whether acceptance of mereological sums inflates our ontology, or whether a reduction of one theory to another also reduces its ontological commitments. In this case, we are using a *relative* interpretation of the concept, in the sense that it is based on a relation between models.[3]

The relative concept, in turn, splits into several, depending on what we mean by one model containing "more than" another. In chapter 3 we identified three relevant relationships here. The first, and strongest, is that which holds iff everything that is in $\mathfrak{M}_1$ is also in $\mathfrak{M}_2$. In this case, we say that $\mathfrak{M}_1$ is *contained* in $\mathfrak{M}_2$. This is also, roughly, the same as saying that $\mathfrak{M}_1$ is part of $\mathfrak{M}_2$. It is expressed by the condition that there is a *canonical* strong monic from $\mathfrak{M}_1$ to $\mathfrak{M}_2$. Since which strong monics are canonical is dependent on which inclusion system we have placed on $\mathcal{M}$, containment of models is relative to such a system.

Luckily, metaphysics usually concerns itself not with the existence of specific entities, but of types of entity. Thus it may be more interesting to ask whether all the *structure* which is in $\mathfrak{M}_1$ is also in $\mathfrak{M}_2$. Mathematically, this means that there is an embedding of $\mathfrak{M}_1$ in $\mathfrak{M}_2$, and seen this way, $\mathfrak{M}_2$ contains as least as much as $\mathfrak{M}_1$ iff $\mathfrak{M}_1$ can be *embedded*

[3] One might also say *doubly* relative, since we are relative to a model space as well. But since all forms of ontological commitment are relative in that way, we use "relative" for the concept of ontological commitment that takes the form of a relation between models.

in $\mathfrak{M}_2$. Categorically, we have decided to explicate this as the condition that there is a strong monic $m : \mathfrak{M}_1 \to \mathfrak{M}_2$. Because this notion is completely category-theoretic, embeddings are fully structural.

Finally, we can also ask a purely numerical question: does $\mathfrak{M}_2$ contain a larger number of things than $\mathfrak{M}_1$? When models are just sets, such as in $\mathcal{V}$, this relationship coincides with the embedding concept, but they come apart in most other spaces. When we interpret the "containing more" clause as simply being about numerosity (it would be a bit misleading to say "cardinality" here, since cardinality is so tied up with the concept of set), we can express it through the existence of a (possibly non-strong) monomorphism from $\mathfrak{M}_1$ to $\mathfrak{M}_2$.

Which of the absolute or relative concepts of ontological commitment is the fundamental one? In certain simple model spaces such as $\mathcal{V}$, it does not matter which one we begin with. Since, in $\mathcal{V}$, models are sets, we can say that $\mathfrak{M}$ contains X's iff $\mathfrak{M} \cap X \neq \varnothing$, and that $\mathfrak{M}_2$ contains at least the things in $\mathfrak{M}_1$ iff for any set X, $\mathfrak{M}_1 \cap X \neq \varnothing$ implies that $\mathfrak{M}_2 \cap X \neq \varnothing$. But this is just the condition $\mathfrak{M}_1 \subseteq \mathfrak{M}_2$.

Already in $\mathcal{T}$, matters are not quite so easy. Does $D_{\mathfrak{M}_1} \subseteq D_{\mathfrak{M}_2}$ imply that $\mathfrak{M}_2$ contains everything that $\mathfrak{M}_1$ contains, for instance? Not necessarily, since different relations may hold in $\mathfrak{M}_1$ and $\mathfrak{M}_2$, and we may be reluctant to say that $\mathfrak{M}_2$ contains everything that $\mathfrak{M}_1$ contains if the things in their domains are radically different in the two models. This is just an instance of the intensionality of Tarskian models that we remarked on in ch. 3: how things are in the model is affected by how they are described.

How do we then determine whether a model contains objects of a given type, or when a model contains another? Part of our difficulty stems trying to use the model-space relative notion of *object* for something it is not fit for. From the viewpoint of $\mathcal{M}$, the models in its object class are the only things that can have self-subsistent existence, and in a certain sense therefore the only things worthy of being called objects. When we relativise ontological commitment to a model space, we should therefore relativise the object concept as well.

The easiest way to do this seems to be to express "$\mathfrak{M}$ contains X's" as "$\mathfrak{M}$ contains some model in $\mathcal{X}$", where $\mathcal{X}$ is a set of objects of $\mathcal{M}$. This containment, in turn, is to be explicated in terms of the existence of a canonical strong monic $m : \mathfrak{X} \to \mathfrak{M}$, where $\mathfrak{X} \in \mathcal{X}$. Thus the only things we can really be committed to are models.ins some model in $\mathcal{X}$", where $\mathcal{X}$ is a set of objects of $\mathcal{M}$. This containment, in turn, is to be explicated in terms of the existence of a canonical strong monic $m : \mathfrak{X} \to \mathfrak{M}$, where $\mathfrak{X} \in \mathcal{X}$. Thus the only things we can really be committed to are models.

This gives rise to a few interesting corollaries. Suppose, as a mathematical structuralist might claim, that the natural numbers make sense only in the context of a natural number system (or a "simply finite system", as Dedekind would have expressed it). Then there is no such thing as being committed to, say, the even numbers, and not the odd numbers. This seems reasonable enough. But suppose we have an Aristotelian metaphysics, in which properties cannot exist on their own, and we have a theory that says that property P exists. In this metaphysics, there is no model that contains *just* P, so the commitment cannot be to $\{P\}$. Instead, we must conceive of it as a commitment to *some* member of the set $\{\{P, a\}, \{P, b\}, \{P, c\}, \ldots\}$ where $a, b, c, \ldots$ are all possible particulars, or even $\{\{P, a, P(a)\}, \{P, b, P(b)\}, \{P, c, P(c)\}, \ldots\}$ where $P(a), P(b), P(c)$ etc. are the facts that a is P, b is P, etc., or some "non-relational tie" of instantiation between P and a particular. The only way to work around this appears to be to embed the model space into a completion of it, thereby introducing the "ideal models", or aspects, which we mentioned in section 3.3. How to do this in detail is far from trivial, however.

Since the only things we can be committed to are models, the relative notion of commitment is more fundamental than the absolute on the single-model level: we need to know when one model contains another in order to be able to say which things are in which models. But on the level of claims, relative commitment turns out to be much more complicated than absolute.

Starting with the absolute commitments of a claim, this concept can be split into a specific one, and a general one. The *specific* commitments of p are those things that are in every model in p. Formally, holding p to be true specifically commits one to the $\mathcal{X}$'s iff the models in $\mathcal{X}$ are embeddable in every model in which p is true. This means that p's truth guarantees the existence of all the $\mathcal{X}$'s, and we write the specific commitments of p as $SC(p)$.

Many claims do not come with specific commitments. “There are gnus” commits one to some gnu, but not any specific one. “Socrates exists” may be an example of a claim with non-empty specific commitments, but even this could be questioned if one has doubts about the identity of Socrates across models or possible worlds.

General commitment is often a much more useful concept, although it turns out to be harder to formalise. Consider, for instance, the claim $g =$“there are gnus”. This does not commit us to any specific gnu, and although it may be held to commit us to *gnuhood*, this is not necessarily so either. If there is something G in every possible model in which there is a gnu, and in no others, then G is a candidate for playing the role of gnuhood. But *any* given gnu can play the role of being a gnu, and it is not really necessary that there be something that “collects” them.

The general commitments of p can be seen as the set of *roles* p requires to be fulfilled. To formulate this properly, it is useful to formalise the role concept. For any two models $\mathfrak{M}, \mathfrak{M}'$, write $\mathfrak{M} \hookrightarrow \mathfrak{M}'$ if there is some strong monic from $\mathfrak{M}$ to $\mathfrak{M}'$. Let a *role* R be a binary relation on $\mathcal{M}$ such that if $\mathfrak{M}R\mathfrak{M}'$, then $\mathfrak{M} \hookrightarrow \mathfrak{M}'$. The intended interpretation is that $\mathfrak{M}R\mathfrak{M}'$ iff $\mathfrak{M}$ can play the role R in the model $\mathfrak{M}'$. The strong monic condition ensures that $\mathfrak{M}$ can be a part of $\mathfrak{M}'$, even if it does not guarantee that it actually *is* a part.

Let $\mathcal{R}(\mathcal{M})$ be the set of all roles on the model space $\mathcal{M}$. We give the formal definition of the general ontological commitment $GC(p)$ of a claim p as follows.

$$GC(p) \underset{def}{=} \{R \in \mathcal{R}(\mathcal{M}) \mid (\forall \mathfrak{M} \in [\![p]\!])(\exists \mathfrak{M} \in \mathcal{M})\mathfrak{M}'R\mathfrak{M}\}$$

The motivation behind this definition is the idea that p’s truth requires all roles that are fulfilled in the models of p to be played by *something.* It is in this sense we ask whether p commits one to the existence of numbers (rather than *the* numbers), physical objects, propositions, etc. For instance, holding p to be true commits one to numbers iff $R_{\mathbb{N}} \in GC(p)$, where

$$\mathfrak{M}R_{\mathbb{N}}\mathfrak{M}' \text{ iff } \mathfrak{M} \text{ plays the role of the natural numbers in } \mathfrak{M}'$$

This role-playing can then be explicated in terms of satisfaction of the Peano-Dedekind axioms. Alternatively, we can give a purely category-theoretic definition, and state that $\mathfrak{M}R_{\mathbb{N}}\mathfrak{M}'$ iff some *natural number object* (Lawvere, 1964) is embeddable in $\mathfrak{M}$. Category-theoretically, commitment to numbers is a *purely* structural property – it does not involve any "internal" attributes of objects.

We may distinguish between *external* and *internal* roles, where we call a role R internal iff $\mathfrak{M}R\mathfrak{M}'$ and $\mathfrak{M}' \hookrightarrow \mathfrak{M}''$ imply that $\mathfrak{M}R\mathfrak{M}''$, and external otherwise. Internal roles are stable under embeddings. One simple example is $\mathfrak{M}$ playing the role of $\mathfrak{M}$ *itself*, which it does when it fulfils the specific commitments of a claim such as "$\mathfrak{M}$ exists". Here, $\mathfrak{M}$ can play the same role in all models in which it is included.

For a possibly external ontological commitment, consider "there is a largest number". One could argue that this commits one to not only some number which is the largest, but also to a *lack* of numbers larger than it. But it is not certain that this should count as an *ontological* commitment, since it does not strictly say that something exists, but also that some things do *not* exist.

Using GC, the relative questions become easy to answer. Say that the role R is *filled* in $\mathfrak{M}$ iff there is some $\mathfrak{M}'$ such that $\mathfrak{M}'R\mathfrak{M}$. q commits one to *at least as much* as p iff $GC(p) \subseteq GC(q)$, and they have the same ontological commitments iff $GC(p) = GC(q)$. This entails the useful theorem that if p entails q, then p's general ontological commitments are at least as large as q's.

Theorem 6.13 : If $[\![p]\!] \subseteq [\![q]\!]$, then p is committed to as least as much as q.

Proof. Let R be any role that q is committed to. Then R is filled in all models in which q are true. But since $[\![p]\!]$ is included in these, R must be filled in these as well. □

Unfortunately, if we do not impose any further conditions on what roles can be, we also have the converse theorem: if $GC(p) \subseteq GC(q)$, then q entails p.[4] Usually we are not interested in *all* roles, however. We may, for example, have a certain set of concepts in mind, in which case these delimit which models can play which roles. In a Fregean spirit, we could hold that Julius Caesar *cannot* play the role of the number 2, since he is a Roman emperor and not a number. We are then only considering a subset of $\mathcal{R}(\mathcal{M})$ to be relevant, and the roles that a claim can be committed to become different.

One important limitation of $\mathcal{R}(\mathcal{M})$ is to consider only internal roles. This is motivated by the thought that ontological commitment is about what *exists*, and not what does *not* exist. Even if "there are no warthogs" is true only in models that lack warthogs, this does not commit us to a *lack* of warthogs in any ontologically significant sense. In the following, we will therefore assume that all roles under consideration are internal.

A further possibility for strengthening our definition is to require the embeddings used in defining a role to be canonical. Let us call a role R *canonical* iff $\mathfrak{M}R\mathfrak{M}'$ entails that at least one of the strong monics from $\mathfrak{M}$ to $\mathfrak{M}'$ is canonical. A canonical role specifies the exact identity of the things that can play it, and not only their structure. The disadvantage of using such roles, however, is that they require us to have access to a specification of which strong monics are canonical.

For internal roles, we can simplify the structure of general ontological commitment somewhat. Since a model can play the internal role R in *any* model it is part of, the role itself can be seen as a set of models, rather than a relation. Thus, we will also say that a claim p commits one to X's iff X is a set of models, and any model in which p is true contains some X. This is equivalent to there being a role R such that $\mathfrak{M}R\mathfrak{M}'$ iff $\mathfrak{M} \in X$ and $\mathfrak{M} \hookrightarrow \mathfrak{M}'$. We can therefore also see general commitment as a function $GCS(p)$, whose values are sets of sets of models, with the interpretation that $\mathcal{X} \in GCS(p)$ iff p commits one to $\mathcal{X}$'s. Symbolically, we have as the definition

$$GCS(p) \underset{def}{=} \{\mathcal{X} \subseteq \mathcal{M} \mid (\forall \mathfrak{M} \in [\![p]\!])(\exists \mathfrak{M}' \in \mathcal{X})(\mathfrak{M}' \hookrightarrow \mathfrak{M})\}$$

[4] This can be shown by considering the role R defined by the condition that $\mathfrak{M}R\mathfrak{M}$ iff $\mathfrak{M} \in [\![q]\!]$.

Theorem 6.14 : For all $\mathcal{X} \in GCS(p)$, if $\mathfrak{M} \in \mathcal{X}$ and $\mathfrak{M} \hookrightarrow \mathfrak{M}'$, then $\mathfrak{M}' \in GCS(p)$. If $\mathcal{X} \in GCS(p)$ and $\mathcal{X} \subseteq \mathcal{X}'$, then $\mathcal{X}' \in GCS(p)$.

Proof. Trivial using set theory and the fact that the composition of strong monics is a strong monic, which entails that embeddability is transitive. □

Theorem 6.15 : $p \vdash q$ entails $GCS(q) \subseteq GCS(p)$, but $p \nvdash q$ does not entail $GCS(q) \nsubseteq GCS(p)$.

Proof. The first part follows directly from thm 6.13 by the correspondence that $R \in GC(p)$ iff $\mathsf{dom}(R) \in GCS(p)$ for internal roles. For the second part, let $\mathcal{M}$ be a model space whose models are non-empty subsets of $\{a, b\}$, and whose only non-trivial embeddings are $\{a\} \hookrightarrow \{a, b\}$ and $\{b\} \hookrightarrow \{a, b\}$. Let $q =$"a or b exist" and let $p =$"a or b, but not both, exist". Under the usual semantics $p \vdash q$ holds, and therefore $GCS(q) \subseteq GCS(p)$. On the other hand, $q \nvdash p$, since q is true in the model $\{a, b\}$ but p is not.

There are three models $u = \{a\}, v = \{b\}$ and $w = \{a, b\}$, and under the usual semantics $[\![p]\!] = \{u, v\}$ and $[\![q]\!] = \{u, v, w\}$. Applying the definition of GCS, we find that

$$GCS(q) = \{\{u, v\}, \{u, v, w\}\}$$

But both of these sets are in $GCS(p)$ as well, since both of them contain some model embeddable in u and some model embeddable in v. Thus $GCS(p) = GCS(q)$. □

Thus each set (i.e. role) in $GCS(p)$ is closed upwards under embeddings, and the whole set of roles is closed upwards under the subset relation. This is due to the facts that any model that contains a model playing the internal role R can itself play that role, and that the filling of all roles in a set ipso facto is the filling of the roles in all its subsets. Furthermore, we do no longer have the trivialising entailment that $GCS(p) \subseteq GCS(q) \Rightarrow p \vdash q$, since not all increases in strength of a claim incur corresponding increases in ontological commitment.

Although they unfortunately are quite opaque, our definitions do seem to capture what we are after. Consider the case where $g =$ "there are gnus" and $gw =$ "there are gnus and warthogs". This is a classic case of one claim intuitively having larger ontological commitments than another, even though both claims' specific commitments are empty. Since $gw \vdash g$, we have that $GC(g) \subseteq GC(gw)$. To show that the converse does not hold, we need to find a role R which is filled in every model in which gw is true, but is unfilled in some model g is true in. Assuming that gnus and warthogs can exist on their own, and also that no gnu ever can be a warthog, we can take each possible gnu and each possible warthog to make up a model of their own. Let $\mathfrak{M}R\mathfrak{M}'$ iff $\mathfrak{M}$ is a warthog model and $\mathfrak{M}$ is embeddable in $\mathfrak{M}'$. By definition, R is internal, and since there are models that contain gnus but not warthogs, R is unfilled in these. Therefore $GC(gw) \nsubseteq GC(g)$.

The framework we have outlined here allows us to approach questions regarding ontological commitments systematically, without presupposing what the model space we are investigating is like. When we do know this, there may be a few more things we can say, as we shall see in the next section.

6.5 Commitment in a Necessitarian Semantics

In one sense, questions of ontological commitment are easy when we are using necessitarian semantics. $\mathcal{N}$ is a construct, and it thus comes with a built-in translation F to $\mathcal{V}$. Thus we can say that a claim p is committed to X's iff the set of all possible X's intersects $F(\mathfrak{M})$, for every model $\mathfrak{M}$ in which p is true. We can say that p has as least as large ontological commitments as q iff p being committed to X's implies that q is committed to X's, for any set X of possible entities. As always, though, the devil is in the details.

First of all, we should ensure that speaking of commitment to *entities* in a necessitarian semantics, as opposed to speaking of commitment to models, actually makes sense. But we can, if desired, translate between the two ways of speaking as soon as we have the forgetful functor F. Let $\mathcal{M}_F(X)$, where X is a set of possible entities, be the set of all models $\mathfrak{M}$ such that $F(\mathfrak{M}) \cap X \neq \varnothing$. Then we can say that p commits one to X's generally iff it commits one to the set $\mathcal{M}_F(X)$ generally, and this allows us to define versions of GCS and SC expressed in terms of commitment to entities. Letting $\mathcal{M}$ be a necessitarian metaphysics with set E of possible entities, we define GCE and SCE through

$$\begin{aligned} SCE(p) &= \mathcal{M}_F^{-1}[SC(p)] \\ &= \bigcap M_F^{-1}\left[[\![p]\!]\right] \\ GCE(p) &= \mathcal{M}_F^{-1}[GCS(p)] \\ &= \{X \subseteq E \mid (\forall \mathfrak{M} \in [\![p]\!])(E_{\mathfrak{M}} \cap X \neq \varnothing)\} \end{aligned}$$

$SCE(p)$ may be read as "the entities that p commits one to", and $GCE(p)$ as "the types of entity that p commits one to", where these "types" are represented by the sets of their instances. It is important to remember that M_F, and thus also SCE and GCE, depend crucially on the forgetful functor F. This functor determines which embeddings are canonical, and thus also what it means for one model to be part of another, rather than merely embeddable therein.

For $\mathcal{N}$ there is a very natural forgetful functor, and thus a natural choice of which monics are canonical. On the other hand, we should always be watchful of "naturalness". Sometimes, focusing on what seems natural hinders one from seeing what is essential or inessential. Thus it is safest to keep in mind that even in $\mathcal{N}$, we have settled on a specific way to interpret these models in $\mathcal{V}$, and this way is external to the model space itself.

Starting with the correspondence variant of necessitarian semantics, we have that p is true iff $c(p)$ exists. This entails that under a correspondence semantics, all claims have specific ontological commitments: their correspondents. The general ontological commitments are determined by this condition through the relationship that p commits one to X's iff $c(p) \in X$.

Already for truthmaking semantics, we have more interesting structures. One of the guiding motivations of truthmaker theory was to allow the possibility that p has a non-specific connection to reality, such as in the paradigmatic case "there are X's". Thus we often do not have specific ontological commitments in a truthmaker semantics. On the other hand, there is an intimate relationship between a claim's truthmakers and that claim's general commitments. As we defined it in section 5.6, truthmaker semantics is characterised by the principle

$$[\![p]\!] = \{\mathfrak{M} \mid \mathit{TM}\,(p) \cap E_{\mathfrak{M}} \neq \varnothing\}$$

Plugging this into our definitions of specific and general ontological commitment, we find that

$$SCE(p) = \{x \in E \mid \mathit{TM}\,(p) \rightarrowtail x\}$$
$$GCE(p) = \{X \subseteq E \mid \mathit{TM}\,(p) \mathrel{\supset\!\!\!\Rrightarrow} X\}$$

i.e. p commits one to X's generally iff the truthmakers of p necessitate some X distributively.

If the necessitarian semantics used furthermore is sound and complete, this means that if X is the set of truthmakers of a claim q (we could call such a claim "existential", since it is true in exactly the models where there is some X), then we have that p commits one to X's generally iff $p \vdash q$. Of course, there is in general no such claim q, since not every set of models needs to be in the image of some claim under the semantics used, but when one exists, it captures nicely what is involved in ontological commitment for truthmaking.

Now, because $X \mathrel{\supset\!\!\!\Rrightarrow} X$, we have that $\mathit{TM}\,(p) \in GCE(p)$. This means that holding p to be true commits one to truthmakers for p. Conversely, since the existence of any truthmaker for p is sufficient for its truth, the existence of some element in each set in $GCE(p)$ is both sufficient and necessary for the truth of p.

The case is very similar for plural truthmaking semantics. Here, we get the result that

$$SCE(p) = \{x \in E \mid Z \rightarrowtail x \text{ for all } Z \in \mathit{TMP}\,(p)\}$$
$$GCE(p) = \{X \subseteq E \mid Z \mathrel{\succ\!\!\!\Rrightarrow} X \text{ for all } Z \in \mathit{TMP}\,(p)\}$$

which also lets us derive the equivalence that p is true iff p's general ontological commitments are fulfilled. In a sense, the question of ontological commitment therefore exhausts the connection between theory and world in a truthmaker semantics. Alternatively, we can say that truthmaker semantics allows us to reduce the question of what the world is like to the question of ontological commitment, i.e. of what exists.

Plural truthmaker semantics (or, equivalently, positive semantics) may fairly be held to be more useful for discussing the ontological commitments of *theories* than singular truthmaker semantics are. Since a theory A in a framework F is determined by its set $\top_A$ of truths, it is natural to see the theory as being made true by the truthmakers for each of the claims in $\top_A$. While A *might* have a single truthmaker as well (for instance, if the metaphysics used is mereological), we may very well be more interested in which things make true which parts of A, rather than the question of what makes true the whole.

Using ontological commitment to derive what the world is like from which theories are true thus involves going from the truth of claims to the existence of entities. But, interestingly, it does not matter whether the truthmaking semantics used is effectivist or not. Recall that the *effective* truthmakers of p are those truthmakers (or verifiers) most "intimately" related to, or actively involved in the truth of p, where these notions seem impossible to define formally. At first, it may seem that requiring all truths to have effective truthmakers would be more demanding than just requiring them to have verifiers. But, as the following theorem shows, this is not so.

Theorem 6.16 : Let $SCE(p)$ and $GCE(p)$ be the specific and general ontological commitments of p under a non-effective truthmaking semantics $\mathcal{S}(A \mapsto \mathcal{M})$, and let $SCE^e(p)$ and $GCE^e(p)$ be the specific and general ontological commitments of p under an effective truthmaking semantics $\mathcal{S}^e(A \mapsto \mathcal{M})$ such that the truthmakers under $\mathcal{S}^e$ of any claim p are a subset of those they are under $\mathcal{S}$. Assume that both semantics are sound and complete. Then

$$SCE(p) = SCE^e(p)$$

$$GCE(p) = GCE^e(p)$$

for any claim $p \in L_A$.

Proof. First, SCE. The specific commitments of a claim are those entities that are in all worlds where the claim is true. But which sets of entities are possible worlds is determined by $\mathcal{M}$, and since both $\mathcal{S}$ and $\mathcal{S}^e$ are sound and complete, they must assign the same truth-value to p in all worlds. Therefore the intersection of all these worlds does not depend on whether $\mathcal{S}$ or $\mathcal{S}^e$ is adopted.

For the general commitments, we have that p commits one to X's under $\mathcal{S}$ iff $V(p) \Rrightarrow X$, and under $\mathcal{S}^e$ iff $TM(p) \Rrightarrow X$. But since $TM(p) \subseteq V(p)$, $TM(p) \Rrightarrow V(p)$, and conversely, if $e \in V(p)$, then p is true in any world in which e exists, and since p requires an *effective* truthmaker under $\mathcal{S}^e$, e must necessitate such a truthmaker, so $V(p) \Rrightarrow TM(p)$ as well. Since $\Rrightarrow$ is transitive, we find that $V(p) \Rrightarrow X$ iff $TM(p) \Rrightarrow X$, so $GCE(p) = GCE^e(p)$. □

What drives this theorem are really the soundness and completeness assumptions: using these, the structure of the theory we used determines the structure of the metaphysics, up to category-theoretic equivalence. There is simply no room for effectiveness to make any difference in what we are committed to in making claims.

But soundness and completeness of the semantics are very important properties when we are to use theories for the purpose of finding out what the world is like. Thus, we find that the question of whether truthmakers have to be effective or not is *irrelevant* to questions of ontology. Since ontology furthermore determines metaphysics in a truthmaker semantics, we find that there is no *metaphysical* reason to prefer either the effective or the non-effective version of truthmaker semantics to the other. But as there are definite practical reasons not to impose an effectivist requirement, this can be seen as an argument for the non-effectivist version.

If effectiveness does not influence metaphysics, where does it play a part? One possibility is that it matters for the philosophy of language. For instance, if we take the meaning of a claim to be its set of truthmakers, we can have a theory in which (logically) equivalent claims may be non-synonymous. This is, however, strictly a difference in expression:

when it comes to what a claim says *about the world*, effectiveness does not matter.

When it comes to general necessitarian semantics, the principle that the ontological commitments determine claims up to equivalence no longer holds. Since claims can be true because of the *lack* of entities, we have that all claims that deny the existence of something have the same empty commitments, for example. Therefore, knowing that certain things exist is not sufficient for deriving the truth of some claims. In contrast to truthmaker theories, general necessitarian semantics does not allow one to reduce metaphysical questions to questions of ontological commitment.

What does effective truthmaking mean in a general necessitarian semantics, then? Supposedly, the circumstance $(X|Y)$ makes p true effectively if the existence of the X's and the non-existence of the Y's are effective in bringing p's truth about. But the important properties for us is that p is true in $\mathfrak{M}$ iff some truth-making circumstance of it holds, and that the effective truthmaking circumstances are a subset of the truthmaking circumstances. Using these, we can again derive that $SCE(p) = SCE^e(p)$ and $GCE(p) = GCE^e(p)$, so even when we are using necessitarian semantics in its most general form, effectiveness remains unimportant to ontology.

Chapter 7
Applications

So far, we have been exploring the connection between the structure of theory and metaphysics imposed by adopting a semantics. In particular, we have been interested in the class of necessitarian semantics, since these allow us to derive the structural relationships formally. But we have also been moving on a very abstract level: nothing has been presupposed about the theories in question, except that they are representable as consequence operators of some kind. It is time to look at some more concrete cases.

In this chapter, we will move gradually from the general to the specific, beginning with propositional logics before giving two versions of first-order logic. The next step is set theory, and after this we show how to approach a physical theory such as quantum mechanics. Finally, we discuss the application of truthmaker theory to two purely philosophical problems: that of qualia, and that of the metaphysical status of moral facts.

7.1 Sentential Logics

The contemporary recognition of how much one's choice of logic influences and is influenced by one's metaphysics is mostly due to Dummett (1978, 1991b). Starting out from the debate between intuitionists and

Platonists in the philosophy of mathematics, Dummett shows us how the intimate connection is between the intuitionists' characterisation of mathematical objects as ones of our own construction, and their rejection of the law of excluded middle.

It is important to be clear that the conflict is *not* over the nature of truth. Both intuitionists and Platonists presuppose that claims are true iff they describe mathematical reality as it is, and we have noted that this simply follows from the meaning of the word "true". However, they disagree on what this reality is like: for the Platonist, it is independent of human activity and thought, and for the intuitionist it is not. The traditional questions of realism vs. idealism or phenomenalism when it comes to external objects can be seen in a similar light. Both the realist and the idealist are entitled to the same notion of truth, and may hold that $P(a)$ is true iff a has the property P. But a, for the idealist, is a collection or construction of sense-impressions, while it commonly is something quite different for the realist.

Interestingly, anti-realists therefore generally do not have to deny that everyday things exist. They may hold, with Moore, that their hands exist. The difference instead turns up in how such a claim is to be interpreted. In our terms, they apply different semantics. More differences may also come into view when theories are placed inside larger theories, such as ones that include metaphysical notions. For an antirealist, the inference from "I have hands" to "there exist at least two things independently of my mind" is questionable.

To illustrate, let us consider two theories C and I over a common language L, freely generated from a set of atomic sentences using the connectives $\bot$, $\to$, $\wedge$ and $\vee$. Let $\neg p$ be an abbreviation for $p \to \bot$, and let

$$C_C(X) = \{p \in L \mid p \text{ follows from } X \text{ in classical logic}\}$$

and

$$C_I(X) = \{p \in L \mid p \text{ follows from } X \text{ in intuitionistic logic}\}$$

We have that C is a theory in I, obtainable by adding all sentences of form $p \vee \neg p$ to $\top_I$. Now suppose that we adopt a truthmaker semantics for the claims in L. This means that we get two model spaces $\mathcal{M}_C$ and $\mathcal{M}_I$, and by the fundamental theorem of truthmaking, we have that $p \vdash_C q$ iff $TM\,(p) \mathrel{\rightarrowtail\!\!\!\twoheadleftarrow}_C TM\,(q)$ and $p \vdash_I q$ iff $TM\,(p) \mathrel{\rightarrowtail\!\!\!\twoheadleftarrow}_I TM\,(q)$.

It might be thought that the most salient difference between $\mathcal{M}_C$ and $\mathcal{M}_I$ would have to do with truthmakers for sentences of the form $p \vee \neg p$. But in $\mathcal{M}_C$ we do not need truthmakers specifically for $p \vee \neg p$ at all, but only ones for p and for $\neg p$. On the other hand, we have that $\varnothing \mathrel{\succ\!\!\!\twoheadleftarrow}_C TM\,(p) \cup\ TM\,(\neg p)$, so every model contains some truthmaker for p, or some truthmaker for $\neg p$.

Of course, $p \vee \neg p$ *has* truthmakers as well, since every possible entity guarantees its truth, but these are not unique to $p \vee \neg p$. This is important to recognize in order to explain a purported oddity about truthmaker theories brought up by Restall (1996, 2003). Restall asks us to consider two principles of truthmaking, here expressed in our terminology:

Entailment Principle: If $p \vdash q$ and $x \Vdash p$ then $x \Vdash q$.

Disjunction Principle: If $x \Vdash p \vee q$ then $x \Vdash p$ or $x \Vdash q$.

It can be shown that these together lead to a trivialisation of truthmaker theory. Since everything classically entails $p \vee \neg p$, everything needs to be a truthmaker for $p \vee \neg p$. But since only one of p or $\neg p$ can be true, the disjunction principle then implies that everything either is a truthmaker for p, or for $\neg p$, and thus every truthmaker makes every truth true.

As we have characterised non-effective truthmaking, the entailment principle is a theorem. The reason why we do not get triviality is that the disjunction principle is false, and furthermore quite unreasonable for classical logic. There is nothing special about "p" in $p \vee \neg p$, since all sentences of that form are equivalent. If we require logically equivalent sentences to have the same truthmakers, $p \vee \neg p$ should therefore have the same truthmakers as $q \vee \neg q$, and why should we take the truth of that sentence to entail the truth of p or of $\neg p$? The attraction felt for the disjunction principle comes from being misled by the surface form of a sentence.

Things are quite different when we leave the safe confines of classical logic, however. Restall suggests modelling the truthmaking relation on the notion of relevant entailment, rather than the classical kind, and proposes the following principle.

> *Relevant Entailment Principle*: $TM(p) \subseteq TM(q)$ iff p relevantly entails q.

But this really makes sense only when the theory we are applying the semantics to is based on relevance logic as well. Assume that the classical-logical equivalence of p and q entails that $TM\ (p)\ = TM\ (q)$. We have that $p \vdash_C q \vee \neg q$, but we would like to avoid the conclusion that $TM\ (p)\ \subseteq TM\ (q \vee \neg q)$. However, since $q \vee \neg q \dashv\vdash_C p \vee \neg p$, the Hertzianness of our semantics together with the assumption that truthmaker sets are closed under classical equivalence forces us to accept that conclusion, if we hold p to relevantly entail $p \vee \neg p$, as is usually done.

On the other hand, if we accept that classically equivalent claims can have different sets of truthmakers, we may well ask ourselves why. Logically equivalent claims are merely grammatical variations of one another, and essentially "say the same thing". Should this purely syntactical feature correspond to anything as metaphysical as differences in truthmaking?

For another perspective, consider the theory I above, which is intuitionistic logic. *This* logic is compatible with the disjunction principle, even if it does not entail it on its own. Intuitionistic logic does not, by itself, require all disjunctions to entail one of their disjuncts. This holds only for a special class of formulae called *Harrop formulae* (Harrop, 1959). However, the disjunction principle *is* fairly natural from an intuitionistic perspective, and this may be one reason why it is seen as natural for truthmaking as well: truthmaking is, at bottom, an intuistionistic principle.

This can even be made into an argument for adopting intuitionistic logic rather than classical. Let us call two claims p and q *metaphysically independent* iff there is some $x \in V\ (p)$ such that $x \notin V\ (q)$ and $x \notin V\ (\neg q)$, and some $y \in V\ (q)$ such that $y \notin V\ (p)$ and $y \notin V\ (\neg p)$. This means that metaphysically independent claims have at least some truthmakers whose existence do not settle whether the other claim is true. Take the following premisses:

(*i*) Truthmaker semantics.

(*ii*) Logically independent sentences are metaphysically independent.

(*iii*) The only necessitation is deterministic necessitation.

(*iv*) Logic is *at least* intuitionistic.

From (*iv*), it follows that it is sufficient to prove that $\vdash p \vee \neg p$ entails that $\vdash p$ or $\vdash \neg p$ in order to show that $\vdash$ is the consequence relation of intuitionistic logic. Assuming, for contradiction, that $\nvdash p$ and $\nvdash \neg p$, there must be some sentence q which is logically independent of p and $\neg p$. By (*i*) and (*ii*), there is then a verifier x for q such that $x \notin V(p)$ and $x \notin V(\neg p)$. But since $p \vee \neg p$ is true in every world, it follows that $x \mathrel{\succ\!\!\!\in} V(p \vee \neg p)$, and since intuitionistic and stronger logics are distributive, $V(p \vee \neg p) = V(p) \cup V(q)$, so $x \mathrel{\succ\!\!\!\in} V(p) \cup V(\neg p)$. By (*iii*), we must then have that $x \rightarrowtail y$, for some $y \in V(p) \cup V(\neg p)$. But any such y must be a verifier for either p or $\neg p$ (or both), which contradicts (*ii*). Therefore, such a sentence q cannot exist, and since the only sentences that have no logically independent sentences are the theorems and their negations, we must have that either $\vdash p$ or $\vdash \neg p$. It follows that the logic in question cannot be stronger than intuitionistic.

Of course, all the assumptions used could be challenged. Beginning last, (*iv*) does not hold if one thinks that the true logic is a relevant one, although it is possible that one could patch up the proof anyway to show something similar. It definitely holds for those who espouse classical logic, though. (*iii*) has some plausibility: in a way, indeterministic necessitation seems even more mysterious than the deterministic kind. It may also be appropriate for a deterministic theory, such as classical mechanics.

To motivate (*ii*), suppose that there *were* logically independent sentences p and q which are metaphysically dependent, i.e. such that every verifier of p is a verifier either of q or $\neg q$, or vice versa. But, this means that knowing *what* makes p true always allows one to infer whether p or q is true. While there are some truthmakers for p for which this is reasonable (such as whole worlds), it is also very reasonable to hold that not all of p's truthmakers are like this.

Finally, the assumption of truthmaker semantics can be challenged. While the argument goes through for plural truthmaker semantics as well, in both its effective and non-effective versions, it does not hold for general necessitarian semantics (or equivalently, when we allow truthmaking by circumstances). In that case, a lack of entities can make either p or not-p true. It follows that $p \vee \neg p$ can be true in every world even if not every world contains something making p true, or something making $\neg p$ true.

There is also another advantage with using general necessitarian semantics rather than truthmaker semantics. In classical logic, truthmaker semantics is incapable of distinguishing between worlds of different cardinalities. Since every world must contain a truthmaker either for p or for $\neg p$, for every assignment of $\{\mathit{true}, \mathit{false}\}$ to the language, and every world contains at least one truthmaker and one falsemaker, every formally complete theory in classical logic admits models of every cardinality from 2 upwards. This is like the Löwenheim-Skolem theorem, extended to finite cardinalities as well as infinite ones.

But while the implications of this fact are counterintuitive, their importance should not be overestimated. Almost no theory is formally complete, and the number of entities a *specific* theory commits one to can vary from theory to theory. The question of how the world is, given a formally complete theory, is not one that we have to face in practice.

There are downsides to general necessitarian semantics as well, as compared to singular or plural truthmaking. For one thing, we can no longer see the truthmaking relation as a form of "grounding" that ties sentences to entities, since the lack of entities is enough to make sentences true, and lacks are not themselves entities in necessitarian metaphysics. More significantly, there is a serious asymmetry inherent in such a scheme. If p's truth requires a truthmaker, but $\neg p$'s does not, one could well ask oneself why. Which sentences we see as "negated" is a matter of convention, so why should that reflect some deep underlying difference in nature? If we see truthmakers as involved in the explanation of why p is the case, we would like to know why only one of p or $\neg p$ calls for such explanation.

To sum up: if we want to hold on to both truthmaker theory and classical logic, we have to accept nondeterministic necessitation. If we are suspicious about that, we should begin thinking about which of the others to give up. This is interesting because truthmaker theory is often associated with realism (cf. Bigelow, 1988; Armstrong, 1997), and nonclassical logic with antirealism. But as Chris Daly points out in an article, truthmaker theory itself is quite silent on whether there is a world independent of our thoughts and theories (Daly, 2005). However, it may still be that, contrary to common belief, it fits more naturally with an antirealistic metaphysics than a realistic one.

7.2 Classical First-order Logic, from Above

The last section discussed propositional logic using truthmaker semantics. It may at first glance seem like the step to first-order predicate logic ought to be fairly easy: just interpret existential and universal quantification as infinite disjunction and conjunction, as Wittgenstein proposed. Since necessitation—unlike derivability—does not have any inherent problems with infinite sets of antecedents, the metaphysical nature of the necessitation relation might be thought to allow such an interpretation. But this would give us *another* logic, in which we would be allowed to infer from the truth of all instances of a generalisation to the truth of the generalisation itself. It would not be FOL.

In fact, quantification would be nothing like conjunction or disjunction even if it actually was the case that only finitely many things existed. From a universal quantification $(\forall x)P(x)$, we can draw the conclusion that nothing that satisfies $\neg P(x)$ exists. This means that there is *no* model in which $(\forall x)P(x)$ is true, but in which some thing c exists such that $\neg P(c)$. On the other hand, from the truth of $P(c_1) \wedge P(c_2) \wedge P(c_3) \wedge \ldots$ we can draw only the conclusion that in any model in which *these* things exist, none of them satisfy $\neg P$. We can say nothing about whether things that are not in the sequence $c_1, c_2, c_3, \ldots$ are P or not, and neither can we say that $c_1, c_2, c_3, \ldots$ are all the things there are.

In this section, we will approach the problem of how to give a truthmaker semantics for first-order logic from a very general perspective. The next section will be devoted to a more concrete approach. Let L^* be the extension of L that contains not only sentences, but also the open formulae used in generating these recursively, and has consequence defined as usual on these so that $\Gamma \vdash_{L^*} \varphi$ iff every sequence of objects that satisfies the formulae in Γ also satisfies φ, in all models. Where $\mathbf{\Gamma}$ is a set of sets of formulae, let

$$\bigwedge \mathbf{\Gamma} \underset{def}{=} C_{L^*}(\bigcup \mathbf{\Gamma})$$

It follows that $\bigwedge \mathbf{\Gamma}$ is the smallest closed set of sentences in L^* that contains all sets in $\mathbf{\Gamma}$. Let a *partition* $\mathbf{\Pi}$ of L^* be a set of subsets of L_{L^*} such that

(i) Each set $\Gamma \in \mathbf{\Pi}$ is a deductive filter, i.e. is closed under logical consequence.

(ii) For each consistent subset $\mathbf{\Pi}' \subseteq \mathbf{\Pi}$, $\bigwedge \mathbf{\Pi}' \in \mathbf{\Pi}$.

(iii) $\bigcup \mathbf{\Pi} = L_{L^*} \backslash \bot$, where $\bot$ is the set of logical falsehoods.

Each element of such a partition can be taken to determine a kind of *property* or *condition* uniquely. The first condition assures us that if something satisfies a condition, its possession is sufficient to make true everything that follows logically therefrom. The second, which is necessary since the truthmaking theory we use is singular rather than plural, means that conditions are closed under (possibly infinite) conjunctions. The final condition guarantees that the set of conditions is large enough to express everything we need.

We call each element of a partition a *cell.* One example of a partition on a classical *propositional* logic is the set that contains all principal filters of the form $C(\{p\})$ and $C(\{\neg p\})$, where p is any atomic sentence. Another is the partition whose cells are the closures of the sets containing each atomic sentence or its negation, which corresponds to Carnap's so-called state-descriptions. In this case, the filters in question are all *ultrafilters*, i.e. those proper filters that are maximal in the language.

Partitions of predicate logics take somewhat more effort to describe, and we will give examples below.

Given any partition $\mathbf{\Pi}$ of L^*, we let a *direct truthmaking function for* Π be a one-to-one correspondence $dtm : \mathbf{\Pi} \to E$, where $\langle E, \succ\!\!\rightrightarrows \rangle$ is a necessitarian metaphysic, such that

$$dtm[\mathbf{\Pi}'] \rightarrowtail dtm(\Gamma) \text{ iff } \Gamma \subseteq \bigwedge \mathbf{\Pi}'$$

for every logically consistent $\mathbf{\Pi}' \subseteq \mathbf{\Pi}$ and every $\Gamma \in \mathbf{\Pi}$. This condition ensures that direct truthmakers are related correctly to be able to capture the logical relationships between the sentences they make true. It is clear that a necessitarian metaphysics that fulfils it exists, since we can take the elements of $\mathbf{\Pi}$ themselves to be the entities, and the deterministic part of the necessitation relations to be governed by the above condition. The nondeterministic necessitation relation itself is then determined as the closure of this relation, using theorem 4.4.

The partition determines which sentences have unique (direct) truthmakers, and which do not. These direct truthmakers have a mereological structure.

Theorem 7.1 : The direct truthmakers for a partition form a mereology, i.e. a metaphysic in which every non-empty set X of compatible entities have a sum $\widehat{X}$. Furthermore, $dtm(\bigwedge \mathbf{\Pi}') = \widehat{dtm[\mathbf{\Pi}']}$ for all compatible sets of cells $\mathbf{\Pi}' \subseteq \mathbf{\Pi}$.

Proof. We need to show that every non-empty compatible set of direct truthmakers has a sum, i.e. an entity that exists in exactly those worlds where the truthmakers in question exist. Let X be an arbitrary such set. Then $dtm^{-1}[X]$ is a set of logically compatible members of $\mathbf{\Pi}$ (or otherwise, we would have had that $X \succ\!\!\rightrightarrows \varnothing$, and X wouldn't have been compatible). Let $\Sigma = \bigwedge dtm^{-1}[X]$, and let $\widehat{X} = dtm(\Sigma)$. Since Σ contains $dtm^{-1}(x)$, for all $x \in X$, $\widehat{X} \circ\!\!\!\longrightarrow x$ for every $x \in X$.

Assume now that some world contains all entities in X. It is enough to show that if $dtm^{-1}[X]$ is a subset of $\Gamma \in \mathbf{\Pi}$, then $\Sigma \subseteq \Gamma$. But this is follows from the existence of $\bigwedge dtm^{-1}[X]$ for sets of compatible cells. □

Using direct truthmaking, we can define the *truthmaking function* based on $\mathbf{\Pi}$ as the function $\mathit{TM} : L \to \wp(E)$ such that

$$\mathit{TM}\,(\varphi) = \bigcup \{dtm(\Pi) \mid \Pi \in \mathbf{\Pi} \text{ and } \varphi \in \Pi\}$$

This means that the truthmakers of a sentence φ are the direct truthmakers of the cells that φ are in, or as we also may say, the *ways* in which φ can be made true. Let $NC(X)$ be the set of entities $\{y\}$ such that $\{y\} \succ\!\!\!\in X$, i.e. the maximal set of entities that necessitate X distributively. The definition we have given satisfies the condition that $\mathit{TM}\,(\varphi) = NC(\mathit{TM}\,(\varphi))$, so the truthmakers are *all* such ways. A different way of expressing the same thing is to say simply that on this interpretation, truthmaking is non-effective.

Since the partition determines the necessitation relation, it also determines which sets of entities make up possible worlds. Here, however, we encounter a surprise: it turns out that *all* sets of direct truthmakers that are closed under deterministic necessitation are possible worlds. Since W is a possible world iff $W \not\succ\!\!\!\in W^C$, and $\succ\!\!\!\in$ is determined purely from its deterministic part, we have that W is a world iff $W \rightarrowtail e$ implies $e \in W$. This, in turn, means that in general neither φ or $\neg\varphi$ need to be true in a possible world, even if $\varphi \vee \neg\varphi$ is.

The proper way to handle this is to use a bivalent semantics, which means that the version of first-order logic we use must be bivalent as well. The easiest way to define such a version is to use the regular one-valued consequence relation for truth (i.e. assume that $t : \varphi \vdash t : \psi$ iff $\varphi \vdash_{L*} \psi$), and add the inferences

$$t : \varphi \dashv\vdash f : \neg\varphi$$

$$f : \varphi \dashv\vdash t : \neg\varphi$$

It then follows that worlds are maximal sets of entities, as expected.

Theorem 7.2 : For bivalent predicate logic, a set of entities W is a possible world iff it is consistent and maximal, i.e. iff $W \not\succ\!\!\!\in \varnothing$ and $W' \succ\!\!\!\in \varnothing$ for all $W' \supset W$.

Proof. Since the metaphysics is necessitarian, possible worlds are models. For any possible world ω, $\mathcal{S}$ assigns all sentences values in $\{t, f\}$. Because ω is a world, it includes its sum $\hat{\omega}$, and this sum is the direct truthmaker for some cell Π. Suppose that ω is *not* maximal, i.e. that there is some other world ω' that includes it. Then the sum $\widehat{\omega'}$ is the direct truthmaker for some cell Π', and we must have that $\Pi \subset \Pi'$. This means that there must be some assignment $v : \varphi$ that holds in Π' but not in Π, but because either $t : \varphi$ or $f : \varphi$ have to be in Π (or ω would not have been a model), Π' must include both the assignments $t : \varphi$ and $f : \varphi$, for some sentence φ. Using the inferences we introduced, we derive that $\{t : \varphi, t : \neg\varphi\} \subseteq \Pi'$, so such a cell cannot exist, since we have assumed all cells to be consistent. This in turn entails that the direct truthmaker $\widehat{\omega'}$ cannot exist either, and because the metaphysics is mereological, then neither can ω'. □

Now, as we have defined these concepts, every partition gives rise to a sound and complete truthmaker semantics. This is proved in the following theorem.

Theorem 7.3 : Let $\mathcal{S}(L \mapsto \mathcal{M})$ be a truthmaking semantics based on the partition $\mathbf{\Pi}$. Then $\mathcal{S}$ is sound and complete.

Proof. By the fundamental theorem of truthmaking, a truthmaker semantics is sound and complete iff it is Hertzian, so what we need to prove is that *TM*, as defined, fulfils the condition

$$\Gamma \vdash \varphi \text{ iff} \otimes \mathit{TM}\,[\Gamma] \mathrel{\succ\!\!\!\Rrightarrow} \mathit{TM}\,(\varphi)$$

for all $\Gamma \subseteq L$ and $\varphi \in L$.[1] Assume first that Γ is consistent, and that $\Gamma \vdash_L \varphi$. Then $\mathit{TM}\,(\psi)$, for $\psi \in \Gamma$, is the set of direct truthmakers for cells that contain ψ. Since $\hat{X}$, where X is a set of direct truthmakers, is the direct truthmaker for $\bigwedge dtm^{-1}[X]$, we have that $\otimes$ TM $[\Gamma]$ is the set of direct truthmakers for all cells that contain the whole of Γ. But any such cell must also contain φ, because $\Gamma \vdash_L \varphi$ is equivalent to $C_L(\{\varphi\}) \subseteq C_L(\Gamma)$. Therefore, any world that contains truthmakers for

[1] The cross-sum is well-defined since the metaphysics is mereological.

all sentences in Γ must contain a truthmaker for φ as well, so $\otimes\ \mathit{TM}\,[\Gamma] \Rrightarrow \mathit{TM}\,(\varphi)$.

Conversely, let $\otimes \mathit{TM}\,[\Gamma] \Rrightarrow \mathit{TM}\,(\varphi)$. Then any world in which a truthmaker for the whole of Γ exists contains a truthmaker for φ. But truthmakers for Γ are direct truthmakers for cells that contain all sentences in Γ, and truthmakers for φ are the direct truthmakers of those cells that contain φ. We want to show that $\varphi \in C_L(\Gamma)$. But any consistent closed set of formulae in a classical logic can be written as the intersection of the maximal consistent closed sets that contain it, and all maximal closed sets are cells in $\mathbf{\Pi}$. Therefore it is sufficient to show that any maximal cell that contains Γ must contain φ as well. But these cells are, as we have seen, exactly the possible worlds, and if there was a possible world that contained truthmakers for Γ but not for φ, then we would not have had $\otimes \mathit{TM}\,[\Gamma] \Rrightarrow \mathit{TM}\,(\varphi)$ by the representation theorem for necessitation relations.

Finally, the case where Γ is inconsistent. Then $C_L(\Gamma) = L$, and since there is no cell that contains Γ, it has no truthmaker. Therefore $\otimes\ \mathit{TM}\ [\Gamma] = \varnothing$, and $\otimes \mathit{TM}\,[\Gamma] \Rrightarrow \mathit{TM}\,(\varphi)$ holds trivially, since for no world ω, $\omega \cap \varnothing \neq \varnothing$. □

How are we to interpret a semantics such as this? The sets in $\mathbf{\Pi}$ represent the sets of sentences that have *primitive* truthmakers, i.e. truthmaking relations that hold because of direct world-language ties, rather than because of logical relations in the language. These direct ties are all one-to-one, and the selection of a partition can be seen as the imposition of a type of logical atomism, where the cells are the "atoms".

Which such partitions are *correct*, then? This problem is analogous to that of finding the truly "effective" truthmakers. Take first the example where we let each element of $\mathbf{\Pi}$ be an ultrafilter on L^*. In such a case, the elements of E are interpretable as possible worlds, and we get a discrete world semantics, as we called it in section 5.4. Alternatively, we can take $\mathbf{\Pi}$ to consist of *all* consistent filters in L^*. This gives us a correspondence semantics, since every logically closed set of sentences then corresponds to a unique entity.

Discrete world semantics and correspondence semantics represent two limiting cases in the continuum of truthmaker semantics for first-order logic. In between these, we have all the cases where truthmakers are *not* maximally specific, but also not the weakest possible. Since we are considering L^* rather than the more limited language L, a semantics where literals and their universal generalisations are taken to be made true directly can be constructed quite simply as follows:

(i) For each n-place predicate P and each n-tuple of terms $\tau_1, \ldots,$ τ_n, let the logical closures of $P(\tau_1, \ldots, \tau_n)$, $\neg P(\tau_1, \ldots, \tau_n)$, $\tau_1 = \tau_2$ and $\tau_1 \neq \tau_2$ be in $\mathbf{\Pi}$.

(ii) For each cell $\Pi \in \mathbf{\Pi}$, let the cell $\forall\xi(\Pi)$ containing all formulae of form $(\forall\xi)\varphi$ such that $\varphi \in \Pi$ be in $\mathbf{\Pi}$.

(iii) For each consistent set $\mathbf{\Pi}'$ of cells in $\mathbf{\Pi}$, let $\bigwedge \mathbf{\Pi}'$ be in $\mathbf{\Pi}$.

Theorem 7.4 : These rules make $\mathbf{\Pi}$ partition of L^*.

Proof. The only criterion that is not evident is $\bigcup \mathbf{\Pi} = L^* \backslash \bot$. The language can be defined recursively from the atomic formulae by the application of $\neg$, $\wedge$, $\vee$, $(\forall\xi)$, and $(\exists\xi)$, so we prove this by induction on the complexity of formulae. All atomic formulae and their negations are in $\bigcup \mathbf{\Pi}$ by definition. Where φ, $\psi \in \bigcup \mathbf{\Pi}$, we have

- $\varphi \wedge \psi \in \bigcup \mathbf{\Pi}$ because $C(\Gamma \cup \Delta)$, for any $\Gamma, \Delta \in \mathbf{\Pi}$, is in $\mathbf{\Pi}$. Therefore any cell that contains both φ and ψ also contains $\varphi \wedge \psi$, so long as φ and ψ are consistent.
- $\varphi \vee \psi \in \bigcup \mathbf{\Pi}$ because it follows from φ (and ψ).
- $(\forall\xi)\varphi \in \bigcup \mathbf{\Pi}$ because of rule (ii).
- $(\exists\xi)\varphi \in \bigcup \mathbf{\Pi}$ because it follows from φ.
- $\neg\varphi \in \bigcup \mathbf{\Pi}$, because all literals and combinations of them with quantification, conjunction and disjunction are in $\bigcup \mathbf{\Pi}$, and all formulae are equivalent to ones in negation normal form, i.e. where negations occur only in front of atomic formulae.

□

Because $\mathbf{\Pi}$ is a partition of L^*, it gives rise to a sound and complete truthmaker semantics for first-order logic. We may call the members of $dtm[\Delta]$, where Δ is the set of cells generated using the rules (i) and (ii), the *atoms*. These are the direct truthmakers of literals and their universal generalisations. The metaphysics as such contains these atoms, sums of them, and nothing else. Specifically, we need to include the universal generalisations, since first-order logic does not allow one to derive any general formula from particular instances, except when these instances happen to be theorems.

The truthmakers for non-general facts are all direct truthmakers of *open* formulae. But how can an open formula have a truthmaker, since it cannot be *true*? The reason is that we have not defined truth for anything but claims, so a claim is true iff it has a truthmaker, but we are free to say something else about non-claims such as open formulae. This property is shared with Tarski's definition of truth as *satisfaction by all sequences*, since this will assign truth to some non-sentences as well. The only reasonable thing, in our case, seems to be to say that if $P(x_1)$ has a truthmaker, then the thing x_1 refers to satisfies P, so that these entities work more like *satisfaction*-makers than truthmakers. What x_1 refers to will then have to be taken as ambiguous, and possibly to be settled by the context.

7.3 Classical First-order Logic, from Below

There is no question that there *are* truthmaker semantics for first-order logic: the last section gave a method to make such a semantics for any chosen partition of the language. However, the methodology required us to assign truthmakers to open formulae, and it furthermore did not give us a *recursive* way to define truth, unlike Tarski's definition. Naturally, it is the quantifiers that cause the problems. In the last section, we handled them as primitive, and universal quantification as strictly stronger than any combination of non-quantified sentences. But there is another way to approach the problem as well, which lends itself to a slightly different characterisation.

Instead of treating universal quantification as primitive, we can exploit the compactness of first-order logic. However, this requires us to limit the space of possible objects. For instance, we could assume that only the natural numbers, and nothing else, are possible. In such a case, a so-called ω-rule, according to which we may draw the conclusion that $(\forall x)P(x)$ from the set of premisses $P(0), P(1), P(2), \ldots$ may be reasonable. But such a rule only makes sense if we have assumed that nothing but natural numbers can exist.

We will handle FOL in a similar fashion here, but instead of limiting ourselves to natural numbers or any other specific set of things, we will start with an arbitrary set. Let I_b be any infinite set, which we will call the *set of basic individual concepts*. Define a set I such that it satisfies

- For any $c \in I_b$, $c \in I$.
- For any n-ary function symbol f^n in L, and any n-tuple $c_1, \ldots, c_n$ of elements of I, there is a unique object

  $$appl(f^n, c_1, \ldots, c_n)$$

  in I called *the result of applying* f^n *to* $c_1, \ldots, c_n$.

We call I the *set of individual concepts* (cf. Carnap, 1956, pp. 41–42) and the elements of I that are not in I_b the *functional concepts*. These are the concepts created from the concepts in I_b by combining them with function symbols. Let $\mathcal{M}$ be a necessitarian metaphysic $\langle E, \mathrel{\succ\!\!\Subset} \rangle$. Let E, for each n-place predicate P^n of L and each n-tuple $c_1, \ldots, c_n$ of elements of I, contain entities $at(P^n, c_1, \ldots, c_n)$ and $\overline{at}(P^n, c_1, \ldots, c_n)$, such that

$$\{at(P^n, c_1, \ldots, c_n)), \overline{at}(P^n, c_1, \ldots, c_n)\} \mathrel{\succ\!\!\Subset} \varnothing$$

$$\varnothing \mathrel{\succ\!\!\Subset} \{at(P^n, c_1, \ldots, c_n), \overline{at}(P^n, c_1, \ldots, c_n)\}$$

The entities $at(P^n, c_1, \ldots, c_n)$ and $\overline{at}(P^n, c_1, \ldots, c_n)$ are called the *positive* and *negative atomic facts* about whether $P^n(c_1, \ldots, c_n)$ holds.

We also need variants of these for equality, so for each pair c, c' of elements of I, we assume E to contain two atomic facts $eq(c, c')$ and $\overline{eq}(c, c')$ that fulfil the condition that

$$\varnothing \rightarrowtail eq(c, c)$$

$$\{eq(c, c')\} \rightarrowtail eq(c', c)$$

$$\{eq(c, c'), eq(c', c'')\} \rightarrowtail eq(c, c'')$$

$$\{at(P^n, c_1, \ldots, c, \ldots, c_n), eq(c, c')\} \rightarrowtail at(P^n, c_1, \ldots, c', \ldots, c_n)$$

These conditions on the necessitation relation ensure that these facts can do the work of truthmakers and falsemakers for identity statements in FOL. Finally, unless we reduce functions to predicates, we also need atomic facts for the results of function application. For each n-place function symbol f^n and each n-tuple $c_1, \ldots, c_n$ of elements of I, we therefore assume the existence of atomic facts $at(f^n, c, c_1, \ldots, c_n,)$ and $\overline{at}(f^n, c, c_1, \ldots, c_n,)$ for which

$$\{at(f^n, c, c_1, \ldots, c_n), \overline{at}(f^n, c', c_1, \ldots, c_n)\} \rightarrowtail eq(c, c')$$

$$\{at(f^n, c, c_1, \ldots, c, \ldots, c_n), eq(c, c')\} \rightarrowtail at(P^n, c, c_1, \ldots, c', \ldots, c_n)$$

hold. All these kinds of atomic facts will also be referred to as *atoms*. Since our truthmaking is singular rather than plural, we also need sums of compatible atomic facts to make true complex sentences. We take E to include all such sums, and to make life simpler for us, we furthermore assume these sums to be unique so that no set of atoms has more than one sum.

Let At be the set of atomic facts. Because of our atomic basis, all facts are uniquely determined by the elements of At that enter into them.

Theorem 7.5 : There is a one-to-one function $at : E \to \wp(At)$, which is surjective on the non-empty compatible subsets of At, such that $\widehat{at(f)} = f$.

Proof. What we need to do is to show that if $\widehat{X} = \widehat{Y}$, where X and Y are sets of atoms, then $X = Y$. But any world that contains all of X will contain $\widehat{X}$, and the same world will then also contain all of Y, since $\widehat{X} = \widehat{Y}$. Symmetry shows that we must have that any world that contains all Y contains all X as well. It is then trivial to show that At, defined as the inverse of the sum operator $\widehat{\cdot}$, is surjective on the values it is defined on, and that $\widehat{at(f)} = f$. □

This metaphysics contains not only sums, but also *logical complements.* As in the last chapter, let $X \perp\!\!\!\perp Y$, where X and Y are subsets of E, mean that no world contains something both from X and from Y, but any world contains something from one of them. We can prove the following.

Theorem 7.6 : For every $X \subseteq E$, there is a set $Y \subseteq E$ such that $X \perp\!\!\!\perp Y$.

Proof. Every element of E is a sum of compatible atomic facts. But for every atomic fact a, it is easily seen that its negation (i.e. positive/negative variant) $\overline{a}$ is such that $\{a\} \perp\!\!\!\perp \{\overline{a}\}$. For any set of atoms Z, let $\overline{Z}$ be the set of all negations of the elements in Z, and where $\mathcal{Z} = \{Z_1, \ldots Z_n\}$ is a set of such sets, let $\overline{\mathcal{Z}}$ be the set $\{\overline{Z_1}, \ldots, \overline{Z_n}\}$.[2] Let $Y = \otimes\overline{at[X]}$. We want to show that $\Omega^{\exists}(Y) = \Omega \backslash \Omega^{\exists}(X)$. Let $X = \{f_1, \ldots, f_n\}$. We have that

$$\begin{aligned}
\Omega^{\exists}(X) &= \bigcup_{i=1}^{n} \Omega^{\exists}(\{f_i\}) \\
&= \bigcup_{i=1}^{n} \Omega^{\forall}(\{f_i\}) \\
&= \bigcup_{i=1}^{n} \Omega^{\forall}(at(f_i))
\end{aligned}$$

and that

[2] We do not intend to rule out $n = \infty$ here.

$$
\begin{aligned}
\Omega^{\exists}(Y) &= \Omega^{\exists}(\otimes\overline{at[X]}) \\
&= \bigcap_{i=1}^{n} \Omega^{\exists}(\overline{at(f_i)}) \\
&= \Omega\backslash\bigcup_{i=1}^{n}(\Omega\backslash\Omega^{\exists}(\overline{at(f_i)}))
\end{aligned}
$$

We now show that for any fact f, $\Omega^{\forall}(at(f)) = \Omega\backslash\Omega^{\exists}(\overline{at(f)})$, from which the theorem follows directly. Let ω be any world in $\Omega^{\forall}(at(f))$. Then $at(f) \subseteq \omega$, so $\overline{at(f)} \cap \omega = \varnothing$, and thus $\omega \notin \Omega^{\exists}(\overline{at(f)})$ Conversely, let $\omega \in \Omega^{\exists}(\overline{at(f)})$ Then there is some $\overline{a} \in \overline{at(f)}$ such that $\overline{a} \in \omega$. But then we must have that $a \notin \omega$, so we cannot have that $at(f) \subseteq \omega$. □

Since sums are unique in our metaphysics, logical complements are unique as well, and we write the logical complement of the set X as $X^{\perp\!\!\!\perp}$. This concept allows us to express the important necessitation relation $\succ\!\!\!\Lleftarrow$ in two somewhat simpler ways, which we will have use for when proving completeness.

Theorem 7.7 : $X \succ\!\!\!\Lleftarrow Y$ is equivalent to $E \succ\!\!\!\Lleftarrow X^{\perp\!\!\!\perp} \cup Y$ and to $X \otimes Y^{\perp\!\!\!\perp} \succ\!\!\!\Lleftarrow \varnothing$.

Proof. Obtainable by simple set-theoretical manipulation from the facts that $X \succ\!\!\!\Lleftarrow Y$ iff $\Omega^{\exists}(X) \subseteq \Omega^{\exists}(Y)$, $\Omega^{\exists}(X \cup Y) = \Omega^{\exists}(X) \cup \Omega^{\exists}(Y)$, $\Omega^{\exists}(X \otimes Y) = \Omega^{\exists}(X) \cap \Omega^{\exists}(Y)$, and $\Omega^{\exists}(X^{\perp\!\!\!\perp}) = \Omega\backslash\Omega^{\exists}(X)$. □

Worlds correspond to maximal consistent sets of facts:

Theorem 7.8 : A subset $W \subset E$ is a world iff $W \not>\!\!\!\Lleftarrow \varnothing$ and, for any $e \notin W$, $W \cup \{e\} >\!\!\!\Lleftarrow \varnothing$.

Proof. From the definition, $W \in \Omega$ iff $W \not>\!\!\!\Lleftarrow W^{C}$. Suppose this holds. Then $W \not>\!\!\!\Lleftarrow \varnothing$, or $W >\!\!\!\Lleftarrow W^{C}$ would have held, by Dilution. Let $e \notin W$. Then we must have that $at(e) \cap W = \varnothing$, for e necessitates its atoms.

Let a be any one of these. Now, a must be one of the types of atomic fact, but if W is a world, it already contains a maximal number of these (this is trivial from the conditions we have laid down on atoms). Thus there can be no such entity e.

In the other direction, assume that $W \not\succ\!\!\!\Rrightarrow \varnothing$ and that $W \cup \{e\} \succ\!\!\!\Rrightarrow \varnothing$ for all $e \notin W$, and for contradiction that $W \succ\!\!\!\Rrightarrow W^C$. Then any world W' that strictly contains W must contain some entity $e \notin W$, but this means that $W \succ\!\!\!\Rrightarrow \varnothing$, so W could not have been a world to start with. Thus there is no world that contains W, and this means that $W \succ\!\!\!\Rrightarrow \varnothing$, contrary to our first assumption. Therefore $W \not\succ\!\!\!\Rrightarrow W^C$. □

We have taken *facts* as primitive, rather than individuals. But given any world, we can still define a set of individuals. For a world ω, let $\sim_\omega$ be a relation on I such that $c_i \sim_\omega c_j$ iff $eq(c_i, c_j) \in \omega$. This relation, expressible as "c_i and c_j are identical in the world ω", has all the properties we should expect from an identity relation.

Theorem 7.9 : $\sim_\omega$ is an equivalence relation, and if $c_i \sim_\omega c_j$, then $at(P^n, c_1, \ldots, c_i, \ldots, c_n) \in \omega$ iff $at(P^n, c_1, \ldots, c_j, \ldots, c_n) \in \omega$, for any predicate P^n.

Proof. A simple verification from the postulates laid down on $eq(c_1, c_2)$. □

Let $Ind(\omega)$ be the set of equivalence classes of I under the relation $\sim_\omega$. This set gives us a kind of representation of which individuals exist in the world ω, in terms of which individual concepts they instantiate. It is easily seen that this allows us to represent domains of any cardinality from 1 up to $|I|$.

We will now describe the truthmakers for arbitrary sentences in L_L recursively. Let a *basic assignment* be a function $v_b : Var \to I_b$, where Var is the set of variables of L. Let an *assignment* be a function $v : Term \to I$, where $Term$ is the set of terms of L, such that

- $v(\tau) \in I_b$ if τ is a variable, and
- $v(f^n(\tau_1, \ldots, \tau_n)) = appl(f^n, v(\tau_1), \ldots, v(\tau_n))$, for any function symbol f^n and any n-tuple of terms $\tau_1, \ldots, \tau_n$.

As before, where v is any assignment, let $v[c/\xi]$ be the assignment exactly like v except at every occurrence of the variable ξ, where it takes the value c. For each such assignment, define a truthmaking function TM_v: $L_{L^*} \to \wp(E)$ recursively, using the following clauses.

$$\begin{array}{rcl}
TM_v\ (P^n(\tau_1, \ldots, \tau_n)) & = & \{at(P^n, v(\tau_1), \ldots, v(\tau_n))\} \\
TM_v\ (\tau = \tau') & = & \{eq(v(\tau), v(\tau'))\} \\
TM_v\ (\neg\varphi) & = & TM_v\ (\varphi)^{\perp\!\!\!\perp} \\
TM_v\ (\varphi \wedge \psi) & = & TM_v\ (\varphi) \otimes TM_v\ (\psi) \\
TM_v\ (\varphi \vee \psi) & = & TM_v\ (\varphi) \cup TM_v\ (\psi) \\
TM_v\ ((\forall\xi)\varphi) & = & \bigotimes_{c\in I}(TM_{v[c/\xi]}\ (\varphi)) \\
TM_v\ ((\exists\xi)\varphi) & = & \bigcup_{c\in I}(TM_{v[c/\xi]}\ (\varphi))
\end{array}$$

Lemma 7.10 : If v and v' are assignments, and φ is a sentence, then

$$TM_v\ (\varphi) = TM_{v'}\ (\varphi)$$

Proof. By induction on the number of variables in φ. □

Because of this lemma, we can define the *non*-assignment-relative truthmakers $TM\ (\varphi)$ of a *sentence* φ to be $TM_v\ (\varphi)$, for any assignment v.

Theorem 7.11 : This semantics is sound and complete for first-order logic.

Proof. Let us call the truthmaker semantics described here $\mathcal{TM}$, and use $\mathcal{T}$ to refer not only to the model space of Tarskian models, but to the Tarskian semantics as well. We show that there are functions $\mathfrak{m} : \Omega \to \mathcal{T}^{\aleph_0}$ and $\omega : \mathcal{T}^{\aleph_0} \to \Omega$, where $\mathcal{T}^{\aleph_0}$ is the category of Tarskian models with countable domains, such that $\omega \vDash_{\mathcal{TM}} \varphi$ iff $\mathfrak{m}(\omega) \vDash_{\mathcal{T}} \varphi$ and $\omega(\mathfrak{M}) \vDash_{\mathcal{TM}} \varphi$ iff $\mathfrak{M} \vDash_{\mathcal{T}} \varphi$, for any sentence φ, any world ω, and any countable Tarskian model $\mathfrak{M}$. This means that if φ is true in some world in Ω, then it is true in some model in $\mathcal{T}^{\aleph_0}$, and vice versa. Soundness and completeness follow from the soundness and completeness of countable Tarskian semantics.

Define $\mathfrak{m}(\omega)$ as the model whose domain D is the set $Ind(\omega)$ of individuals of ω, such that the extension of the predicate P^n is the set of n-tuples $d_1, \dots d_n$ such that $at(P^n, c_1, \dots, c_n) \in \omega$, for some individual concepts $c_1 \in d_1, \dots, c_n \in d_n$. Let the extension of the function symbol f^n be the function f on D such that $d = f(d_1, \dots, d_n)$ iff $at(f^n, c, c_1, \dots, c_n) \in \omega$, where again $c_k \in d_k$ for $k = 1 \dots n$. For any assignment v on I, let v^* be an assignment on D such that $v^*(\tau)$, for any term τ, is the element of D in which $v(\tau)$ is included.

For any countable Tarskian model $\mathfrak{M}$ with domain D, let h be any injective function from D to I. Let $W \subseteq E$ be the set that contains the atom $at(P^n, h(d_1), \dots, h(d_n))$ iff $\langle d_1, \dots, d_n \rangle$ is in the extension of P^n in $\mathfrak{M}$, and likewise for the function symbols. Let $\equiv$ be any equivalence relation on I such that $c_1 \equiv c_2 \Rightarrow c_1 = c_2$ for all $c_1, c_2 \in h[D]$, and $\bigcup[h[D]]_{\equiv} = I$. It is clear that such a relation always exists, and that each class in $I/\equiv$ contains exactly one element of $h[D]$. Extend W to a set W' such that for each $c_1, c_2 \in I$ for which $c_1 \equiv c_2$, $eq(c_1, c_2) \in W'$, and such that W' satisfies the postulates on equality atoms above. Finally, define $\omega(\mathfrak{M})$ to be the unique world that contains W' as its set of positive atoms (there is such a world since W' has to be consistent, and it is unique because of the fact that the positive atoms are fixed by the definition of W). For each assignment s on D, let s^* be any assignment on I such that $s^*(\tau) \in h(s(\tau))$, for each term τ.

We now wish to prove that the functions $\mathfrak{m}$ and ω preserve which sentences are true or false. For this, it is clearly sufficient (and also necessary, by the preceding lemma) that this holds for all assignments. We proceed recursively, by induction on the complexity of formulae. Let ω be any world and $\mathfrak{M}$ any Tarskian model of L, let v be any assignment on I, and let s be any assignment on $\mathfrak{M}$'s domain D. For atomic formulae (i.e. those of complexity 0), we have one of the cases

- φ is of the form $P^n(\tau_1, \dots, \tau_n)$. Assume that $\omega \cap \ TM_v\ (\varphi) \neq \varnothing$. Then ω contains the atom $at(P^n, v(\tau_1), \dots, v(\tau_n))$, so $\langle v^*(\tau_1), \dots, v^*(\tau_n) \rangle$ must be in the extension of P^n in $\mathfrak{m}(\omega)$, and thus $\mathfrak{m}(\omega) \vDash_{v^*} \varphi$. If $\omega \nvDash_v \varphi$, on the other hand, then $\mathfrak{m}(\omega)$ does not have $\langle v^*(\tau_1), \dots, v^*(\tau_n) \rangle$ in the extension of

P^n, so $\mathfrak{m}(\omega) \not\models_{v*} \varphi$.

For the function ω, let $\mathfrak{M} \models_s \varphi$. Then, by construction of $\omega(\mathfrak{M})$, $\omega(\mathfrak{M})$ must contain the atom $at(P^n, s^*(\tau_1), \ldots, s^*(\tau_n))$, so $\omega(\mathfrak{M}) \models_{s*} \varphi$. Conversely, if $\mathfrak{M} \not\models_s \varphi$, then

$$at(P^n, s^*(\tau_1), \ldots, s^*(\tau)) \notin \omega(\mathfrak{M})$$

must hold, so then $P^n(\tau_1, \ldots, \tau_n)$ cannot be true in $\omega(\mathfrak{M})$.

- φ is of the form $\tau_1 = \tau_2$. If $\omega \models_v \varphi$, then $eq(v(\tau_1), v(\tau_2))$ is in ω, so $v(\tau_1)$ and $v(\tau_2)$ are members of the same individual in $Ind(\omega)$, and $v^*(\tau_1) = v^*(\tau_2)$. If $\mathfrak{m}(\omega) \models_{v*} \varphi$, then $v^*(\tau_1) = v^*(\tau_2)$, so $eq(v(\tau_1), v(\tau_2)) \in \omega$.

  If $\mathfrak{M} \models_s \varphi$, then $s(\tau_1) = s(\tau_2)$, so we must have $s^*(\tau_1) = s^*(\tau_2)$ as well. But since $eq(c, c)$ is in all worlds, for any individual concept c, we have that $eq(s^*(\tau_1), s^*(\tau_2))$, so $\omega(\mathfrak{M}) \models_{s*} \varphi$. In the other direction, assuming that $\omega(\mathfrak{M}) \models_{s*} \varphi$, it follows from the injectivity of h that $\mathfrak{M} \models_s \varphi$.

Now assume that we have proved that $\mathfrak{m}(\omega) \models_{v*} \psi$ iff $\omega \models_v \psi$ and $\omega(\mathfrak{M}) \models_{s*} \psi$ iff $\mathfrak{M} \models_s \psi$ for all formulae ψ of complexity n, and wish to prove that the same holds for formulae of complexity $n + 1$. We give the proofs for all rules except those of $\wedge$ and $\exists$, since these are very similar to those of $\vee$ and $\forall$.

- φ is of the form $\neg\psi$. Assuming that $\omega \models_v \varphi$, ω must contain some element e such that $e \in TM_v\ (\psi)^{\perp\!\!\perp}$. But no world can contain both such an element and an element that makes ψ true, so ψ is false. We must therefore have that $\mathfrak{m}(\omega) \not\models_{v*} \psi$, so $\mathfrak{m}(\omega) \models_{v*} \varphi$.

  Let $\mathfrak{M} \models_s \neg\psi$. Then ψ is not true, so it does not have a truthmaker in $\omega(\mathfrak{M})$, and since any world contains truthmakers either for ψ or for $\neg\psi$, we must have that $\omega(\mathfrak{M}) \models_{s*} \neg\psi$. Conversely, if $\mathfrak{M} \not\models_s \neg\psi$, then $\mathfrak{M} \not\models_s \psi$, so $\omega(\mathfrak{M})$ has some truthmaker for ψ. Therefore it cannot have one for $\neg\psi$, so $\omega(\mathfrak{M}) \not\models_{s*} \neg\psi$.

- φ is of the form $\psi_1 \vee \psi_2$. Let $\omega \vDash_v \varphi$. Then ω contains some element in $TM(\psi_1) \cup TM(\psi_2)$, but it can do so only if it contains some truthmaker either for ψ_1 or ψ_2. Therefore ψ_1 or ψ_2 must be true, so $\mathfrak{m}(\omega) \vDash_{v*} \varphi$. If $\omega \nvDash_v \varphi$, then ω contains truthmakers neither for ψ_1 nor ψ_2, so both of these are false, and therefore $\mathfrak{m}(\omega) \nvDash_{v*} \varphi$.

  $\mathfrak{M} \vDash_s \varphi$ is true iff ψ_1 or ψ_2 is. But these are true iff some truthmaker for ψ_1 or ψ_2 exists, so $\mathfrak{M} \vDash_s \psi_1 \vee \psi_2$ iff $\omega(\mathfrak{M}) \vDash_{s*} \psi_1 \vee \psi_2$

- φ is of the form $(\forall\xi)\psi$. From the earlier theorems we have proved about cross-sums, it follows that

  $$\omega \vDash_v (\forall\xi)\psi \text{ iff } \omega \vDash_{v[c/\xi]} \psi \text{ for all } c \in I$$

  But $(\cdot)^*$ is a *surjective* function from assignments on I to assignments on $Ind(\omega)$, so this holds iff $\mathfrak{m}(\omega) \vDash_{s[c/\xi]} \psi$, for all assignments s on the domain of $\mathfrak{m}(\omega)$.

  If $\mathfrak{M} \vDash_s (\forall\xi)\psi$, then, for any assignment s' on D that is like s except possibly at ξ, $\mathfrak{M} \vDash_{s'} \psi$. We need to show that for no assignment $s^*[c/\xi]$, $\omega(\mathfrak{M}) \nvDash_{s^*[c/\xi]} \psi$. But suppose that there were such an assignment. Then, since all individual concepts of an individual d satisfy the same formulae, we must have that for some c in the image of h, $\mathfrak{M} \nvDash_{s^*[c/xi]} \psi$, so $\mathfrak{M} \nvDash_{s[d/\xi]} \psi$ for some $d \in D$, contrary to assumption. Likewise, if there is some $d \in D$ such that $\mathfrak{M} \nvDash_{s[d/\xi]} \psi$, then it follows by the construction of ω that $\mathfrak{M} \nvDash_{s[d/\xi]*} \psi$.

□

What kind of metaphysics is this? The fundamental entities we have are facts, or at least fact-like, since they can do the work of making true. Identity, as we interpret it, does not have the same role as in Tarskian semantics. Instead, it is a relation between individual concepts, and these concepts, in turn, are used only to specify facts. It is only the facts themselves that exist, and individuals and individual concepts are merely linguistic aids for us to talk about them.

Another feature worth commenting on is that we seemingly *do* have the disjunction thesis, since $TM\ (\varphi \vee \psi) = TM\ (\varphi) \cup\ TM\ (\psi)$. But, of course, this is not so. The identity holds only for complex formulae, and not for atomic ones. In this semantics, we have taken the atomic formulae to be those traditionally seen as atomic and their negations, but as we saw in the last section, this is only one of the infinity of choices we could have made.

7.4 Set Theory and Mathematics

Mathematics is a subject whose metaphysics has inspired and intrigued philosophy since Plato. If numbers, sets and functions are not out there in the physical world, then where are they? Does the use of mathematics really presuppose the ontological excesses of Platonism, as Quine argued? To simplify matters here, and to allow the application of the methods of this book, we shall assume not only that mathematics is useful, but also that it is *true.*

At first glance, it might seem that this settles the matter: if "there are primes larger than 100" is true, does that not entail that there are primes? And does that, in turn, not entail that there are numbers, since primes are numbers? Yes and no. The first entailment certainly holds in Peano arithmetic, and the second holds if one has a wider theory that incorporates the concept of number, such as ZFC. But this just concerns which sentences we may infer from which. Why does the sentence "there are numbers" commit one to numbers?

Put this way, this question looks almost silly. The sentence commits one to numbers because that is what it means! But, the only thing we can infer from this is that we are entitled to infer "there are numbers" from it. We are moving in a circle, inside one and the same theory. In Carnap's terms, the *internal* question of whether there are numbers is trivial, and it can lead us to no metaphysical insights.

For the external question, not so. Carnap is certainly right that it does not make sense to ask a *purely* external question, separated from any language or theory. Such a question is just noise, or marks on paper without any significance, and a theory is needed to interpret it. But we may still be interested in questions that are external to mathematics. Seen this way, it is trivial that numbers exist in a mathematical sense, but do they exist in some other sense as well?

The way to get a clear picture of such a "sense" is by semantics, and in this chapter, we focus on truthmaker semantics. The question of whether holding Peano arithmetic to be true commits one to numbers then transforms into the question of what the truthmakers of true statements in Peano arithmetic are.

But here we encounter a surprise. If mathematics is truly necessary, i.e. if mathematical statements hold in every possible world, then they need no truthmakers of their own. Equivalently, every possible entity is a truthmaker for every mathematical statement. Seen from an "explationist" viewpoint, p only needs explaining (or at least only needs an *ontological* explanation) if it could have been false. If we hold that $5 + 7 = 12$ couldn't have been otherwise, there is nothing to explain, and the statement needs no specific truthmakers. Thus, mathematics does not involve us in any ontological commitments at all, above those we already had.

So suppose that mathematics is *not* necessary, at least in some sense. It is, for instance, not usually seen as *logically* necessary, as there are models of predicate logic where $5+7 \neq 12$. Since the Peano axioms (at least as both Dedekind and Peano presented them) presuppose some kind of set or class theory, it will be more useful for us to discuss such a theory. We take ZFC as our example. What kind of ontological commitments does one involve oneself in by holding ZFC to be true?

Starting with the language, ZFC is expressible using only the primitive two-place predicate $\in$ and no function symbols. We do not even need an identity predicate, since identity can be defined in terms of having the same members, but to keep the formalism general, we will include identity as a primitive concept as well. The language thus consists of sentences in a variant of first-order logic without function symbols, and without any other relations. We assume that the regular consequence relations hold, so that $\Gamma \vdash \varphi$ iff φ is a first-order consequence of Γ. ZFC then comprises a theory in this framework, which we shall call $FZFC$.

The problem that interests us here is what commitments we impose by taking ZFC to be true, rather than merely $FZFC$. Since ZFC is a theory in $FZFC$, it is true iff $\top_{ZFC}$ is true, so what we need is to find truthmakers for these sentences: the theorems of ZFC. Since ZFC is formally incomplete, these do not decide all claims in L_{ZFC}, unless ZFC is inconsistent. This entails that any semantics we use must have more than one model.

Take the following axiomatisation of ZFC, here given in English as the translations to FOL are trivial and standard:

Extensionality: Sets with the same members are identical.

Replacement: The image of any set under a functional relation is a set.

Powerset: The subsets of any set together form a set.

Union: The elements of the sets in any set of sets together form a set.

Regularity: Every non-empty set contains some element disjoint from it.

Infinity: There is an infinite set.

Choice: Every set of disjoint sets has a choice function.

There is no need for axioms for separation, pairing, or null set, since these follow from the others due to the presence of Replacement. First of all, we note that it is sufficient for the truth of ZFC that there exists at least one truthmaker for each of the axioms, and for finite conjunctions of them. In the case of replacement, which is an axiom schema, this calls for a countable infinity of truthmakers. This means that we cannot presuppose that there is a single entity ("the whole of set-theoretical reality") that makes all of ZFC true, which would have been the case if ZFC had been logically equivalent to a single first-order axiom. ZFC thus is an example of a theory with no minimal truthmaker: every single thing that makes ZFC true, will make something else true as well.

Truthmaker semantics gives us a strikingly different picture of the ontological commitments of mathematics than the standard Tarskian semantics. An axiom such as Choice, as it is usually read, postulates the existence of one choice function *for every set of disjoint sets.* But truthmaker-theoretically, the sets involved have little influence on the ontology. We may compare their role with the individual concepts of the last section, which are used only to specify facts. *Prima facie*, Choice can be made true by a single entity.

Things are however not quite as simple as this. The individuation of truthmakers is, by the fundamental theorem of the last chapter, intimately tied to the logical relations among claims. Since ZFC is a first-order theory, we have to look at the first-order consequences of the axioms. Taking Choice as our example, every non-equivalent sentence that follows from this axiom has to have its own set of possible truthmakers. While nothing precludes the actual world from being such that it contains a single thing, and that thing makes Choice and all sentences that follow from it true, there needs to be other possible worlds where more things are involved as well.

More specifically, whenever Choice $\dashv\vdash \varphi_1 \vee \ldots \vee \varphi_n$ for some logically independent set $\varphi_1, \ldots, \varphi_n$ of FOL sentences, there are possible entities $a_1, \ldots, a_n$ such that $a_1 \Vdash \varphi_1, \ldots, a_n \Vdash \varphi_n$, and thus such that these together (plurally) make Choice true. But, since Choice requires a *single* truthmaker in classical truthmaking semantics, these must necessitate such a truthmaker for Choice as well.

So from the truth of Choice, we can draw the conclusion that there is some entity c such that c makes Choice true. But there may also be other things, which together necessitate c. How do we find out if this is the case? Taking a specific consequence φ of Choice as an example, the question becomes one of whether there is some world ω that contains one of the things c which make φ true in the actual world, but where Choice is false. Or differently put, does TM (φ) contain the actual world?

We have already noticed the difficulties with using a phrase such as "the actual world". All that our theories and semantics can do is to separate out a world in which the same claims are true or false as in the actual world. We can close in on $\mathfrak{A}$, gradually, by adding more and more theory, although we have no reason to ever expect to be able to identify it uniquely. Since the only access we have to metaphysics is through our theories, we have to resist the temptation to try to answer questions such as “what does *actually* make p true”? Given truthmaker theory, all we can say is that *something* does, and then proceed to investigate the structural properties of this *something*.

The problem can be simplified somewhat through the imposition of a more specific semantics: if not every conceivable truthmaker of p is taken to be possible, it may be that the necessitation structure is enough to single out particular truthmakers. Take, for instance, a classical-logical truthmaker semantics for $FZFC$ based on a partition $\mathbf{\Pi}$. In this semantics, each set in $\mathbf{\Pi}$ requires its own direct truthmaker. We can prove the following:

Theorem 7.12 : The general commitments of ZFC are the sets of entities that contain some set $dtm(\Gamma)$, where Γ is any member of $\mathbf{\Pi}$ which is contained in $C_{FZFC}(ZFC)$.

We are thus committed to something playing the role of truthmaker for every theory contained in ZFC. Unless we know something about $\mathbf{\Pi}$ and the function dtm, we can say nothing about the specific commitments. For the semantics described in the last section, the truthmakers will have to depend on the nature of the set I_b of basic individual concepts. For each assignment v such that $v(x_1) \neq v(x_2)$, the literals “$x_1 \in x_2$” and “$x_1 \notin x_2$” will have unique assignment-relative truthmakers. But they do not, of course, have proper *non*-assignment-relative truthmakers, since only sentences do so.

Take, for instance, the Null Set theorem and the Extensionality axiom.

$$(\exists x_1)(\forall x_2)x_2 \notin x_1$$

$$(\forall x_1)(\forall x_2)((\forall x_3)(x_3 \in x_1 \leftrightarrow x_3 \in x_2) \rightarrow x_1 = x_2)$$

These are usually interpreted as showing that there is a unique null set. But in the truthmaker semantics we have used here, they say no such thing. Let v be an assignment such that $x_1 \mapsto c_1$ and $x_2 \mapsto c_2$. We have that

$$TM_v \ (x_2 \notin x_1) = \{\overline{at}(\in, c_1, c_2)\}$$

so TM_v $((\forall x_2)(x_2 \notin x_1))$ is the set consisting of the single mereological sum of all atoms of the form $\overline{at}(\in, c_1, c_2)$, where $c \in I$. The truthmakers for $(\exists x_1)(\forall x_2)(x_2 \notin x_1)$ are then *all* such sums, for any value of x_1. But these are as numerous as the elements of I, and no model of ZFC has to contain *just* one of them.

The Extensionality axiom, which “should” have given us uniqueness, does not do so either. This is obvious if we remember that it does not follow from the Null Set theorem, so it represents a strengthening of the theory generated by that theorem. But adding more theory can never reduce ontologies in a truthmaker semantics. Instead, the extensionality axiom invokes truthmakers of its own, and does nothing to reduce the ontological indeterminacy of the Null Set theorem.

For truly specific commitments, we need to consider sentences without existential quantifiers or disjunctions. An example of such a sentence is

$$(\forall x_1)(x_1 \notin x_1)$$

$\mathcal{P}$

which follows from the axiom of regularity. This is made true by the sum of atoms of the form $\overline{at}(\in, c, c)$, for all $c \in I$. Since we have assumed sums to be unique in our metaphysics, this entity indeed exists in every model in which ZFC is true. It is thus part of ZFC’s specific commitments.

It is clear that truthmaker theory lets us paint a picture of set-theoretical reality far removed from how it is traditionally conceived. We can find truthmakers for all of ZFC’s truths, but one is left with the question: *what are these objects*? We are here invited to move away from the model-theoretical approach to metaphysics, and into the more expressionistic areas of trying to interpret truthmakers in terms of more well-known objects. We have found a collection of objects identified through descriptions such as “the truthmaker for $a \in b$”, and we want to see if there are more informative or intuitive ways to describe them.

Unfortunately, it appears we cannot take the truthmakers of ZFC to be sets in the traditional sense, so that $x \in b$ is made true by b, for any x. The reason is that this would make the inference of $a' \in b$ from $a \in b$ valid, for any a, a' in b, since they would have the same truthmaker. But this inference is not valid in $FZFC$. On the other hand, if use ZFC itself to settle validity, then all sentences could be taken to have the same truthmaker.

What if we use some theory *between* $FZFC$ and ZFC as a framework, then? Any collection of consequences of the axioms of ZFC generates such a theory. The difficulties encountered lie both in identifying exactly which sentences we should take as truths of the framework, as well as in motivating why exactly this framework, rather than $FZFC$ or any other, is to be used.

Of course, there is nothing that hinders us from settling the problem of couching truthmaker vocabulary in more familiar terms by stipulation. We could, as in the last section, call the truthmakers "facts", or perhaps even "mathematical facts". This is of course just a matter of terminology, but on the other hand, one should never underestimate the power of terminology either.[3] What is important to remember is that just because we have put a name on some class of things, that does not make these things into a well-defined ontological category, separate from everything else. By calling something a "fact", we do not thereby rule out that it may also be an "object", for example.

Another perspective becomes available when we take a step back and look at the place of ZFC in other theories. Presumably, a first-order language for physics may include $\in$, but it will also contain other predicates. Could we have that the truthmakers of sentences of the form $a \in b$ are identical to truthmakers of some other sentence $\varphi(a, b)$? While this would not necessarily give us a *reduction* of ZFC to non-mathematical vocabulary, it *would* show that ZFC does not add to the ontological commitments of physics.

However, as long as our framework is purely first-order logical, this cannot be. If $a \in b$ were to have the same truthmakers as $\varphi(a, b)$, where φ does not mention $\in$, then we would be allowed to infer $a \in b$ from $\varphi(a, b)$, as they are true in exactly the same models. But in FOL, *no*

[3] Cf. Feynman's famous comment "We could, of course, use any notation we want; do not laugh at notations; invent them, they are powerful. In fact, mathematics is, to a large extent, invention of better notations." (Feynman, 1963, p. 17-7)

literal is inferable from a sentence that does not mention the literal's predicate.

Again, the matter is different if we consider frameworks stronger than first-order ones. If the framework licenses the derivations $\varphi(a, b) \dashv\vdash a \in b$, we are free to adopt a semantics in which $\mathit{TM}\,(\varphi(a, b)) = \mathit{TM}\,(a \in b)$. But this just pushes the question back, to one of which framework we should adopt. The problem is that certain inferences *are* accepted in mathematics, and other are not. If we were to, for instance, identify truthmakers of mathematical sentences with certain behavioural patterns among mathematicians, it would be allowed to refer to these patterns in a mathematical proof. But it is not: no type of argument external to mathematics itself or classical logic is allowed in mathematics. The ontological question will, as the intuitionists saw, have to have influence on the logical.

7.5 Quantum Mechanics

The applications we have considered so far have traditionally been seen as a priori subjects. While we have not made anything of the a priori/a posteriori or analytic/synthetic distinctions, it is instructive to also consider a theory that falls within the usual concept of "empirical". One of the most fundamental of these theories is quantum mechanics. This case also has intrinsic interest, since quantum mechanics is often seen as requiring us to adopt new ways of thinking about metaphysics.

As in section 2.4.3, let QM be a theory whose language L_{QM} is a collection of all sentences of the three forms

Preparation:	the system is prepared in state ϱ at t.
Measurement:	observable $\mathbf{A}$ is measured at t.
Observation:	the value of observable $\mathbf{A}$ at t is in the set V.

where ϱ is a density operator, $\mathbf{A}$ is an observable, t is a time, and V is a Borel set of real numbers. As before, we use $p, p_1, p_2, \ldots$ for preparation sentences, $m, m_1, m_2, \ldots$ for measurement sentences, and $o, o_1, o_2, \ldots$ for observation sentences. Recall that for any sentence s, $t(s)$ be the time mentioned in such it, and for any sentence of measurement or observation, $\mathbf{O}(s)$ be the observable involved in it. Call the set of all preparation sentences P, that of all measurement sentences M, and that of all observation sentences O.

Define the probabilistic consequence operator C^{π}_{QM} recursively, as indicated in ch. 2. This means that $C^{\pi}_{QM}(X)$ is obtained by time-ordering the sentences of X, and then letting $C^{\pi}_{QM}[0](X)$ be the observation sentences that have probability π given the preparation and measurement sentences first in X, and $p \in C^{\pi}_{QM}[k](X)$ the observation sentences that have probability π, given the observation, measurement/or and preparation sentences at point k. These probabilities are calculable by use of the formulae we gave in section 2.4.3.

Adopting a probabilistic truthmaker semantics, we want to have that

$$p \in C^{\pi}_{QM}(X) \Rightarrow \otimes\ TM\,[X] \overset{\pi}{\rightrightarrows} TM\,(p)$$

for all X, p and π. As we mentioned in the last chapter, the converse is generally too strong, and we will see why later. Let $\mathcal{N}$ be a probabilistically necessitarian metaphysics $\langle E, N \rangle$, whose entities will be referred to as *events*. These events will be truthmakers for all statements of QM. Just as there are three different kinds of sentence, these entities can be classified according to which kind of sentences they are truthmakers for. Let $E = E_P \cup E_M \cup E_O$, where

$$\begin{aligned} E_P &= \bigcup TM\,[P] \\ E_M &= \bigcup TM\,[M] \\ E_O &= \bigcup TM\,[O] \end{aligned}$$

We should not assume from the outset that E_P, E_M and E_O are disjoint. For one thing, if sums of events make up events, then the sum of a measurement event and an observation event would make true sentences both about measurement and about observation. What we *can* know about E are things like the following: if e is a measurement that makes true the sentence "observable $\mathbf{A}$ is measured at t", then all of the possible worlds that contain e will contain *some* entity e' that makes true a sentence of the form "the value of observable $\mathbf{A}$ at t is in the set V", for some Borel set V. e', in turn, will exist in all worlds that contain some entity e'' which makes true the sentence "The value of observable $\mathbf{A}$ at t is in the set V'", where $V \subseteq V'$.

Some typically quantum-mechanical theorems can also be extracted from this scheme. For instance, we have that $TM\ (m_1) \overset{0}{>\!\!\!\Rrightarrow} TM\ (m_2)$ whenever $t(m_1) = t(m_2)$ and $\mathbf{O}(m_1)$ does not commute with $\mathbf{O}(m_2)$, so that the performance of a measurement excludes the possibility that a measurement incompatible with it has been performed at the same time. We also have that the entities whose times occur before an entity e generally only give probabilities for the occurrence of those after them.

However, in the necessitarian metaphysic, *all* combinations of entities have probabilities. This follows from our having defined the probabilistic necessitation relation $\overset{\pi}{>\!\!\!\Rrightarrow}$ so it can be interpreted as "the proportion of worlds that contain all of X which also contain some Y is π". But the quantum mechanics itself does not specify all probabilities, but only those of observations, given measurements. We do not in general do not have $o_1 \vdash^{\pi}_{QM} o_2$ for *any* value of π, for different observation statements o_1 and o_2. This means that in the quantum theory, we can *not* have a probability that a certain observation will be followed by another. Such probabilities are only available in the case where some measurement is actually made, and when the system has been prepared correctly.

This fact is sometimes disguised by the wording "the value of observable $\mathbf{A}$ at t is in the set V" of an observation statement o. There are two radically different ways this can be interpreted: actualistically and subjunctively. On the actualist reading, o entails that the experiment required for observing $\mathbf{A}$ actually has been carried out, and thus allows inference to a corresponding measurement sentence. But that measurement sentence does not, on its own, entail any measurements at any other time. This reading, which may be traced to Bohr's interpretation of QM, is the one we have adopted here.

On the other hand, if we read o subjunctively—as saying that if $\mathbf{A}$ *were* measured at t, then its value would be in V—this does not allow us to infer probabilities for another observation o' such that $t(o) \neq t(o')$ either. That the value of o would have a value in V *if* $\mathbf{A}$ were measured does not mean that it *is* measured. And since whether $\mathbf{A}$ is measured or not at t makes a crucial difference to the probabilities of observations after t, it is not possible to assign probabilities in the "lateral" way, straight from observation to observation.

The difference between the readings is fundamental, and the subjunctive interpretation is largely to blame for the unclatiry in the characterisation of what constitutes an "element of physical reality" in the famous EPR paper (Einstein et al., 1935). As Bohr points out, it is only in the context of a concrete measurement that it makes sense to talk about observables having values (Bohr, 1958, pp. 59–61; for criticism cf. Bell, 2004, pp. 155–156). But this has to be an *actual* measurement, and not merely a counterfactual one. This is why our theory QM allows inference of observation sentences only from sets of sentences that contain measurement sentences.

Just as measurement sentences are irreducible to observation sentences, measurements are distinct from observations. While every observation presupposes some definite measurement and thus could be seen as a part of that measurement, a measurement only gives probabilities to observations. Suppose that we attempted to "split up" a measurement m, so that each observation was to correspond one-to-one with a variant of this measurement:

$$\begin{tikzcd}
 & o_1 \arrow[dl, shift left] & & & m_1 \arrow[r, shift left] & o_1 \arrow[l, shift left] \\
m \arrow[ur, dotted, shift left] \arrow[r, dotted, shift left] \arrow[dr, dotted, shift left] & o_2 \arrow[l, shift left] & {} \arrow[r, Rightarrow] & {} & m_2 \arrow[r, shift left] & o_2 \arrow[l, shift left] \\
 & o_3 \arrow[ul, shift left] & & & m_3 \arrow[r, shift left] & o_3 \arrow[l, shift left]
\end{tikzcd}$$

Here, dotted arrows indicate probabilistic necessitations, and solid arrows deterministic ones. In the right-hand diagram, we *can* identify each observation with a unique measurement, so suppose we took o_1 to make true not only "observation 1 was made", but also "measurement 1 was made"? This, however, would distort the inferential structure of quantum mechanics. Suppose that we *do* know that a certain measurement one was made, and want to calculate the probabilities of making a certain observation. Then there are no probabilistic necessitation relations to ground those probabilities, for if we identify measurements and observations, we lose the probabilistic information.

The impossibility of reducing measurements to observations is closely connected to the fact that the probabilistic semantics for QM is incomplete. If it *was* complete, we would also have a way of calculating the probability that a certain measurement is made, given earlier observations.

Incompleteness in a semantics is a sign of weakness of the theory, rather than of the semantics. There are *extensions* of QM for which completeness seems attainable, such as the GRW theory which introduces spontaneous wave function collapses (Ghirardi et al., 1986), or the de Broigle-Bohm theory, which is completely deterministic (Bohm, 1952; Bohm and Hiley, 1993). On its own, however, even a density matrix for the whole world does not give probabilities for the occurence of measurements: they are outside the standard theory.

We also have that preparations, in the absence of further principles, are irreducible to observations and measurements. Every density matrix is, by Gleason's theorem, interdefinable with a probability measure on the algebra of observables (Gleason, 1957; Mackey, 1963). But the *result* of an observation is not in itself sufficient to determine such a probability measure, unless we have a probability measure which describes the state before that observation was made.

To illustrate, suppose that we have made an observation o_1 through a measurement m_1, and that we wish to make a second measurement m_2. Let o_2 be a possible observation of m_2. To calculate the chance of o_2 to occur given m_1, o_1 and m_2, we need something playing the role of a quantum state, which is determined by a density operator. But the only thing we have available is whether the result of m_1 is in the Borel set specified in o_1, and even if we can determine a density operator after such an operation, this requires us to know what the density operator was before.[4]

In his seminal work on quantum mechanics, von Neumann (1955, pp. 328–346) anticipates these problems, but holds the density operator to be specifiable through use of the principle of insufficient reason. However, apart from the many intrinsic problems with the principle in question, this also has the problem that it rules out states not obtainable from a homogeneous mixture through the application of a finite number of measurements. Therefore it seems to me that we are well advised to take preparations to be entities in their own right, different from observations, measurements, or sums thereof.

On the other hand, the so-called *quantum state* does not need any ontological basis. It is sufficiently determined by the preparation event, together with subsequent measurements and observations, and can be seen as an attribute of these. Since there is no state, it does not undergo time evolution. Time enters in the specification of a measurement or an observation, just as in the Heisenberg picture.

The metaphysics of quantum mechanics that follows from adopting probabilistically necessitarian semantics thus commits us to a certain type of entities, which we have called events, and three types of these, which we have called preparations, measurements, and observations. These are connected with probabilistic necessitation relations, and it is these that ground the validity of quantum-mechanical inferences. Among the things that we are *not* committed to are the following:

- A wave function. This function may be useful for us when we want to calculate probabilities, but the quantum theory itself does not need to mention it, and we do not require specific truthmakers for statements about it.

- Microscopic particles. Although these possibly could be constructed from the truthmakers we have (for instance, through

[4] There are exceptions to this, such as where the outcome of a measurement is a pure state. In that case, we can calculate the density operator trivially by assigning that state probability 1 and all others probability 0. But pure states are uncommon in practice, and if the observables in question are continuous, they are unattainable even in theory.

the equivalence that a particle is to be seen as any truthmaker for a certain set of sentences about a given class of observables), these do not play any essential role. It is also not the case that all systems can be separated into independent particles, so even the general usefulness of a particle language can be questioned.

- An observer. This characteristic is shared with any Bohr-style interpretation. While we have truthmakers for observations, nothing in these mentions an actual observer. Whether observations are to be interpreted in physical, mental, or otherwise terms is a question that appears first when we try to place *QM* inside a larger framework, which includes such events as well. The quantum mechanical metaphysics itself is silent on this point.
- An existent multitude of worlds or minds (Everett, 1957; De Witt, 1971), a quantum potential (Bohm and Hiley, 1993), a dynamic state (van Fraassen, 1991), etc.

Despite this, the metaphysics is sufficiently rich, in the sense that all truths have truthmakers. Now, it may be thought that this is because of the poverty of our language: without connectives, we cannot say things like "if measurement $\mathbf{A}$ is performed at t, then there is a probability π that the result will be in (a, b)". But we can add connectives. For instance, $TM\ (\neg o)$, for an observation sentence, can be taken to be the set of observations whose results are incompatible with s. The truthmakers of $o_1 \wedge o_2$ are certain observations o_x such that $\mathbf{O}(o_x)$ is an observable that corresponds to a projection onto a subspace included in both the subspaces projected onto by $\mathbf{O}(o_1)$ and $\mathbf{O}(o_2)$.

Adequate connectives for material implication are notoriously hard to design for logics that capture quantum-mechanical reasoning (see Dalla Chiara and Giuntini, 2002, §3). A connective for *strict* implication is different, however. We can take $s_1 \xrightarrow{\pi} s_2$, for any real number π in $[0, 1]$, to be true iff $TM\ (s_1) \overset{\pi}{\rightrightarrows} TM\ (s_2)$. But in any probabilistically necessitarian semantics, this holds in all worlds, or it holds in none. Therefore, all true instances will have everything as truthmakers. It is, so to say, a truth of the framework rather than of the world.

We will not go into detail concerning how all connectives are to be defined. It appears, however, that we do not need to introduce any new kinds of entities in order to settle the truth-values of complex sentences as well.

An interesting point of note is that the metaphysics, while necessarily nonlocal (as any metaphysics of quantum theory must be, see Bell, 2004), is not completely holistic. A holistic metaphysics would be one in which worlds never overlap, so that from the knowledge of one thing one can draw inferences about everything else that exists. If truthmaking is non-effective, all true sentences have the same truthmakers in such worlds. To show that this metaphysics is non-holistic, it is sufficient to find entities e_1, e_2 such that $\{e_1, e_2\} \overset{1}{\not\succ\!\!\Rrightarrow} \varnothing$ and $\{e_1\} \overset{1}{\not\succ\!\!\Rrightarrow} \{e_2\}$. Such entities do not allow the inference of the existence or non-existence of the other, from the existence of the one. But almost any observations made at different times fulfil these conditions, and even at the same time, almost all observations whose observables commute fulfil them as well.

This means that we very well can see a quantum mechanical system as made up from independent parts. However, these parts are not spatial, but *logical*. Some of them have essential spatial extension, such as measurements and observations of a particle's position. Others do not, such as momentum or spin measurements. While these, in practice, are always made *somewhere* (a Stern-Gerlach apparatus, for instance, is certainly a spatiotemporal object), this does not entail that the system measured itself is spatial.

If one may be allowed a bit of wild speculation, this could be interpreted as an indication that the metaphysics of quantum mechanics, while not necessarily holistic, is not in itself spatial either. Space (or more generally spacetime) could appear as a macroscopic statistical phenomenon of the same class as, say, temperature. If this is so, then it could be possible to find rare violations of relativity on the microscale. This would, in turn, explain how relativity fits with quantum mechanics, despite the fact that they seem to depend on contradictory presuppositions.

One interpretation of quantum mechanics along these lines is the theory of causal sets by Sorkin and his collaborators (Sorkin, 1989;

Bombelli, 1987). According to causal set theory, spacetime has the structure of a locally finite partial order $\preccurlyeq$ with the interpretation that $a \preccurlyeq b$ iff a can influence b causally. The condition of local finiteness then allows the derivation of a volume for each set of points, and this, according to a theorem of Malament (1977), is enough to determine the entire structure of spacetime.

The causal set interpretation is formulable as a necessitarian metaphysic, since a partial order is nothing but a particularly simple form of necessitation relation (i.e. a singular deterministic one). But the necessitation structure also opens up for generalizations. For instance, it allows one to pull the dynamics into the model itself, since a necessitarian metaphysics can describe possible models as well as actual ones. Furthermore, it would be possible to use spacetime regions rather than points as a basis for the structure. We would then get a kind of spacetime mereology, which could prove to be useful for framing questions on the relation between quantum mechanics and general relativity.

7.6 Mind and Metaethics

Many parts of philosophy have metaphysical underpinnings. In this final section, we will take a brief look at applications of truthmaker semantics to the philosophy of mind and metaethics. Starting out with the philosophy of mind, one of the fundamental metaphysical problems in this area can be posed as: what is the mind, and how is it related to the brain? In contemporary philosophy, the question has often centered on *qualia*: purported subjective qualitative properties of experience. What are these? Are they somehow reducible to physical entities?

We will focus on three influential types of answer, among which one comes in at least two important varieties.

Identity: Qualia are physical entities.

Distinctness: Qualia are distinct from physical entities.

—*Supervenient*: Qualia are determined by physics.

—*Non-supervenient*: Qualia are independent of physics.

Eliminativism: Qualia do not exist.

We here consider supervenience only in the form of supervenience on the *physical*. It, however, is easy to generalise the discussion to the case of supervenience on the *non-mental*, and in our second trio of view below, to supervenience on the *non-moral*.

The first of the answers is often associated with the earlier identity theories of mind (Lewis, 1966; Armstrong, 1968), but also some later functionalist theories fit in here (Block and Fodor, 1972). Other forms of functionalism, however, are supervenient dualisms, such as Putnam's (Putnam, 1975b, pp. 429–440). A non-supervenient distinctness thesis is defended in Chalmers's *The Conscious Mind* (Chalmers, 1996). Finally, eliminativism about qualia is most well-known as advocated by Dennett (1988).

To proceed, and to place our spotlight on the metaphysical question proper, we assume that positive claims about experiences (such as "Mary has an experience of red at t") can be true. While not uncontroversial, this assumption seems defensible so long as we do not presuppose any specific interpretation of "experience". It is a fact that experience claims, just as any other claims, can stand in inferential relations. Let

$$\begin{aligned} a = {} & \text{Mary has an experience of red at } t \\ c_1 = {} & \text{Mary has a visual experience at } t \\ c_2 = {} & \text{Mary has an auditory experience at } t \end{aligned}$$

Then, unless Mary has synaesthesia, we have that $a \vdash c_1$ and $a \nvdash c_2$. In a truthmaker semantics, it follows that

$$TM\left(a\right) \rightarrowtail TM\left(c_1\right)$$

$$TM\left(a\right) \not\rightarrowtail TM\left(c_2\right)$$

Let us call whatever makes an experience claim true a *quale*. Whatever qualia are, this makes them fulfil several of the expected desiderata. For one thing, they satisfy the *esse est percipi* principle: by being truthmakers for claims about experiences, they cannot exist unless their corresponding experience claims are true, and this will entail that someone in fact experiences them.

We can see at once that Eliminativism is incompatible with the truth of positive experience claims given truthmaker semantics. If a positive experience claim is true, then there must be something that makes it true, and any such thing is a quale. So let us henceforth concentrate on Identity and Supervenient Distinctness. Both of these postulate a certain type of relationship between physical states and qualia. The supervenience thesis inherent in *both* of these can be expressed as the claim that the physical entities determine the qualia in the sense that if ω_1 and ω_2 contain the same physical entities, they also contain the same qualia. This entails that there is a function $\mathcal{Q} : \wp(P) \to \wp(Q)$, where P is the set of all possible physical entities, and Q the set of all possible qualia, such that for any world ω,

$$\omega \cap Q = \mathcal{Q}(\omega \cap P)$$

This is what Kim refers to as *global* supervenience (Kim, 1984). It is a very weak kind of relationship, and in many cases we are interested not only in the condition that the whole of physical reality determines the whole of qualia space, but also whether parts of the physical world determine parts of the mental. When this holds for single qualia, we have

$$\omega \cap Q = q[\omega \cap P]$$

for some partial function $q : \wp(P) \to Q$ with the interpretation that $a \in q(X)$ iff a occurs in any world in which the X's occur. This means that not only is the set of all qualia in a world determined by its physical entities, but qualia are so determined individually as well.

For the identity theory, we have that q is the identity function wherever it is defined, and thus it follows that $Q \subseteq P$, i.e. all qualia are physical. But even in the distinctness theory, qualia claims have physical truthmakers—at least if we allow truthmaking by pluralities, or have a mereological metaphysics. As we have defined it, a truthmaker for p is something that is sufficient for the truth of p. But if a collection of physical entities is sufficient for the existence of a quale, and that quale is sufficient for the truth of a positive experience claim p, then the collection of physical entities itself must be sufficient for p's truth as well.

What work do the *qualia* do then? It is hard to say; their supervenience base can do the work of truthmaking just as well as the qualia themselves, so it is not clear what the qualia are for. If we consider two worlds, exactly alike physical

Indeed, in a truthmaker semantics, the purported differences between the identity view and the supervenient distinctness view almost disappear. The distinction itself really presupposes a traditional correspondence theory, in which non-equivalent claims cannot be made true by the same thing. In a truthmaker theory, problems such as the so-called "multiple realisability" argument tend to lose their bite: *all* claims can be made true by different kinds of things, so there is nothing special about mental states being thus multiply realisable.

To really distance oneself from the identity theory, one has to deny supervenience as well, and hold that the qualia are underdetermined by the physical world. But this means that one has to go for a fully-fledged duality theory, on which the mental floats free of the physical. The problem with this is, of course, that we can no longer avail ourselves of interpersonal comparisons or other paradigms of scientific inquiry if we are to explore such domains. All our evidence of others' mental life is physical, and if physical observation does not allow us to infer things about the mental, then nothing else will either.

Let us now briefly consider metaethics. Much 20th century debate in the metaphysical parts of this field bears close resemblance to that going on in the philosophy of mind. Instead of the triad we discussed earlier, we have

Naturalism: Moral facts are physical facts.
Realism: Moral facts are non-physical.
—*Supervenient*: Moral facts are determined by physics.
—*Non-supervenient*: Moral facts are independent of physics.
Nihilism: Moral facts do not exist.

Of these, Naturalism corresponds to the Identity Theory in the philosophy of mind, Realism corresponds to Distinctness, and Nihilism to Eliminativism. Nihilism is sometimes confused with non-cognitivism, which holds that moral statements do not permit of truth and falsity, but is different, since one very well can hold that moral claims are true iff their corresponding facts exist, but since no such facts exist, all moral claims are false. The most well-known philosopher to put forth such a theory is John Mackie, in *Ethics: Inventing Right and Wrong* (Mackie, 1977).

As long as we adopt truthmaker theory, Nihilism is incompatible with the truth of positive moral claims, just as Eliminativism is incompatible with positive truths about qualia. As for Naturalism and Realism, the same lessons can be drawn as in the philosophy of mind. A Realism that holds moral facts to supervene on the physical facts, such as Moore's (Moore, 1922), runs into exactly the same problems as Supervenient Distinctness in the philosophy of mind: the "moral" truthmakers do not play any role, as any true moral claim must have physical truthmakers as well. They can be cut off using Occam's razor, without any loss in representative power.

These considerations seem to indicate that *supervenient* dualism realism and moral realism are red herrings, at least if we adopt a truthmaker semantics (this actually holds for general necessitarian semantics as well). But how would a *non*-supervenient dualism or moral realism work? Considering dualism first, if a quale does not supervene on the physical entities, there are possible worlds—exactly alike physically—both where this quale exists, and where it does not. Whether this makes sense or not naturally depends on what we mean by "possible" here. Perhaps these worlds are *physically* possible, but not *psycho*physically so, as Chalmers (1996, p. 213) argues?

Something similar can be said for moral realism. If Naturalism is wrong, there can be no determination of the moral by the physical. But we could envisage moral-physical laws nevertheless, even if these cannot have the same modal force as the physical laws themselves. The problem is to explain how we can have epistemic access to such laws.

If we back down on the claim that experience claims and moral claims are *true*, we regain the possibility of adopting Eliminativism or Nihilism. Both claims about qualia and claims about right and wrong lack several of the properties that we generally associate with truth, most important of which are intersubjective criteria of confirmation or disconfirmation. On this reading, they are to be taken not claims at all, but as what may be called *pseudoclaims* – things that have the *syntactic* structure of claims, but may not be such.[5]

For pseudoclaims, *truth* may not be the semantic value we are after. For morals, *imperance* could be more important, as Hare (1952) argued. For experience pseudoclaims, we can take inspiration from the later Wittgenstein, and say that these do not play the role of assertions either (Wittgenstein, 1953, §244). Thus we could say that experience pseudoclaims can be *avowed*, but not strictly *true* or *false*, since they are non-assertory.

We can use moral or experiential pseudoclaims to make inferences, as long as they are placed in the language of a many-valued theory. Given the right consequence relations, we can infer from claims to pseudoclaims and back. It is only the semantic interpretation that differs, so pseudoclaims can still have as central a logical role as claims do. Denying the possibility that they can be true or false therefore does not need to expose us to the so-called Frege-Geach problem (Geach, 1965). This problem only appears if we require all forms of semantics to be of the Tarskian kind, or if we disregard the possibility of using semantic values other than t and f.

[5] We have, of course, not assumed that claims need to have a syntactic structure at all. Nevertheless, it is difficult to imagine that the question of truth or falsity would even be posed for experience and moral claims, unless we put these in sentential form, and compare them structurally to more paradigmatic claims such as "there are three apples in this basket". I conjecture that much of the attraction in assigning truth-values to these claims comes from thinking of them syntactically.

Epilogue: Models and Metaphysics

After describing the various types of scientific mind in his 1870 address to the mathematical and physical sections of the British Association, James Clerk Maxwell went on to state their relevance to scientific practice:

> For the sake of persons of these different types, scientific truth should be presented in different forms, and should be regarded as equally scientific whether it appears in the robust form and the vivid colouring of a physical illustration, or in the tenuity and paleness of a symbolical expression. (Maxwell, 1870, p.220)

"This is almost the most important thing Maxwell ever wrote" comments John Gribbin—himself a physicist—in his history of science (Gribbin, 2002, p.430). But wherein lies this great importance?

Maxwell was one of the originators of the model-building (or *pictorial*) view of science, which we already have mentioned Hertz's adherence to. During the 20th century, physics grew more and more abstract. But still, as we have entered the 21st, most physicists primarily work using representations: for quantum mechanics often a wave function, a set of matrices, or a collection of paths (Dirac, 1958, ch. 3, Feynman and Hibbs, 1965). The pictures, or models, are crucial to our thinking.

This, however, should not lead us into the temptation to close our eyes to their conventionality.

Poincaré took up this thread, first with regard to the structure of space, and then with regard to science in general (Poincaré, 1905). This influenced Carnap already in his doctoral dissertation, although the influence was not to manifest itself in full until later, when he formulated his principle of tolerance and his characterisation of *external* questions as ones to be settled by convention rather than empirical or deductive investigation.

Despite his animadversions to metaphysics, Carnap has been a large influence throughout this book. The metaphysics we have advocated is not the metaphysics that Carnap revolted against.[6] I have proposed that we base metaphysics on model theory in order to make it relevant to science, and in extension, to human affairs, knowledge and understanding. But a theory of models is a kind of language, and as such it is shot through with convention. This makes all metaphysics conventional at heart. But does this mean that we have lost the world? That metaphysics, on our interpretation, does not after all concern how reality is, but only how we represent it?

Although I agree that this question is natural, it is based on a faulty and misleading picture of thought, language and theory. It is impossible to think about, talk about, or even experience the world without conceptualisation, so all metaphysics will presuppose a certain amount of convention just due to the conventionality of concepts. Where conventionality stops and fact starts is, as Quine rightly pointed out, quite vague. The line itself is thus a matter of convention, and metaphysics, rather than being purely about the world or purely about our representations, is just like any theory a pale grey lore of them both.

[6] In *Meaning and Necessity* (Carnap, 1956, p. 43) Carnap explicitly took exception to Quine's use of the word "ontology" for the set of objects falling under the range of a language's variables. He thought that such a word would invite philosophers to attack such questions using metaphysical speculation, rather than considerations of theoretical usefulness. Now we know that he was right to worry: Quine's position in *On what there is* has been misused as an invitation to intuition-based speculative metaphysics ever since. I can only hope that my appropriation of the word "metaphysics" here will not be taken in the same way.

Furthermore, metaphysics as advocated here is not incompatible with realism. We have assumed rather than denied that we can have reason to believe theories (scientific or otherwise) to be *true* rather than merely empirically adequate, useful, well-corroborated, or anything else. We have also accepted the fact that truth means agreement with reality, or in the more general case, agreement with what the theory is about. This does not, on itself, determine that reality very much, but it does so in the presence of a semantics. The choice of semantics is where conventionality enters.

A point of importance to classical realism, on which we have been largely silent, is to what extent the world is dependent or independent of what we think or say about it. One of our model spaces (the space $\mathcal{Ch}$ of coherence models) is definitely dependent on belief, since its models *are* sets of beliefs. But the other spaces are neither belief dependent nor belief independent on their own. Truthmaker semantics, for instance, can can be dependent or independent depending on what the metaphysics' possible entities are.

Still, it may seem that the conventional aspects involved would invoke a necessary dependence between reality and convention – as if reality itself somehow was a product of stipulation. But this worry is hard to even state coherently. To raise the question of truth for a claim, we generally need to place it inside a framework, i.e. a theory. Only inside such a theory does a claim have enough inferential connections to allow it to be tested. When used as a framework, however, the theory is not true or false, but sound or unsound. Soundness means that no matter what the theory's subject may be like, the theory's inferences about it are truth-preserving. To raise the question of whether the *framework* is true, it will have to be considered as a theory inside a larger framework.

When *testing* a theory, we thus always need to place it inside a framework. But truth itself does not entail anything about testing—it is not in itself an epistemic concept. We have explicated "p is true", when p is a claim in a framing theory F about the world, as "the world is one of p's models". This does presuppose that the world has a certain kind of structure, as given by the model space we use in the semantics. What if the world does not have this kind of structure? For instance, suppose we use a Tarskian semantics, and the world is *not* a Tarskian model?

We must not lose sight of the fact that even this question is framed in a language, or otherwise we would not understand it. To ask "is the world a Tarskian model?", if it is not to be trivial, one has to entertain the possibility that it is something else. This means that the semantics we use for interpreting that question have to employ a model space that contains more than just the Tarskian models. Nevertheless, it *is* semantics-relative. Relative to a certain model space, the world is a Tarskian model, and relative to another, it isn't.

A staunch old-fashioned realist may of course, at this point, attempt to hold that what matters is whether the *intended* semantics contains the world as a model or not. But how are we to interpret this? The possible "intendedness" of a certain semantics would be a property attached to it by the proponent of a certain theory. This property must be verbalisable if it is to be relevant to communication and science. There must therefore be a way to discuss semantics, and to hear if one interpretation or another is intended. This discussion will however itself have to be conducted in a language, and its results will depend on how this language is interpreted. We cannot "step outside" language or conceptualisation, and no theory ever interprets itself.

Of course, this does not mean that we cannot have objectivity *relative* to a theory (or a model space). This is, as Quine noted, just like geometry. But while large parts of geometry are accessible in a coordinate-invariant form, there is no such thing as a framework-invariant theory. We need theory to discuss theory.

One coordinate system that we have spent much time on in this book is the one spanned by necessitarian semantics. This is a way to interpret theories that lends itself well to metaphysical investigations: limiting questions about models to questions about what exists in them means that it is easy to keep track of the information encoded in such a model. In the case where all combinations of entities make up possible worlds, the information content I (in Shannon's well-known sense of the word) in being told that a specific model $\mathfrak{M}$ is the actual one is simply the number of possible entities $|E|$, since we can specify which world is the actual one by saying for each $e \in E$ whether it exists or not, and this takes answers to $2^{|E|}$ yes-or-no questions.

In metaphysics where some combinations of entities are impossible, saying which model is the actual one gives less information. According to the usual definition, we have that $I = \log_2(|\Omega|)$, where Ω, as before, is the set of possible worlds. These information-theoretical notions are relevant to the question of truth, since another reasonable way to express the correspondence criterion would be the definition

$$p \text{ is true } \underset{def}{=} \text{ the world encodes the information given by } p$$

This pinpoints realism as the requirement that the information given in a true claim must come from the *world*, rather than from anything else. Encoding can of course be done in more than one way, and we must therefore settle on an encoding scheme, or rather an *algorithm* which converts the data, as it is in the world, to the form it has in p. Such an algorithm plays the same role as a semantics.

The direct information-theoretical characterisation of truth is available for all kinds of necessitarian semantics, but the one that we have concentrated on primarily here is truthmaker theory. The reason for this is the current popularity of it in metaphysics, and it is therefore time for us to come to some kind of verdict on its advantages and disadvantages.

General necessitarian semantics can be summarised in the slogan "truth supervenes on being". This is properly taken as a stipulation rather than a substantial thesis. It lays down principles for how to individuate objects, and thus also for what we mean by the word "object" (or in our case, "entity"). This stipulation should therefore be criticised according to its utility, rather than to pretentions of factual correctness. It definitely conforms to the correspondence criterion, since it characterises truth as dependent on the what the world is like.

It is harder to motivate truthmaker theory in the same way, either in its singular or its plural form. *Why* should every truth be based on the existence of something? Rodriguez-Pereyra's argument that truthmaking is a relation and relations relate entities does not work if one takes truthmaking to *not* be a relation in any substantial sense. It may very well be *true* that a makes p true, without there being any metaphysically "thick" relation holding between them, if this is something that follows from our adaption of a semantics rather than from any fact about the world. To make an analogy, I am able to imagine a certain crater on the far side of the moon, but this does not entail that there is any substantial connection between my imagination and said crater, even if it does exist. In fact there cannot be, as no information can travel instantaneously like that.

All that correspondence *really* requires is that truths somehow are related to the world, and this does not mean that they have to be related to specific things, rather than to reality as a whole. To assume that they do relate to parts of the world is to assume a form of truthmaker theory, and thus it cannot be used as an argument for said theory the way Rodriguez-Pereyra does.

Thus I hold that rather than by the kind of rationalistic arguments traditionally given for it, truthmaker theory must be motivated by its usefulness. How enlightening are the pictures we can paint using it? How useful are they to science? Its main claim to these is, I believe, the correspondence it sets up between the necessitarian structure of metaphysics and the logical structure of a true theory. But to some extent, it shares this property with general necessitarian semantics as well. In chapter 6, we proved the isomorphism not only of truthmakers and claims, but of truthmaking circumstances and claims as well.

Still, truthmaker theory has the advantage that the correspondence between world and theory becomes particularly simple in it. Not as simple as in, say, a straight correspondence theory, but simpler than in general necessitarian metaphysics, and a straight correspondence theory seems even harder to motivate. In a truly Carnapian fashion, we can *decide* to require truth to be grounded in entities. We should not imagine that this decision itself gives us any information about the world. It simply sets up a convenient framework for us to discuss how the world is, and relative to this framework, questions about the world can be asked. When we wish to ask instead whether truthmaker theory itself is true—for example, by asking whether every world in which p is true contains something that does not exist in any world in which p is false—we should use a wider framework. But as we saw in section

2.5, we cannot expect there to be a *widest* one. All theory, and metaphysics as well, is perspective-dependent. We always theorise from the perspective of a framework, but nothing stops us from changing that perspective to a more enlightening one, if the one we are viewing the problems from at the moment does not give us the vantage point we desire.

Bibliography

Adámek, Jiří, Herrlich, Horst, and Strecker, George E. (2004). *Abstract and Concrete Categories.* Wiley, online ed.

Armstrong, David (1968). *A Materialistic Theory of the Mind.* Routledge & Kegan Paul.

—— (1989). *Universals: An Opinionated Introduction.* Westview Press.

—— (1997). *A World of States of Affairs.* Cambridge University Press.

—— (2004). *Truth and Truthmakers.* Cambridge University Press.

Austin, John L. (1950). "Truth". In "Philosophical Papers", Oxford University Press.

Awodey, Steve (2006). *Category Theory.* Oxford University Press.

Bacon, John (1995). *Universals and Property Instances: The Alphabet of Being.* Oxford University Press.

Bednarczyk, Marek A., Borzyszkowski, Andrzej M., and Pawolwski, Wieslaw (2007). "Epimorphic Functors".

Bell, John S. (2004). *Speakable and Unspeakable in Quantum Mechanics.* Cambridge University Press, 2nd ed.

Belnap, Nuel and Anderson, Alan Ross (1975). *Entailment: The Logic of Relevance and Necessity.* Princeton University Press.

BIBLIOGRAPHY

Bigelow, John (1988). *The Reality of Numbers: A Physicalist's Philosophy of Mathematics*. Oxford University Press.

Birkhoff, Garrett (1967). *Lattice Theory*. AMS Colloquium Publications, 3rd ed.

Block, Ned and Fodor, Jerry (1972). "What Psychological States are Not". *Philosophical Review*, 81:159–181.

Bohm, David (1952). "A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables, I and II". *Physical Review*, 85:166–193.

Bohm, David and Hiley, Basil J. (1993). *The Undivided Universe: An Ontological Interpretation of Quantum Theory*. Routledge.

Bohr, Niels (1958). *Atomic Physics and Human Knowledge*. Science Editions.

Bolzano, Bernard (1837). *Wissenschaftslehre*. Seidel.

Bombelli, Luca (1987). *Space-Time as a Causal Set*. Ph.D. thesis, Syracuse University.

Boolos, George (1975). "On second-order logic". *Journal of Philosophy*, 72:509–527. Also in Boolos (1998).

—— (1984). "To be is to be the value of a variable (or some values of some variables)". *Journal of Philosophy*, 81:430–450. Also in Boolos (1998).

—— (1998). *Logic, Logic and Logic*. Harvard University Press.

Bovens, Luc and Hartmann, Stephan (2003). *Bayesian Epistemology*. Oxford University Press.

Brandom, Robert (1994). *Making it Explicit*. Harvard University Press.

Carnap, Rudolf (1937). *The Logical Syntax of Language*. Kegan Paul.

—— (1942). *Introduction to Semantics*. Harvard University Press.

—— (1943). *The Formalization of Logic*. Harvard University Press.

—— (1950). *Logical Foundations of Probability*. University of Chicago Press.

—— (1956). *Meaning and Necessity*. University of Chigaco Press, 2nd ed.

Chalmers, David (1996). *The Conscious Mind.* Oxford University Press.

Chang, C.C. and Keisler, H. J. (1973). *Model Theory.* North Holland.

Church, Alonzo (1944). “Review of ‘Formalization of Logic”’. *Philosophical Review*, 53:493–498.

Cummins, Robert (1998). “Reflection on Reflective Equilibrium”. In DePaul, Michael R. and Ramsey, William (eds.) “Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry”, Rowman and Littlefield.

Dalla Chiara, Maria Luisa and Giuntini, Roberto (2002). “Quantum Logics”. In Gabbay, Dov and Guenthner, Franz (eds.) “Handbook of Philosophical Logic, vol. 6”, Kluwer Academic Publishers.

Daly, Chris (2005). “So Where’s the explanation?” In Beebee, Helen and Dodd, Julian (eds.) “Truthmakers: the contemporary debate”, Oxford University Press.

Davey, B. A. and Priestley, H. A. (2002). *Introduction to Lattices and Order.* Cambridge University Press, 2nd ed.

De Witt, Bryce (1971). “The Many-Universes Interpretation of Quantum Mechanics”. In d’Espagnat, Bernard (ed.) “Foundations of Quantum Mechanics”, Academic Press.

Dennett, Daniel (1988). “Quining Qualia”. In Marcel, A and Bisiach, E. (eds.) “Consciousness in Contemporary Science”, Oxford University Press.

Dirac, Paul A. M. (1958). *The Principles of Quantum Mechanics.* Oxford University Press, 4th ed.

Dodd, Julian (2001). “Is Truth Supervenient on Being?” *Proceedings of the Aristotelian Society*, 102:69–86.

Duhem, Pierre (1954). *The Aim and Structure of Physical Theory.* Princeton University Press.

Dummett, Michael (1976). “What is a Theory of Meaning? (II)”. In Evans, Gareth and McDowell, John (eds.) “Truth and Meaning”, Oxford University Press. Also in Dummett (1993).

—— (1978). *Truth and Other Enigmas.* Harvard University Press.

—— (1981). *Frege: Philosophy of Language.* Harvard University Press, 2nd ed.

—— (1991a). *Frege: Philosophy of Mathematics.* Duckworth.

—— (1991b). *The Logical Basis of Metaphysics.* Harvard University Press.

—— (1993). *The Seas of Language.* Oxford University Press.

—— (2000). *Elements of Intuitionism.* Oxford University Press, 2nd ed.

Einstein, Albert (1905). "On the Electrodynamics of Moving Bodies". *Annalen der Physik*, 17:891–921.

Einstein, Albert, Poldolsky, Boris, and Rosen, Nathan (1935). "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review*, 47:777–780.

Ellis, Brian (2001). *Scientific Essentialism.* Cambridge University Press.

Etchemen dy, John (1990). *The Concept of Logical Consequence.* CSLI Publications.

Everett, Hugh (1957). "'Relative State' Formulation of Quantum Mechanics". *Reviews of Modern Physics*, 29:454–462.

Feferman, Solomon and Feferman, Anita Burdman (2004). *Alfred Tarski: life and logic.* Cambridge University Press.

Feynman, Richard P. (1963). *The Feynman Lectures on Physics, 3 vols.* Addison-Wesley Publishing Company.

Feynman, Richard P. and Hibbs, Albert R. (1965). *Quantum Mechanics and Path Integrals.* McGraw-Hill.

Field, Hartry (1977). "Logic, meaning and conceptual role". *Journal of Philosophy*, 69:379–409.

Fine, Kit (2002). "The Varieties of Necessity". In Gendler, T. S. and Hawthorne, J. (eds.) "Conceivability and Possibility", Oxford University Press. Also in Fine (2005).

—— (2005). *Modality and Tense.* Oxford University Press.

Fox, John (1987). "Truthmaker". *Australasian Journal of Philosophy*, 65:188–207.

Frege, Gottlob (1884). *Die Grundlagen Der Arithmetik.* Breslau.

Fremlin, David H. (2000). *Measure Theory, vol 1.* Torres Fremlin.

French, Steven and Ladyman, James (2003). "Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure". *Synthese*, 36:31–66.

Gärdenfors, Peter (1988). *Knowledge in Flux.* MIT Press.

Geach, Peter T. (1963). "Quantification Theory and the Problem of Identifying Objects of Reference". *Acta Philosophica Fennica*, 16:41–52.

—— (1965). "Assertion". *Philosophical Review*, 74:449–465.

—— (1967). "Identity". *Review of Metaphysics*, 21:3–12.

Gentzen, Gerhard (1934). "Untersuchungen über das logische Schließen". *Mathematische Zeitschrift*, 39:410–431.

Ghirardi, G. C., Rimini, A., and Weber, T. (1986). "Unified dynamics for microscopic and macroscopic systems". *Physical Review D*, 34:470–491.

Giere, Ronald N. (1979). *Understanding Scientific Reasoning.* Holt, Rinehart and Winston.

Gleason, Andrew (1957). "Measures on the Closed Subspaces of a Hilbert Space". *Journal of Mathematics and Mechanics*, 6:885–893.

Goodman, Nelson (1951). *The Structure of Appearance.* Harvard University Press.

Grätzer, George (1979). *Universal Algebra.* Springer-Verlag, 2nd ed.

Gribbin, John (2002). *Science: A History.* Penguin.

Guay, Alexandre and Hepburn, Brian (2009). "Symmetry and Its Formalisms: Mathematical Aspects". *Philosophy of Science*, 76:160–178.

Hare, Richard Mervyn (1952). *The Language of Morals.* Oxford University Press.

Harman, Gilbert (1974). “Meaning and Semantics”. In Munitz, M. K. and Unger, P. (eds.) “Semantics and Philosophy”, NYU Press.

Harrop, Ronald (1959). “Concerning Formulas of the types $A \to B \vee C$, $A \to (Ex)B(x)$ in intuitionistic formal systems”. *Journal of Symbolic Logic*, 25:27–32.

Heil, John (2003). *From an Ontological Point of View.* Oxford University Press.

Henkin, Leon (1961). “Some remarks on infinitely long formulas”. In “Infinitistic Methods”, Pergamon Press.

Hertz, Heinrich (1899). *The Principles of Mechanics Presented in a New Form.* Macmillan and Co.

Hilbert, David and Bernays, Paul (1934). *Grundlagen der Mathematik I.* Springer-Verlag.

Hintikka, Jaakko (1996). *The Principles of Mathematics Revisited.* Cambridge University Press.

—— (1999). “The Emperor’s New Intuitions”. *The Journal of Philosophy*, 96:127–147.

Hodges, Wilfrid (1993). *Model Theory.* Cambridge University Press.

—— (2005). “Model Theory”. In Zalta, Edward N. (ed.) “Stanford Encyclopedia of Philosophy”, The Metaphysics Research Lab, CSLI, Stanford University. URL `http://plato.stanford.edu/archives/fall2005/entries/` `/model-theory/`.

Horwich, Paul (1998). *Truth.* Oxford University Press, 2nd ed.

Hume, David (1739). *A Treatise of Human Nature.*

—— (1779). *Dialogues Concerning Natural Religion.*

Humphreys, Paul (1985). “Why Propensities Cannot be Probabilities”. *The Philosophical Review*, 94:557–570.

Ishiguro, Hidé (1990). *Leibniz’s Philosophy of Logic and Language.* Cambridge University Press, 2nd ed.

Johansson, Ingebrigt (1936). “Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus”. *Compositio Mathematica*, 4:119–136.

Kant, Immanuel (1783). *Prolegomena to any Future Metaphysics.* Hackett Publishing Company. Trans. by James W. Ellington, 1977.

Kim, Jaegwon (1984). “Concepts of Supervenience”. *Philosophy and Phenomenological Research*, 45:153–176.

Kitcher, Philip (1981). “Explanatory unification”. *Philosophy of Science*, 48:507–531.

—— (1989). “Explanatory unification and the causal structure of the world”. In Salmon, Wesley and Kitcher, Philip (eds.) “Scientific Explanation”, University of Minnesota Press.

Kripke, Saul (1981). *Naming and Necessity.* Blackwell Publishers.

Kyburg, Henry E. (1965). “Discussion: Salmon’s paper”. *Philosophy of Science*, 32:174–151.

Ladyman, James and Ross, Dan (2007). *Every Thing Must Go: Metaphysics Naturalised.* Oxford University Press.

Lawvere, F. William (1964). “An elementary theory of the category of sets”. *Proceedings of the National Academy of Sciences of the United States of America*, 50:1506–1511.

Leibniz, Gottfried Wilhelm von (1679). “Elements of Calculus”. In Loemker, Leroy E. (ed.) “Philosophical papers and letters”, pp. 360–370. University of Chicago Press.

—— (1714). “The Monadology”. In Loemker, Leroy E. (ed.) “Philosophical papers and letters”, pp. 1044–1061. University of Chicago Press.

Leonard, Henry S. and Goodman, Nelson (1940). “The Calculus of Individuals and Its Uses”. *Journal of Symbolic Logic*, 5:45–55.

Leśniewski, Stanisław (1916). “Podstawy ogólnej teoryi mnogości. I”. In “Prace Polskiego Koła Naukowego w Moskwie”, Translated as ‘Foundations of the General Theory of Sets. I’ in Leśniewski (1992).

—— (1992). *Collected Works, vol. 1.* Kluwer.

Levi, Isaac (1991). *The Fixation of Belief and Its Undoing.* Cambridge University Press.

Lewis, David (1966). "An Argument for the Identity Theory". *Journal of Philosophy*, 63:17–25.

—— (1973). *Counterfactuals.* Harvard University Press.

—— (1986). *On the Plurality of Worlds.* Blackwell Publishers.

—— (1991). *Parts of Classes.* Basil Blackwell.

—— (2001). "Truthmaking and Difference-Making". *Noûs*, 35:602–615.

Lewitzka, Steffen (2007). "Abstract Logics, Logic Maps, and Logic Homomorphisms". *Logica Universalis*, 1:243–276.

Lowe, E. Jonathan (1998). *The Possibility of Metaphysics.* Oxford University Press.

Luce, R. Duncan, Krantz, David H., Suppes, Patrick, and Tversky, Amos (1990). *Foundations of Measurement*, vol. III. Academic Press.

Lyndon, Roger C. (1959). "Properties Preserved under Homomorphism". *Pacific Journal of Mathematics*, 9:143–154.

MacFarlane, John (2005). "Logical Constants". In Zalta, Edward N. (ed.) "Stanford Encyclopedia of Philosophy", The Metaphysics Research Lab, CSLI, Stanford University. URL `http://plato.stanford.edu/archives/sum2005/entries/logical-constants/`.

Machery, E., Mallon, R., Nichols, S., and Stich, S. (2004). "Normativity and Epistemic Intuitions". *Cognition*, 92:1–12.

Mackey, George W. (1963). *Mathematical Foundations of Quantum Mechanics.* W.A. Benjamin.

Mackie, John L. (1974). *The Cement of the Universe: A Study of Causation.* Oxford University Press.

—— (1977). *Ethics: Inventing Right and Wrong.* Viking Press.

Malament, David B. (1977). "The class of continuous timelike curves determines the topology of spacetime". *Journal of Mathematical Physics*, 18:1399–1404.

Marcus, Ruth Barcan (1961). “Modalities and Intensional Languages”. *Synthese*, 13:303–322.

Martin, Richard M. (1980). “Substance Substantiated”. *Australasian Journal of Philosophy*, 58:3–20.

Maxwell, James Clerk (1870). “Address to the Mathematical and Physical Sections of the British Association”. In Niven, W. D. (ed.) “The Scientific Papers of J. Clerk Maxwell”, Cambridge University Press.

McLarty, Colin (1992). *Elementary Categories, Elementary Toposes.* Oxford University Press.

Melia, Joseph (1995). “On What There’s Not”. *Analysis*, 55:223–229.

Minkowski, Hermann (1908). “Raum und Zeit”. *Physikalische Zeitschrift*, 10:104–111.

Montague, Richard (1970). “English as a Formal Language”. In “Linguaggi nella Società e nella Tecnica”, Edizioni di comunità. Also in Montague (1974).

—— (1973). “The proper treatment of quantification in ordinary English”. In Hintikka, Jaakko, Moravcsik, J. M., and Suppes, Patrick (eds.) “Approaches to Natural Language”, Reidel. Also in Montague (1974).

—— (1974). *Formal Philosophy.* New Haven.

Moore, George E. (1922). “The Conception of Intrinsic Value”. In “Philosophical Studies”, Routledge & Kegan Paul.

Mulligan, Kevin, Simons, Peter, and Smith, Barry (1984). “Truth-Makers”. *Philosophy and Phenomenological Research*, 44:287–321.

Olsson, Erik J. (2005). *Against Coherence: Truth, Probability and Justification.* Oxford University Press.

Pearl, Judea (2009). *Causality: Models, Reasoning, and Inference.* Cambridge University Press, 2nd ed.

Poincaré, Henri (1905). *Science and Hypothesis.* Walter Scott Publishing.

Popper, Karl R. (1959). “The Propensity Interpretation of Probability”. *British Journal for the Philosophy of Science*, 10:25–42.

Putnam, Hilary (1968). “Is Logic Empirical?” In Cohen, Robert S. and Wartofsky, Marx W. (eds.) “Boston Studies in the Philosophy of Science, vol. 5”, Reidel. Also in Putnam (1975b).

—— (1975a). “The meaning of ‘meaning”’. In Gunderson, K. (ed.) “Languauge, Mind and Knowledge”, Minnesota Studies in the Philosophy of Science. University of Minnesota Press. Also in Putnam (1975b).

—— (1975b). *Mind, Language and Reality.* Cambridge University Press.

Quine, Willard van (1948). “On What There Is”. *Review of Metaphysics*, 2:21–38. Also in Quine (1980).

—— (1951). “Two Dogmas of Empiricism”. *Philosophical Review*, 60:20–43. Also in Quine (1980).

—— (1960a). “Carnap and Logical Truth”. *Synthese*, 12. Also in Quine (1976).

—— (1960b). *Word and Object.* M. I. T. Press.

—— (1969). *Ontological Relativity.* Columbia University Press.

—— (1972). “The Variable”. In Parikh, Rohit (ed.) “Logic Colloquium”, Springer Lecture Notes in Mathematics. Also in Quine (1976).

—— (1976). *The Ways of Paradox.* Harvard University Press, 2nd ed.

—— (1980). *From a Logical Point of View.* Harvard University Press, 3rd ed.

—— (1981a). “Responding to David Armstrong”. *Pacific Philosophical Quarterly*, 61. Also in Quine (1981b), pp. 108–184.

—— (1981b). *Theories and Things.* Harvard University Press.

—— (1986). *Philosophy of Logic.* Harvard University Press, 2nd ed.

Rényi, Alfréd (1955). “On a new axiomatic theory of probability”. *Acta Mathematica Hungarica*, 6:285–333.

Restall, Greg (1996). “Truthmakers, Entailment and Necessity”. *Australian Journal of Philosophy*, 74:331–340.

—— (2003). “Modelling Truthmaking”. *Logique et Analyse*, 169–170:211–230.

Rodriguez-Pereyra, Gonzalo (2005). “Why Truthmakers”. In Beebee, Helen and Dodd, Julian (eds.) “Truthmakers: the contemporary debate”, Oxford University Press.

—— (2006). “Truthmaking, entailment, and the conjunction thesis”. *Mind*, 115:957–982.

Roeper, Peter and Leblanc, Hughes (1999). *Probability Theory and Probability Semantics*. University of Toronto Press.

Russell, Bertrand (1985). *The Philosophy of Logical Atomism*. Open Court. Also available in *Logic and Knowledge*.

Ryle, Gilbert (1957). “Theory of Meaning”. In Mace, C. A (ed.) “British Philosophy in Mid-Century”, George Allen & Unwin.

Salmon, Wesley (1989). *Four Decades of Scientific Explanation*. University of Minnesota Press.

Scott, Dana (1971). “On engendering an illusion of understanding”. *Journal of Philosophy*, 68:787–807.

Sellars, Wilfrid (1953). “Inference and Meaning”. *Mind*, 62:313–338.

—— (1954). “Some Reflections on Language Games”. *Philosophy of Science*, 21:204–228. Also in Sellars (1963).

—— (1963). *Science, Perception and Reality*. Routledge & Kegan Paul.

Shapiro, Stewart (1991). *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford University Press.

Shoesmith, D. J. and Smiley, Timothy (1978). *Multiple Conclusion Logic*. Cambridge University Press.

Shogenji, Tomoji (1999). “Is Coherence Truth-Conducive?” *Analysis*, 59:338–345.

Skolem, Thoralf (1919). “Untersuchungen über die Axiome des Klassenkalkuls und über Produktations- und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen”. *Skrifter utgit av Videnskapsselskapet i Kristiania, I. Matematisk-naturvidenskabelig klasse*, 3.

Sneed, Joseph (1971). *The Logical Structure of Mathematical Physics.* Reidel.

Sorkin, Raphael (1989). “Does a Discrete Order underly Spacetime and its Metric?” In Coley, A., Cooperstock, F., and Tupper, B. (eds.) “Proceedings of the Third Canadian Conference on General Relativity and Relativistic Astrophysics”, World Scientific.

Sosa, Ernest (2007). “Experimental philosophy and philosophical intuition”. *Philosophical Studies*, 132:99–107.

Stegmüller, Wolfgang (1979). “The Structuralist View: Survey, Recent Developments and Answers to Some Criticisms”. In Niiniluoto, Ikka and Tuomela, Raimo (eds.) “The Logic and Epistemology of Scientific Change”, North Holland.

Strawson, Peter F. (1959). *Individuals: an Essay in Descriptive Metaphysics.* Methuen.

Suppes, Patrick (1959). “Measurement, empirical meaningfulness, and three-valued logic”. In Churchman, P. D. and Ratoosh, P. (eds.) “Measurement: definitions and theories”, Wiley.

Tarski, Alfred (1936). “Über den Begriff der logischen Folgerung”. *Actes du Congrès International de Philosophie Scientifique*, 7:1–11. Also in Tarski (1956).

—— (1956). *Logic, Semantics, Metamathematics.* Oxford University Press.

—— (1986). “What are Logical Notions?” *History and Philosophy of Logic*, 7:143–154.

van Fraassen, Bas C. (1980). *The Scientific Image.* Oxford University Press.

—— (1989). *Laws and Symmetry.* Oxford University Press.

—— (1991). *Quantum Mechanics: An Empiricist View.* Oxford University Press.

—— (2002). *The Empirical Stance.* Yale University Press.

von Mises, Richard (1981). *Probability, Statistics and Truth.* Dover, 2nd ed.

von Neumann, John (1955). *Mathematical Foundations of Quantum Mechanics.* Princeton University Press.

Weinberg, J., Nichols, S., and Stich, S. (2001). “Normativity and Epistemic Intuitions”. *Philosophical Topics*, 29:429–459.

Williams, D. C. (2004). “The Elements of Being”. *Review of Metaphysics*, 7:3–18, 171–192.

Wittgenstein, Ludwig (1922). *Tractatus Logico-Philosophicus.* Routledge & Kegan Paul Ltd.

—— (1953). *Philosophical Investigations.* Blackwell Publishers.

Wójcicki, Ryszard (1988). *Theory of Logical Calculi.* Kluwer Academic Publishers.

Index