MODELS AND MAPS: AN ESSAY ON EPISTEMIC REPRESENTATION

Gabriele Contessa

Please cite this work as follows:
Contessa, Gabriele (2013). *Models and Maps: An Essay on Epistemic Representation*, unpublished manuscript, Carleton University, Ottawa, ON.

For Stella and Mia (who are very disappointed that this book does not contain any pictures)

Foreword

This book has had a long and tormented history. A version that is not too different from the one you are currently reading was ready in the early summer of 2009. All I wanted to add at that point was a few more scientific case studies to illustrate the various aspects of my account. Then, a health problem in my family caused a significant reduction in the time I could devote to my research and this book ended up on the back burner for a few years, as I tried to work on new and more stimulating projects in the time I had. When I finally was able to came back to the book, I realized that I would have to re-write large swaths of the book to incorporate the huge literature on the topic that has accumulated since 2009 and I decided that perhaps the best thing would be to make this version of the manuscript available in its current form. I believe that the book makes some contributions that are still relevant today and I hope the readers will think so too. Please feel free to circulate and cite this version of the manuscript with proper attribution. However, this manuscript cannot be used for any commercial purposes. If you would like to cite the manuscript, please do so as follows:

Contessa, Gabriele (2013). *Models and Maps: An Essay on Epistemic Representation*, unpublished manuscript, Carleton University, Ottawa, ON.

Preface

In *Le Cittá Invisibili*, Italo Calvino describes the city of Eudoxia. Eudoxia is a chaotic maze of ramshackle buildings and narrow, winding alleyways. Somewhere in the city, however, lies a carpet whose colourful threads interweave so as to form a symmetric geometrical pattern. Despite the fact that the carpet and the city do not resemble each other much, the narrator assures us that, if you were to carefully contemplate the carpet, you would be able to discern in it "the true form" of the city.

The representations of the world that science provides us with are not unlike the carpet in Eudoxia. The world is a complicated and multifarious place that does not contain any of the point masses and frictionless planes that inhabit our models of it, and yet those models often prove themselves surprisingly successful at guiding our actions in the world. Moreover, although the world does not appear to us anything like how our most fundamental theories represent it to be, many tend to think that those theories provide us with a truer image of the world than our naïve conception of it.

This book is about the relation between those (and other) representations and the portions or aspects of the world they represent. Initially, it may be tempting to think that representation is a binary relation that holds between two objects in virtue of the way those objects are and independently of us, like the mysterious, almost magical relation that holds between the carpet and Eudoxia in Calvino's story. However, few today would take this magical view of representation seriously. As Hillary Putnam (1981: 1) famously argued, if an ant is crawling on the sand on a desert island and, in so doing, it leaves tracks in the sand that, by pure coincidence, would look to us like a caricature of Winston Churchill, the ant does not seem to have thereby produced a representation of Churchill. Representing is not something things do on their own; it is something *we* do with them. This book tries to explain how we do that—how we turn something into a representation of something else, and what it is that makes that something a more or less accurate representation of that something else.

This book is the culmination of a long research project, which started when I was in graduate school. It is hard to overestimate my intellectual debt to my then-supervisor, Nancy Cartwright. Ever since coming across her book *How the Laws of Physics Lie* as an undergraduate student in Rome, I became fascinated with Nancy's view of the relationship between our theories and the world. I couldn't believe my luck when, a few years later, Nancy agreed to supervise my doctoral thesis. Nancy turned out to be the ideal supervisor, somehow managing to be both my biggest supporter and my staunchest critic. Although I am responsible for all of the shortcomings of this book, if there are any good ideas in it, I owe them mostly to Nancy.

Beside Nancy, I also owe substantial debts to many others, including Otávio Bueno, Anjan Chakravartty, Mauro Dorato, Steven French, Roman Frigg, Ken Gemes, Ron Giere, Robin Hendry, Elaine Landry, Eleonora Montuschi, Mary Morgan, Matteo Morganti, Margaret Morrison, Martin Thompson-Jones, Chris Pincock, Mauricio Suárez, Bas van Fraassen, and John Worrall, as well as Daniela Bailer-Jones and R.I.G. Hughes, who, unfortunately, are no longer with us.

Portions of the Introduction, portions of Chapters 1 and 2, and the entirety of Chapter 3 are adapted from, respectively, (Contessa 2011), (Contessa 2007), and (Contessa 2010). I am grateful to the editors and publishers for permission to use those materials.

I am very grateful to the Social Sciences and Humanities Research Council of Canada for supporting my research through a Standard Research Grant. I would also like to thank Eli Shupe for her invaluable assistance in the preparation of this book and for many helpful comments.

Gabriele Contessa

Toronto
July 2013

Introduction

My daughters would love to go tobogganing down the tobogganing hill by themselves, but they are young and I am an apprehensive parent, so, before letting them do so, I want to make sure that they wouldn't go too fast. But how can I find out how fast they would go? At first, it might be tempting to think that I could approach the question in a purely empirical manner. Simply sending my daughters down the hill on their toboggan a number of times and trying to measure their velocity would probably be the most accurate empirical method to find an answer to my question but, of course, that would miss the whole point of the exercise. If, on the other hand, I were to ride the toboggan myself or let it slide empty down the hill, the question would be whether the toboggan would go as fast if my daughters were to ride it instead. So it would seem that, short of (self-defeatingly) sending my daughters down the hill to find out whether it would be safe to do so, I cannot find an answer to my question in a purely empirical manner—I have to appeal (to some extent) to theoretical considerations. In particular, in this case, I would have to turn to our best theory of the motion of mid-sized physical objects moving at relatively low velocities—i.e. classical mechanics. But what does classical mechanics tell us about my daughters and their possible journey down the hill?

One way to try to answer the question theoretically would be to try to apply the theory to the situation directly. Since my daughters and their toboggan are currently at rest at the top of the hill, classical mechanics tells us that their velocity would be determined by the forces they would be subjected to after the toboggan is released at the top of the hill. The problem is that, during their downhill journey, my daughters and their toboggan would be subjected to an extraordinarily large number of forces—from the gravitational pull of distant stars to the weight of the snowflake sitting on the tip of one of my younger daughter's eyelashes—so that any attempt to apply the theory directly to the real-world situation in all its complexity would seem doomed to failure.

Since applying the theory to the real-world situation *directly* seems to be practically impossible, I might try to apply it *indirectly* instead—that is, I might apply the theory to a simplified model of the real-world situation. Luckily, in this case, I might even be able to use a simple stock model from classical mechanics—the inclined plane model. In the inclined plane model, a box sits at top of an inclined, frictionless plane, where its potential energy, U_i, is equal to mgh (where m is the mass of the box, g is its gravitational acceleration, and h is the height of the plane) and its kinetic energy, KE_i, is zero. When the box is released, it will slide down the plane and, at the bottom of the slope, all of its initial potential energy will have turned into kinetic energy ($E_f = KE_f + U_f = \frac{1}{2}mv_f^2 + 0 = 0 + mgh$). The final velocity of the box, v_f, will thus be $(2gh)^{1/2}$ and depends only on its initial height and on the strength of the gravitational pull. If, for example, we set h to 10 m and g to 9.8 m/s^2 the final velocity of the box will be 14 m/s or about 50 km/h. But what does this tell me about how fast my daughters would go on their toboggan? And why should I believe what the model tells me anyway?

The practice of using models to predict, explain, investigate, or understand the behaviour of aspects or portions of the real world is ubiquitous. Models are widely used by natural and social scientists, engineers, policy-makers, as well as ordinary people (as my example illustrates). However, it is only relatively recently that philosophers of science have started to take models seriously. The received view, which we might call '*the descriptive view*', was that scientific theories were sets of sentences or propositions that related to the world by describing it truly or falsely (or, at least, by entailing other sentences or propositions that describe aspects or portions of it truly or falsely). According to the descriptive view, the principles of the theory and deductive reasoning do all the real work; scientific models play at most an ancillary, heuristic role.

However, as my example shows (and as the critics of the descriptive view have convincingly argued), even the simplest real-world systems are way too messy and complicated for us to be able to apply the abstract concepts of our theories (and the mathematical apparatus that often comes with them) *directly* to them. Usually, we can only apply these theoretical concepts to simplified and idealized models of those systems.[1] In light of these and other considerations, most philosophers of science have now come to abandon what I have called the descriptive view in favour of what I call 'the *representational view*'.

Supporters of the representational view come in (at least) two varieties. Those who adopt what we could call 'the *model view*' (or, as it is often misleadingly called, the "semantic view") deny that scientific theories are collections of sentences or propositions and prefer to think of them as collections of models.[2] Those who opt for what we could call 'the *hybrid view*', on the other hand, are still happy to think of theories as collections of sentences or propositions but deny that models play an ancillary role and maintain, instead, that models play crucial mediating role between our theories and the world.[3] Despite their disagreements, all supporters of the representational view seem to agree on two crucial points. The first is that it is scientific models (not sentences or propositions) that relate directly to the world. The second is that, unlike sentences or propositions, and like tables, apples and chairs, models are not truth-apt—i.e. they are not capable of being true or false. So, whereas according to the descriptive view, scientific theories related to the world just like (declarative) sentences or propositions do (i.e. by being true or false of it), according to the representational view they relate to the world more like maps and pictures do (i.e. by representing aspects or portions of it).

As models gained centre stage in the philosophy of science, a new picture of science emerged (or, perhaps, an old one re-emerged)[4], one according to which science provides us with (more or less faithful) representations of the world as opposed to (true or false) descriptions of it.[5] So, to a first approximation, the representational view holds that theories do not relate to the world directly by describing it but indirectly by describing families of models, which, in turn, are used to represent (parts or portions of) the world.

Now, the problem is that, while a philosopher of science who adopts the descriptive view could reasonably hope to outsource the task of providing an account of how theories relate to the world to philosophers working in other areas by relying on their accounts of semantic notions such as reference, meaning, truth, etc., it is not as easy to do so when it comes to the notion of (non-linguistic) representation used by the representational view. While supporters of the representational view may find very helpful insights and suggestions in the burgeoning literature on mental or pictorial representation, it is not clear whether they can find exactly what they need—i.e. an account of what I call 'epistemic representation'. This book develops and defends one such account, which can be used by supporters of the representational view as well as by other philosophers who, in other areas or contexts, may find the notion of epistemic representation useful.

This book has three parts. In Part I, I distinguish three relevant uses of the term 'representation' and explain in what sense models represent real-world systems. In Chapters 1

---

[1] This point has been pressed most forcefully by Nancy Cartwright (see, in particular, Cartwright 1983 and Cartwright 1999).

[2] The so-called semantic view originated with the work of Patrick Suppes in the 1960s (see, e.g., (Suppes 1960) but also (Suppes 2002)) and can be safely considered the new received view of theories, counting some of the most prominent philosophers of science among its supporters (see, e.g., (van Fraassen 1980), (Giere 1988), (Suppe 1989), (da Costa and French 1990)). How exactly the so-called semantic view of theories relates to the view that theories are collections of models is an exegetical question that is beyond the scope of this book.

[3] This view, sometimes referred to as the models-as-mediators view, is developed and defended, for example, by many of the contributors to (Morgan and Morrison 1999).

[4] See, e.g., (van Fraassen 2008, Ch. 8).

[5] How solid the contrast between representing and describing is obviously depends on one's views on language and truth.

and 2, I introduce the notions of denotation, epistemic representation, and faithful epistemic representation as well as a number of related notions and I discuss some ordinary examples of epistemic representation and faithful epistemic representation. In Chapter 3, I offer an account of what scientific models are.

Part II is devoted to the question of what makes a vehicle an epistemic representation of a certain target. In Chapter 4, I discuss two accounts of epistemic representation—i.e. the denotational account and the inferential account—and argue that neither provides us with a satisfactory answer to that question. In Chapter 5, I develop and defend a third account of epistemic representation—the interpretational account. According to the interpretational account, a vehicle is an epistemic representation of a certain target (for a certain user) only if the user adopts an interpretation of the vehicle in terms of the target.

Part III focuses on the question of what makes an epistemic representation of a certain target a more or less faithful one. In Chapters 6 and 7, I examine, respectively, the similarity and the structural accounts of faithful epistemic representation and I find them both wanting. In Chapter 8, I develop what I call 'the structural similarity account', which, I argue, combines the strengths of the similarity account with those of the structural account without sharing their respective weaknesses.

Part I: Untangling Representation

1 Epistemic Representation

1.1 Denotation and Epistemic Representation

It's your first time in London and you just got off the train at Liverpool Street railway station. You know your hotel is near Holborn station but you don't know how to get there yet. You follow the signs with the London Underground logo to the entrance of Liverpool Street subway station and, just as you are about to ask someone for directions, you notice a pile of folded pieces of glossy paper with coloured lines, small circles, and names printed on them. You pick one of these papers up and, after examining it for a couple of minutes, you conclude that to get to Holborn station you need to take a westbound Central Line train and get off at the fourth stop. But how did you acquire this information? How could examining a small piece of paper covered in symbols provide you with such detailed piece of knowledge?

The short answer is, of course, that the piece of paper is a map of the London Underground network. If the piece of paper were instead a map of Toronto's subway system, a reproduction of one of Francis Bacon's self-portraits, or a £10 banknote, it wouldn't have been of any use in finding your way around the London Underground network. But what is it that makes a piece of paper *into* a map of the London Underground network (as opposed to, say, a portrait of Francis Bacon or a banknote)? One tempting answer is that the piece of paper is a map of the London Underground network because the marks it bears constitute a representation of the London Underground network (rather than of Toronto's subway network or of Francis Bacon). A problem with this answer, however, is that it tries to explain the obscure with the more obscure, as we are still left asking what it is for one thing to count as a representation of another.

So, what does representation mean (in this context)? In many ways, 'representation' is an ambiguous term, and before we can make any progress in understanding what it takes for something to represent something else, we must be clear about which sense (or senses) of 'representation' are relevant here. For our purposes, it is important to initially distinguish between two different senses of 'representation'. In the first sense, both the London Underground logo and the London Underground map can be taken to "represent" the London Underground network. In the terminology I use here, we can say that they both *denote* the network. The London Underground map, however, does more than merely denoting the London Underground network. It also represents the network in a second, stronger sense—viz. in the sense that it is (in my terminology) an *epistemic representation* of the network.

It is in virtue of the fact that the map of the London Underground is an epistemic representation of the London Underground network that (to use the terminology introduced by Chris Swoyer (1991)) one can perform *surrogative inferences* from the map to the network, where:

(1) if v and t are two distinct objects, a *surrogative inference* from v to t is an inference whose only premise is a proposition about v and whose conclusion is a proposition about t.

In our example, the map and the network are clearly two distinct objects. One is a piece of glossy paper, while the other is an intricate network of trains, tunnels, railways and platforms.[6] Yet their users frequently perform surrogative inferences from the one to the other. For example, from 'a red line connects the circle labelled 'Holborn' to the circle labelled 'Liverpool Street'' (which

[6] Or so I assume here. For some issues in the metaphysics of subway systems see Myrtle Willoughby's seminal (2012).

expresses a proposition about the map) a user can validly infer 'Central Line trains operate between Holborn and Liverpool Street' (which expresses a proposition about the network). The same does not apply to the London Underground logo. Users of the London Underground network do not usually use the London Underground logo to perform surrogative inferences from it to the network and there seems to be no obvious way to do so. The logo may denote the network, but the logo itself is not an epistemic representation of the network.[7]

1.2 Epistemic Representation

The notion of epistemic representation can be defined as follows:

> **(2)** *v* is an *epistemic representation* of *t* (for *u*), if and only if:
>
> [2.1] *u* is able to perform (valid though not necessarily sound) surrogative inferences from *v* to *t*.

In what follows, I call *v*, *t*, and *u* respectively 'the *vehicle*', 'the *target*', and 'the *user*' (of the epistemic representation), and I call condition [2.1] (*valid*) *surrogative reasoning*. So, (2) thus tells us that (valid) surrogative reasoning is a necessary and sufficient condition for epistemic representation. But what exactly does this mean?

First, let me explain 'for *u*'. According to (2), an object, *in and of itself*, is not an epistemic representation of anything—it is an epistemic representation of something else only *for someone*. Epistemic representation is therefore not a dyadic relation between two objects but (at least) a triadic relation between a vehicle, a target and a (set of) user(s).[8] For the sake of simplicity, in what follows I often omit mention of the users of an epistemic representation unless it is required by the context. However, this does not mean that a vehicle can be an epistemic representation of a target for no one in particular or in its own right. It should always be understood that a vehicle is an epistemic representation of a certain target only if there is some user for whom it is an epistemic representation of that target.[9]

Second, let me clarify 'is able to'. The fact that someone actually performs a specific surrogative inference from a certain object to another is perhaps the clearest "symptom" of the fact that, for that person, the former object is an epistemic representation of the latter. However, there may be "asymptomatic" cases of epistemic representation. That is, a user does not need to perform any *actual* piece of surrogative reasoning in order for the vehicle to be an epistemic representation of the target for her. For example, even if a user has never performed and will never perform any actual inference from the London Underground map to the London

[7] I say more about denotation in Section 1.4 below.

[8] That what I call epistemic representation is (at least) a triadic relation seems to be one of the few issues on which most contributors to the literature on scientific representation agree (see, e.g. Suárez 2002 and 2003, Frigg 2002, Giere 2004). Suárez (2002) however does not seem to think that this is the case. He thinks that the supporters of the similarity and structural accounts of epistemic representation are trying to "naturalize" epistemic representation in the sense that they are trying to reduce representation to a dyadic relation between the vehicle and the target. Whereas, in the past, some contributors to the debate may have given the impression that they conceived of representation as a dyadic relation, I do not think that this is their considered view. Giere (2004), for example, dispels any doubt by declaring: 'The focus on language as an object in itself carries with it the assumption that our focus should be on *representation*, understood as a two-place relationship between linguistic entities and the world. Shifting the focus to scientific practice suggests that we should begin with the activity of *representing,* which, if thought of as a relationship at all, should have several more places. One place, of course, goes to the agents, the scientists who do the representing" (Giere 2004, p.743).

[9] This is particularly important when the epistemic representation has a large set of users (such as the London Underground map). In those cases, we usually tend to disregard the fact that the vehicle is an (epistemic) representation *for those users*, not *in its own right*. The fact that a vehicle is an epistemic representation for many people or even for everyone does not imply that it is an epistemic representation in and of itself.

Underground network, the map may still be an epistemic representation of the network for her insofar as she would be able to perform surrogative inferences from the map to the network.

Third, let me turn to the parenthetic qualification 'valid though not necessarily sound'. First of all, a surrogative inference is *sound* if and only if it is valid and its conclusion is true (or, at least, approximately true[10]). But when is a surrogative inference valid? Since a precise answer to this question would require the use of concepts that I will not be in a position to introduce until Chapter 5, for the moment we will have to rely on an informal (and somewhat vague) characterization of 'valid surrogative inference'. The intuitive idea behind the notion of a *valid* surrogative inference is that a surrogative inference is valid only if it is in accordance with a systematic set of rules. For example, according to the set of rules associated with what I call 'the standard interpretation of the London Underground map', from the fact that the circles marked 'Holborn' and 'Bethnal Green' are connected by a red line, it is valid to infer that Central Line trains operate between Holborn and Bethnal Green. According to the same set of rules, however, from the fact that the circles marked 'Holborn' and 'Bethnal Green' are three inches apart it is not valid to infer that the distance between Holborn and Bethnal Green station is three miles (or anything else about the London Underground network, for that matter). The standard set of rules associated with the map licenses the former inference but not the latter.[11]

The qualification that the set of rules be systematic is meant to ensure that the result of applying a given rule does not depend on who applies the rule or the circumstances in which the rule is applied but only on the way the vehicle is and what the rule states. A systematic set of rules, for example, cannot include rules such as: 'If the circles marked 'Holborn' and 'Bethnal Green' are connected by a red line, conclude the first thing that goes through your mind' because this kind of rule might (and presumably would) give different results when different users apply it or even when the same user applies it on different occasions.

The valid surrogative inferences from the vehicle to the target do not need to be sound in order for the vehicle to be an epistemic representation of the target—i.e. the conclusions that that user validly draws (or would draw) from the vehicle to the target do not need to be true (or even just approximately true) in order for the vehicle to be an epistemic representation of the target for that user. Insofar as a user is able to perform (valid) inferences from a certain vehicle to a certain target, that vehicle is an epistemic representation of that target for that user, independently of whether or not the conclusions drawn by the user are (approximately) true of the target.

Unless otherwise specified, whenever talking about a surrogative inference from a certain vehicle to a certain target in what follows, I always assume that the inference is valid—i.e. that there is a set of rules standardly associated with the vehicle and that the inference in question conforms to those rules. Whether there are any such rules, what they are, and how they arise are all crucial questions, but I save these questions for Chapter 5, where I outline my account of epistemic representation and introduce the notion of interpretation on which it relies.

Finally, the fact that epistemic representation is defined in terms of surrogative reasoning may misleadingly suggest that *v* is an epistemic representation of *t* for *u because u* is able to perform surrogative inferences from *v* to *t*. However, just the reverse is true—as I argue, it is because *u* can interpret *v* in terms of *t*, and in doing so, make *v* an epistemic representation of *t*, that *u* is able to perform surrogative inferences from *v* to *t*.

---

10 I say more about what I mean by 'approximately true' in §2.2 below.

11 Of course, it is possible to come up with a non-standard set of rules, according to which the latter inference is valid and the former is not (although no set of rules can make the latter inference sound, as the two stations are not three miles apart).

1.3 Representation and Epistemic Representation

The notion of epistemic representation introduced in the previous section is primarily a technical one. However, I intend it to be an explication of what I take to be one of the ordinary meanings of 'representation.' By 'epistemic representation' I mean a representation that is used (or can be used) for epistemic purposes—i.e. a representation that can be used to try to learn something about its target. Although the primary purpose of some representations may not be epistemic (the main purpose of, say, Pablo Picasso's Portrait of Dora Maar, for example, seems to be aesthetic and the main purpose of a toy car is presumably ludic), I think it is a positive result that, according to (2), the vast majority of what we ordinarily consider representations (including the two I just mentioned) turn out to be epistemic representations of their targets. Portraits, photographs, maps, graphs, and other representational devices typically do allow their users to perform (valid) inferences about their targets and, as such, according to (2), they are epistemic representations of their targets (for us). For example, according to (2), if we are able to draw (valid) inferences from a portrait to its subject (as we are usually able to do in the case of traditional portraits), then the portrait is an epistemic representation of its subject (for us).

However, there is also a sense in which the notion of epistemic representation seems to be somewhat weaker than the ordinary notion of representation. This, I think, is due to an ambiguity in the ordinary notion of representation. 'Represent' is sometimes used as a success verb and sometimes not. This is probably why we usually tend to conflate two distinct issues: the issue of whether a certain vehicle is an (epistemic) representation of a certain target, and the issue of the degree to which that vehicle is a faithful (epistemic) representation of that target.

Let me illustrate this point with an example. Suppose that Jack has never met Jill. Nevertheless, Jack has seen a portrait of Jill, from which he was able to infer a description of Jill's appearance. In order to do so, Jack had to perform a number of inferences from the portrait to the subject of the portrait. But why should Jack assume that the description he has inferred from the portrait is an accurate description of Jill? For all we know, the portrait may not be an accurate representation of Jill (the painter might have never met Jill or might be mistaken about her appearance). The portrait may even be a deliberately misleading representation of Jill's appearance (for whatever reason, the painter might have intended to mislead the viewers about Jill's appearance). Alternatively, the portrait may be a "figurative" rather than "literal" representation of Jill (the painter, for example, might have represented Jill as she did to allude to a trait of Jill's personality rather than to reflect Jill's visual appearance).

Whatever the case may be, the painting is an epistemic representation of Jill (for Jack)—what is to be determined is the extent to which it is a *faithful* epistemic representation of Jill. Nevertheless, the painting cannot be an unfaithful epistemic representation of Jill unless it is an epistemic representation of Jill *in the first place*. To put it in a slogan, 'there is no misrepresentation without representation.' Unless Jack is ready to infer from the portrait a description of Jill's appearance, the portrait cannot mislead Jack about Jill's appearance and, insofar as Jack is ready to infer from the portrait a description of Jill, the portrait is, according to (2), an epistemic representation of Jill for him. As this example illustrates, in addition to the notion of epistemic representation, we need a notion of a *faithful* epistemic representation. In the next chapter, I introduce this notion more precisely.

1.4 Denotation

In the last century or so, the notion of denotation (or reference) has played a central role in analytic philosophy, so much so that surveying all the major philosophical accounts of denotation would lead us far astray from the focus of this book. Nevertheless, since denotation plays an important role in what follows, a few remarks are in order.

The notion of denotation is a central notion in the philosophy of language, where it is (minimally) construed as a relation that holds between certain kinds of linguistic expressions—I call them '*denoting expressions*'—and objects (in the broadest sense of the word).[12] To consider a typical example, the proper name 'Napoleon' denotes (or refers to) Napoleon.

In the analytic philosophy of art, however, the notion of denotation has a broader use. On this broader use, the first *relatum* of the denotation relation does not need to be a linguistic expression. Any two objects (in the broadest sense of the word) can be the *relata* of the denotation relation. Of course, these two uses are compatible insofar as we assume that linguistic expressions are only some of the objects that can be used to denote other objects. In this book, I adopt this broader use of 'denotation' according to which anything can be used to denote anything else.

But the question remains: what exactly *is* denotation? What is required for something to denote something else? Here I do not try to answer these questions. Instead, I treat denotation as a primitive notion. This is not because I doubt that philosophers or scientists can provide a deeper account of denotation, but only because the details of any such account are unlikely to affect the view that I develop and defend in this book.

Here, I assume that a vehicle denotes a certain target for a user if and only if the user takes the vehicle to denote (stand for, refer to) the target (I treat these three expressions as synonymous). I take it that it is uncontroversial that, as human beings, we happen to be able to use some objects to denote other objects. In the absence of this ability, it would be difficult to explain even our most basic linguistic and symbolic practices. If there is any disagreement surrounding denotation, the disagreement is about *how* some things get to denote other things for us, and not about *whether* some things do denote other things (for us). Nothing in what I say here, however, depends on *how* some things denote other things for us. It only depends on the fact that some things do denote other things (for us), which, as far as I can see, is a widely shared assumption.

1.5 Representational Content and Representational Scope

I will now introduce a few more notions that will be useful in what follows. The first is the notion of 'representational content'.

> **(3)** The *representational content* of an epistemic representation is the set of propositions about its target, *t*, that it is valid to infer from its vehicle, *v*.

So, for example, since, according to the rules standardly associated with the London Underground map, it is valid to infer from it that Holborn station and Liverpool Street station are connected by Central Line trains, the proposition expressed by 'Holborn station and Liverpool Street station are connected by Central Line trains' is part of the representational content of the map (given the standard set of rules ordinarily associated with it). On the other hand, since it is not valid to infer from the map anything about the distance between those two stations, 'Holborn station is three miles from Liverpool Street station' is *not* part of the representational content of the map (given the standard set of rules ordinarily associated with it).

It is important to emphasize that an object in and of itself does not have a representational content; an object has a representational content only insofar as it is used as the vehicle of an epistemic representation of some target by some users. Since we often use the same term (e.g. 'map') to refer both to epistemic representations and to the objects that serve as their vehicles ('The map has a few rips on it'), it is easy to get confused about this. However, in any ambiguous

[12] I say 'minimally' because I take it that most philosophers of language do not conceive of the denotation relation simply as a dyadic relation.

cases, context should help the reader determine whether terms such as 'map' refer to an epistemic representation or to its vehicle. For example, whenever I talk about the representational content of a map, I intend 'map' to refer to the epistemic representation, not to the material object that serves as its vehicle.

Let me now introduce the notion of incompatible representations.

(4) *v* and *v** offer *incompatible epistemic representations* of (a certain aspect of) the target *t* if and only if:

[4.1] The representational content of *v* includes at least one proposition that is the negation of a proposition included in the representational content of *v**.

So, for example, an old 1930s London Underground map and a new map offer incompatible representations of the London Underground network because, among other things, from the new map it is valid to infer that a direct train service operates between Euston station and Oxford Street station, while, from the old one, it is valid to infer the opposite. Two different copies of the new map, on the other hand, will not offer incompatible representations of the network, as everything that can be inferred from one of them can also be inferred from the other.

I now define the notion of 'representational scope'.

(5) *v* and *v** *have the same scope* if and only if:

[5.1] *v* and *r** are both epistemic representations of *t* for *u*, and

[5.2] for every proposition *p* about *t*, *p* is part of the representational content of *v* if and only if *p* or its negation is part of the representational content of *v**, and vice versa.

(6) *v* and *v** *have different scopes* if and only if:

[6.1] *v* and *v** are both epistemic representations of *t* for *u*, and

[6.2] for some proposition *p* about *t*, *p* is part of the representational content of *v* but neither it nor its negation is part of the representational content of *v** or vice versa.

(7) *v has (strictly) broader scope than v** if and only if:

[7.1] *v* and *v** are both epistemic representations of *t* for *u*,

[7.2] for every proposition *p* about *t*, if *p* is part of the representational content of *v**, *p* or its negation is part of the representational content of *v*, *and*

[7.3] for some proposition *p*, *p* is part of the representational content of *v* but neither *p* nor its negation is part of the representational content of *v**.

(8) *v* has (strictly) narrower scope than v* if and only if:

[8.1] *v* has broader scope than *v**.

(9) *v* and *v** have *overlapping scopes* if and only if:

[9.1] *v* and *v** are both epistemic representations of *t* for *u*,

[9.2] for some proposition *p* about *t*, *p* or its negation are part of the representational content of both *v* and *v**.

(10) *v* and *v** have *partially overlapping scopes* if and only if:

[10.1] for some proposition *p* about *t*, *p* or its negation are part of the representational content of *v* and *v**, and

[10.2] neither *v* nor *v** has broader scope than the other.

For example, even if, as we have seen, an old 1930s London Underground map and a new map offer incompatible representations of the London Underground network, they still have the same scope insofar as, for every proposition about the network that can be validly inferred from the one, it or its negation can be inferred from the other. In other words, any question about the network that can be answered by using one of the two maps (e.g. 'Does a direct train service operate between Euston and Oxford Circus?', 'Is there a station named 'Holborn'?') can also be answered by using the other, even if the two maps provide us with different answers to some of those questions.

London Underground maps from the 1910s, on the other hand, had different scope from more recent ones. Their scope included information about the geographic locations of stations, which falls outside of the scope of standard London Underground maps produced after the 1930s. So, the scope of a 1910s map is both different and broader than the scope of a new map, for there are propositions that are validly inferred from the 1910s map that are such that neither they nor their negations can be validly inferred from new map.

Although the 1910s map has broader scope than the new map, it is not necessarily the case that one of two epistemic representations of the same target will always have broader scope than the other, even in cases where their scopes overlap. For example, although their scopes partially overlap, a postcard with a view of Rome and a map of Rome have largely different scopes. Most of what can be validly inferred about Rome from the postcard cannot be inferred from the map and vice versa. From the postcard we can infer a great deal about what Rome looks like but very little about its topography, while the reverse is true of the map. In other words, the relation 'having the same or broader scope than' is a partial order—i.e. it is possible for neither of two epistemic representations of the same target to have the same or broader scope than the other.

1.6 Conclusion

So far, I have distinguished between three senses of the ordinary notion of 'representation', which I have called denotation, epistemic representation, and faithful epistemic representation. In this book, I take the notion of denotation to be primitive and focus instead on the notions of epistemic representation and faithful epistemic representation. In this chapter, I have defined the notion of epistemic representation and related notions. In the next chapter, I turn to the notion of faithful epistemic representation.

2 Faithfulness

2.1 Faithful Epistemic Representation

As we have seen in Chapter 1, both an old 1930s London Underground map and a new map are epistemic representations of the London Underground network (for us) because from either map we can perform valid surrogative inferences to the network. However, the two maps offer largely incompatible representations of the London Underground network (as it is today) because, from one map, it is valid to infer propositions that are the negations of propositions that can be validly inferred from the other. For example, from the old map, one would validly infer that there is no direct train connection between Euston and Oxford Circus, while from the new map one would infer that Victoria Line trains operate between those two stations. Because these are incompatible states of affairs, both maps cannot be right. As it turns out, all valid surrogative inferences from the new map to today's network are sound—i.e. have true conclusions[13]—but only some of the inferences from the old map to today's network are sound.[14] In this sense, we can say that the old map *misrepresents* some aspects of today's network. But how can the old map both *represent* and *mis*represent today's network?

As I argued in Section 1.3 above, in order to misrepresent the network, a map must represent it in the first place. To avoid confusing the different senses of 'representation', it is thus crucial to carefully distinguish two separate issues. The first is the issue of whether a certain vehicle is an epistemic representation of a certain target (for a certain user). The second is the issue of whether (and to what degree) that vehicle is a *faithful* epistemic representation of that target.[15]

To this end, alongside the definition of 'epistemic representation', we need to introduce the following notions:

(11) *v* is a *completely faithful epistemic representation* of *t* (for a user *u*) if and only if:

[11.1] *v* is an epistemic representation of *t* (for *u*), and

[11.2] all valid surrogative inferences from *v* to *t* are sound (i.e. have true (enough) conclusions).

(12) *v* is a *partially faithful epistemic representation* of *t* (for a user *u*) if and only if:

[12.1] *v* is an epistemic representation of *t* (for *u*), and

[12.2] some (but not all) valid inferences from *v* to *t* are sound.

---

[13] Or so I assume here.

[14] Presumably the reverse would be true of the 1930's London Underground network.

[15] The distinction between epistemic representation and faithful epistemic representation is inspired by and closely related to the distinction Mauricio Suárez (2004) has drawn between representation and 'accurate, true and complete representation' (Suárez 2004, p.767). The importance of such distinctions when doing philosophical work on representation cannot be overestimated. In fact, I think no real progress can be made in our understanding of the notion of representation until these two notions are carefully distinguished. Unfortunately, even Suárez seems to have failed to fully appreciate the importance of those distinctions (but more on this later). I am heavily indebted to Nancy Cartwright for helping me to fully appreciate their importance.

According to (11) and (12), the new map is thus a completely faithful representation of today's network, while the old map is only a partially faithful one. Note that in order to be a *completely faithful* epistemic representation of its target, an epistemic representation does not need to be a *complete* epistemic representation of its target. The new London Underground map, for example, is a completely faithful epistemic representation of today's network because, from it, one can validly draw only true conclusions about today's network. However, the map is not a complete epistemic representation of the network, as there are innumerably many aspects of the London Underground network that fall outside of its representational scope (e.g. the internal structure of the stations or the distances between them).

Note also that the same vehicle can be a faithful epistemic representation of some aspects of a certain target while misrepresenting others. This seems to be the case with the old London Underground map, which is a partially faithful epistemic representation of the London Underground network. From the old map, we can validly draw many true conclusions about today's network as well as many false ones.

2.2 Closeness to the Truth

In Section 1.2, I defined a sound surrogative inference as one that is valid and whose conclusion is true (or approximately true). Although I am unable to offer anything even remotely close to a full-fledged account of approximate truth, in this section I try to clarify what I mean by 'approximately true' and introduce the notion of closeness to truth that plays an important role in what follows.

Intuitively, a proposition is approximately true if it is "close enough" to the truth, where in most cases what counts as "close enough" depends on a number of contextual factors. As I use it here, the expression 'approximately true' applies only to propositions concerning some quantity—e.g. the height of a building, the mass of a stone, or the age of a person. Suppose, for example, that I am describing a room to a friend and I tell her that it is 6 meters wide. If it turns out that the room is actually only 5.9 meters wide, the proposition expressed by my sentence would be strictly speaking false. However, in most ordinary contexts, a difference of 10 centimetres is negligible. Although my description of the room is strictly speaking false, it seems that in most ordinary contexts it would be unreasonable for my friend to protest that my description was false if she were to find out that the room is actually 5.9 meters wide. *In those contexts*, the sentence 'The room is 6 meters wide' expresses an approximately true proposition. In other contexts, however, a difference of 10 centimetres might be crucial. If my friend wanted to know whether a bookshelf that is exactly 6 meters wide would fit in that room, the difference between 6 metres and 5.9 metres would be crucial. *In that context*, the sentence 'The room is 6 metres wide' does not express an approximately true proposition (but it would probably do so if the room were 6.1 metres wide instead).

While the notion of approximate truth (as I use it here) is context-dependent, it seems possible to introduce a notion of *closeness to the truth* that it is not context-dependent. If 'The room is 5.9 meters wide' expresses a strictly true proposition, then, for any two real numbers, r and r^* such that $r^* > r \geq 5.9$ or $r^* < r \leq 5.9$, the proposition expressed by 'The room is r metres wide' *is closer to the truth* than the one expressed by 'The room is r^* metres wide'. The reason why the notion of closeness to truth is not context-dependent is that, while there may be contexts in which the proposition expressed by 'The room is r metres wide' is approximately true but the one expressed by 'The room is r^* metres wide' is not, there seem to be no contexts in which the reverse is the case. If the proposition expressed by 'The room is r^* metres wide' is "true enough", then so is the one expressed by 'The room is r metres wide'.

While the above is just a rough sketch of how the notion of closer to the truth can be analysed in simple cases, I think it is easy to see how this approach can be extended to cover similar or simpler cases. For example, if *p** is a (strictly) true proposition and *p** and *p* are contrary or contradictory propositions, it seems natural to say that *p** is closer to the truth than *p*. Much work would be needed to turn this sketch into a full-fledged, general account of 'closer to the truth'. However, I hope that this sketch can provide us with a sufficiently clear grasp of 'closer to the truth' for our present purposes.

2.3 Comparative Faithfulness

Whether something is an epistemic representation of something else for someone is an all-or-nothing matter. Whether something is a faithful epistemic representation of something else, on the other hand, is a matter of degree. An epistemic representation can be a more or less faithful to its target, and, of two epistemic representations of the same target, one may be more faithful (overall) than the other. At first, it might be tempting to think that (overall) faithfulness is a matter of how many true conclusions can be validly drawn from the vehicle to the target. However, this proposal does not stand up to scrutiny. First, this criterion allows for a completely faithful representation of a certain target to be less faithful than a partially faithful one (all it would take is for the partially faithful epistemic representation to have broader scope and for the set of true conclusions validly drawn from the completely faithful one to be a proper subset of the set of true conclusions validly drawn from the partially faithful one). Second, it seems to be the case that infinitely many true conclusions can be drawn from any partially faithful epistemic representation. From the 1930s London Underground map, for example, it is possible to draw infinitely many true conclusions about the London Underground network (e.g. 'It is not the case that the network includes 1,000 stations', 'It is not the case that the network includes 1,001 stations', 'It is not the case that the network includes 1,002 stations', and so on).

It therefore seems preferable to adopt the following definitions instead:

(13) *v* and *v** are (overall) *equally faithful epistemic representations* of *t* (for *u*) if and only if:

[13.1] *v* and *v** are epistemic representations of *t* (for *u*),

[13.2] *v* and *v** have the same scope, and

[13.3] :

[13.3.1] :

[13.3.1.1] Every true proposition that can be validly inferred from *v* can also be validly inferred from *v**, *and*

[13.3.1.2] Every true proposition that can be validly inferred from *v** can also be validly inferred from *v*, *and*

[13.3.2] :

[13.3.2.1] No false proposition that can be validly inferred from *v* is closer to the truth than the corresponding proposition that can be validly inferred from *v**, *and*

[13.3.2.2] No false proposition that can be validly inferred from *v** is closer to the truth than the corresponding proposition that can be validly inferred from *v*.

(14) *v** is (overall) a (strictly) *more faithful epistemic representation* of *t* than *v* if and only if:

[14.1] v and v^* are epistemic representations of t (for u),

[14.2] v and v^* have the same scope, and

[14.3] :

[14.3.1] :

[14.3.1.1] Every true proposition that can be validly inferred from v can also be validly inferred from v^*, *and*

[14.3.1.2] Some true proposition that can be validly inferred from v^* cannot be validly inferred from v, *or*

[14.3.2] :

[14.3.2.1] Some false proposition that can be validly inferred from v^* is closer to the truth than the corresponding proposition that can be validly inferred from v, *and*

[14.3.2.2] No false proposition that can be validly inferred from v is closer to the truth than the corresponding proposition that can be validly inferred from v^*.

(15) v is (overall) a (strictly) *less faithful epistemic representation* of t than v^* if and only if:

[15.1] v^* is a (strictly) more faithful epistemic representation of t than v,

A few remarks are in order here. First, the above definitions require that two epistemic representations have the same scope in order for them to be comparable in terms of faithfulness. This entails that the predicate 'x is (overall) a (strictly) *more faithful epistemic representation* of t than y' denotes a relation that is a partial order—i.e. it is possible that neither of two representations of the same target is (overall) a (strictly) *more faithful epistemic representation* of it than the other. The postcard of Rome and the map of Rome that I mentioned in Section 1.5, for example, have largely different scopes and there seems to be no clear sense in which either can be considered a more faithful epistemic representation of Rome than the other. By requiring that two representations have the same scope in order for them to be compared, these definitions also avoid situations (such as the one mentioned above) in which a partially faithful representation of a certain target turns out to be more faithful than a completely faithful one with a narrower scope.

Second, (14) only defines what it is for an epistemic representation of a certain target to be *strictly* more or less faithful than another. So, for example, a new London Underground map is strictly more faithful than an old 1930s map because there are no true conclusions that can be drawn from the latter but not from the former. However, there are probably cases in which an epistemic representation is more faithful than another without being *strictly* more faithful. If, for example, from one map of Rome it were valid to infer only true conclusions about a small area of Rome (say, Trastevere) while from another it were valid to infer only true conclusions about all areas of Rome except Trastevere, it might be tempting to regard the latter as an overall more faithful epistemic representation of Rome than the former, even if neither representation would be a *strictly* more faithful epistemic representation of Rome than other. However, with the exception of extreme cases like the one I just described, whenever neither of two representations is a strictly more faithful epistemic representation of a certain target than the other, we do not seem to have clear intuitions as to how the two compare in terms of overall faithfulness. Although I doubt that there are any general rules, the account I propose focuses on uncontroversial cases while leaving open the possibility that some of the more controversial cases may still be cases in which one epistemic representation is overall more faithful than the other (even if it is not *strictly* more faithful).

Third, the above definitions are only concerned with what I call 'the *overall* faithfulness of an epistemic representation'. However, we are often interested in whether an epistemic representation is *specifically* faithful—i.e. it represents faithfully some specific aspect of the target in which we happen to be interested. So, for example, although one of the two maps of Rome mentioned above may be more faithful than the other *overall* (even if not strictly so), it may not be the most faithful epistemic representation *for our specific purposes*. If the only thing we are interested in is finding our way through Trastevere, for instance, the map that is the less faithful overall might happen to be the one that is the most faithful for our purposes. Unless otherwise stated, in what follows I am concerned with the overall faithfulness of epistemic representations.

2.4 Conclusion: "The Problem of Scientific Representation"

In the literature about models, the question of how models represent real-world systems is often referred to as "the problem of scientific representation". From what I have said in the last two chapters, however, it should be clear that I take this to be a misnomer for at least two reasons. The first reason is that the problem does not seem to be specifically a problem about *scientific* representation—in fact, at best it is unclear whether there is something specific about how scientific models represent real-world systems that sets "scientific" representation apart from epistemic representation more generally. In other words, it seems advisable to assume that the question 'In virtue of what does a certain scientific model represent a certain real-world system?' seems to be just an instance of the more general question 'In virtue of what is a certain vehicle an epistemic representation of a certain target (for a certain user)?'. The second reason is that the problem is really not a single problem (as the definite article seems to suggest). There are at least two questions we should address in trying to solve "the problem of scientific representation"—i.e. (a) 'In virtue of what is a certain vehicle an epistemic representation of a certain target?' and (b) 'In virtue of what is a certain vehicle a *faithful* epistemic representation of a certain target (to the extent that it is)?'. In Part II and Part III of this book, I address each question in turn. Before doing so, however, in the next chapter I try to clarify what I mean by 'scientific models' here.

3 Models

3.1 Scientific Models as Epistemic Representations

How does what I have said so far relate to scientific models and the so-called problem of scientific representation? One of the main theses underlying this book is that scientific models are epistemic representations of certain portions or aspects of the world and, as such, we can use them to perform surrogative inferences about those portions or aspects of the world. This thesis is nicely illustrated by my initial example. In it, I used a stock model from classical mechanics as an epistemic representation of my daughters going downhill on their toboggan. In Part II and Part III of this book, I develop and defend general accounts of, respectively, epistemic representation and faithfulness that apply to models as well as other epistemic representations. But, before trying to account for what models *do*, it will be helpful to have a clearer grasp on what models *are*. In this chapter, I explain what I mean by 'model' and briefly discuss the relationship between what I call 'epistemic representation' and what is usually called 'scientific representation'. Before proceeding, however, I should note that what follows is only a sketch of an account of the ontological status of models; developing a full-fledged account would involve delving into a number issues in metaphysics, philosophy of language, and philosophy of mathematics that fall far outside the scope of this book. Nevertheless, I hope that this sketch provide us with a better grasp of my use of 'model'.

3.2 Mathematical Models, Material Models, Fictional Models

Scientists often refer to models in journal articles, textbooks, class notes, conference presentations, and conversations with colleagues, and they make assertions about those models, assertions that they seem to deem capable of being true or false. The author of a physics textbook, for example, might use apparently referring expressions such as 'the ideal pendulum' and 'the Rutherford model of the atom', and assert sentences such as: 'The ideal pendulum is not affected by friction' and 'In Rutherford model of the atom, the electrons move in well-defined orbits'.

Scientists do not seem to take the practice of referring to models and making assertions about them to be much more problematic than the practice of referring to apples and asserting they are red. And, indeed, this practice displays all the external indicators of being a successful linguistic practice. When talking about a certain model, scientists generally seem to agree about which object they are referring to, and they rarely seem to disagree about the truth or falsity of their claims about those objects. Yet, in most cases, it is far from clear what the entities supposedly referred to by those expressions are, what makes propositions about them true or false (if such propositions actually are capable of being true or false), or how one can find out which of them are true and which of them are false.

Given all the attention that philosophers of science have devoted to scientific models in the last few decades, one would expect to find in the literature some attempt to shed light on these questions. Somewhat surprisingly, however, this is not the case. Questions concerning the ontology and epistemology of scientific models are only rarely raised in the literature,[16] and serious attempts to answer those questions are rarer still. This phenomenon is even more surprising if one considers the amount of interest generated by analogous questions about the ontology and epistemology of mathematical objects in the philosophy of mathematics.

This lack of interest, I suspect, is partly explained by the fact that it is commonly believed that 'scientific models' is a catchall phrase for what is actually a heterogeneous collection of objects. It is commonly assumed that, if all scientific models have something in common, this is not their ontological nature but rather their function. As R.I.G. Hughes puts it, '[…] perhaps the only characteristic that all [representational] models have in common is that they provide representations of parts of the world' (Hughes 1997, S325).

In the literature, one finds two main functional characterisations of what models are, which Ronald Giere (1999) dubs, respectively, the instantial and the representational conception of scientific models. The *instantial conception* characterises a model of a certain theory as anything that satisfies that theory—that is, as anything of which that theory is a true description. The *representational account*, on the other hand, conceives of a scientific model as something that is used by us to represent some system in the real world.[17] Both the instantial and the representational accounts of scientific models, however, remain almost completely silent as to what kinds of entities do in fact perform the relevant function.

Personally, I do not object to functional characterizations of scientific models. In fact, I take the representational conception to be a step in the right direction. However, even if, from an ontological point of view, scientific models are a mixed bag and the best general characterization that one can give of them is a functional characterization, it does not follow that it is impossible to develop an informative account of the ontology of scientific models. Even if not all scientific models belong to a single ontological "kind", they might nonetheless belong to a few such kinds and we might be able to formulate satisfactory accounts of each of these. In other words, the heterogeneity of models certainly makes the task of formulating an account of their ontology more difficult, but it does not exempt us from that task.

Nor does it follow that the questions concerning the ontology of scientific models are any less pressing. To the contrary, even if models are characterized purely functionally, it is difficult to understand *how* a certain object can perform the relevant function, if we have no idea of what that object is.[18] Moreover, it seems that, if models play a central role in science, as most contemporary philosophers of science seem to agree that they do, we need at least the sketch of an answer to questions such as 'What is a scientific model?' in order to be able to answer questions such as 'What makes claims about a certain model true or false?,' 'How do we learn about models?' or 'How do we use models to represent the world?'

3.3 A (Provisional) Taxonomy

What kind of object, then, is a scientific model? At least in some cases, the answer to this question seems to be relatively straightforward. When my high-school biology teacher was talking about the model of DNA that still stands on one of the shelves in the school lab, for instance, she was referring to an actual concrete object that is about a meter tall, stands on a wooden pedestal, and consists of coloured plastic balls and thin metal rods arranged in a certain way around a metal pole. If we were in my high-school laboratory, I could point the model out

[16] For an exception see (da Costa and French 2003).

[17] The instantial account has been championed by the likes of Patrick Suppes (1960) and Bas van Fraassen (1989, Ch. 9), among others. Despite coming under attack from various fronts, it is still widely popular among the supporters of the semantic view of theories. Supporters of the representational account, which is becoming increasingly popular, include Ronald Giere (1988) and R.I.G. Hughes (1997).

[18] See, for example, (Martin Thomson-Jones 2010).

to you. We could touch it or take photographs of it. If we were to accidentally drop it on the floor, it might break. However, it would be possible to replace it by buying a new one for about a thousand dollars. In other words, the DNA model is a material object, just like the shelf on which it stands. I call models of this kind '*material models*'. Since material models do not seem to pose any ontological questions that are not already posed by other ordinary material objects, I will not focus on them here.

In the case of *mathematical models*, however, answering the above question is slightly more problematic. When talking about the logistic growth model, for example, biologists usually seem to be referring to an equation of the form:

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right).$$

In this and other cases, scientists seem to be referring to a mathematical object (often an equation, or a set of equations). If it is harder to say what kind of object a mathematical model is, that is because, from an ontological point of view, mathematical objects (if there even are any) are more elusive than material objects. If, when talking of mathematical models, scientists are just referring to mathematical objects, then the philosopher of science can delegate the task of investigating the nature of mathematical models to the philosopher of mathematics.

The vast majority of scientific models, however, do not seem to fall into either of the two above-mentioned categories. Consider a homely example from classical mechanics: the ideal pendulum. When referring to the ideal pendulum, scientists apparently refer to an object whose description can be found in many basic physics textbooks. This object is usually described as a point mass of mass m suspended from an inextensible, massless string of length l, which when displaced from its rest position of an angle θ, swings back and forth under the influence of a uniform gravitational field and which, not being affected by frictional forces, will continue to oscillate back and forth indefinitely.

The ideal pendulum is obviously not a material object. One cannot point at it, nor break it, nor photograph it (even if, in some textbooks, it is possible to find drawings of it). As the name suggests and the textbooks are usually keen to remark, no real pendulum has all the characteristics attributed to the ideal pendulum—no real pendulum bob (no matter how small) is a point particle; no real string (no matter how light and strong) is completely massless or inextensible; no pendulum (no matter how well lubricated) is absolutely frictionless, and so on.

Yet the ideal pendulum model is not just a mathematical object either. Although equations (and other mathematical objects such as state-space diagrams) can be used to describe some aspects of the behaviour of the ideal pendulum, the equations (or state space representation) should not be mistaken for the ideal pendulum itself. For example, the oscillations of the ideal pendulum can be described by means of the differential equation

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\sin\theta = 0.$$

However, this equation itself is not the pendulum (contrary to what people seem to think)—the equation is one way to describe the way the pendulum moves—and, if the pendulum moves in the way described by that equation, it is because of the characteristics that the pendulum is said to have—it is because the bob of the pendulum has a mass and it is in a gravitational field, that it oscillates in the characteristic way described by the above equation. The pendulum is not a set of trajectories in a state-space either (as, for example, van Fraassen (1989, p. 223) seems to suggest). A trajectory in a state-space is just a perspicuous way to represent certain aspects of the behaviour of the pendulum.

To sum up, scientific models such as the ideal pendulum can be described, can be drawn, are said to have characteristics that are typically ascribed to concrete objects, and yet they are not actual concrete objects. These features make models such as the ideal pendulum very similar to another sort of entity philosophers puzzle over—i.e. fictional characters. Fictional characters such as the Gruffalo are not actual concrete objects and yet they are said to have characteristics that only actual concrete objects seem to be capable of having (the Gruffalo, for example, is said to have curled out toes and a poisonous wart at the end of its nose). Also, it is not unusual to find drawings of the Gruffalo printed on the covers of books. Additionally, both fictional characters and scientific models seem to be human artefacts created by one or more people at some time (although it is not always easy or possible to identify who their authors are or when the entities in question were created). The Gruffalo, for example, was created by Julia Donaldson and Axel Scheffler sometime in the late 1990s and the Bohr model of the atom was created by Niels Bohr in the early 1910s. Due to these striking resemblances here I call models such as the ideal pendulum, the inclined plane, and the Bohr model of the atom '*fictional models*'.

The claim that there is a striking resemblance between some scientific models and fictional characters is not entirely novel. Nancy Cartwright, for one, famously claimed (though, I suspect, somewhat figuratively) that "[a] model is a work of fiction" (Cartwright 1983, p.153—but see also (Giere 1985), (Giere 1988) and (Godfrey-Smith 2006) for similar remarks). In this chapter, I argue that the resemblance between fictional models and fictional characters is not accidental—it is, so to speak, a family resemblance. Fictional models and fictional characters are two species of the same ontological genus—that of *fictional entities*.

In the absence of a specific account of fictional entities, however, the claim that models are fictional entities risks being an instance of explaining the obscure by way of the more obscure. In the philosophical literature, one can find a number of rival accounts of the nature of fictional entities. So, unless one specifies what one means by the claim that fictional models are fictional entities, that claim is condemned to remain little more than a stimulating metaphor. In this chapter, I intend to sketch such an account.

Two remarks are in order here. First, to say that both fictional entities and scientific models belong to the same ontological genus is not to say that there is no difference whatsoever between fictional models and fictional characters. To the contrary, there are many differences between them. For example, these two kinds of objects are found in completely different contexts (i.e. fictional literature and scientific discourse) and perform largely distinct functions. Therefore, when claiming that both fictional models and fictional characters are fictional entities, I do not mean to claim that the scientific practice of modelling is a form of fictional literature. Nor do I mean to claim that talking about models and talking about fictional entities serve the same purposes. Rather, I am claiming that that talking about models and talking about fictional entities are analogous linguistic practices, concerning objects that belong to the same ontological genus but that are used for largely different purposes.

Second, the claim that fictional models belong to the same ontological kind as fictional entities is to be clearly distinguished from the variety of scientific antirealism that is usually referred to as fictionalism (see (Fine 1993)). I take it that fictionalism is a form of instrumentalism according to which theories should not be construed literally but only as fictions aimed at "saving the phenomena" (whatever that means). So, a fictionalist believes that, when a theory states that, say, there are electrons, we should not take that statement literally. Rather, we should take the statement as part of a useful fiction that allows us to make certain empirical predictions. While a fictionalist might want to espouse the thesis that fictional models belong to the same ontological category as fictional entities, one does not need to be a fictionalist to do so. In fact, one can be a scientific realist and still believe that the *models* that we use to represent atoms or protons are fictional entities even if the atoms and protons themselves are actual concrete entities.

3.4 Actual Concrete Objects

In this and the next two sections, I consider three unsatisfactory proposals about the nature of fictional models. The first proposal is that fictional models are actual concrete systems, which we pretend have different characteristics than those they in fact have. According to this view, for example, 'the ideal pendulum' refers to some real pendulum or other whose string we pretend to be massless and whose bob we pretend to have no extension. If we were to take this proposal seriously, however, it would seem legitimate to ask *which* actual concrete pendulum is the ideal pendulum—is it the pendulum of the grandmother's clock in the hallway, or the one that stands on the table in one of the labs in the physics department? This is clearly a ludicrous question, for there seems to be no single concrete pendulum that is the referent of the expression 'the ideal pendulum' (or, at least, if there is one such pendulum, the vast majority of the users of the expression 'the ideal pendulum' ignore which concrete actual pendulum that expression refers to).

A more plausible variant of this proposal is that 'the ideal pendulum' does not always refer to the same actual concrete pendulum but to different actual concrete pendula on different occasions. So, for example, if we are investigating the behaviour of the tire-swing hanging from the tree in the backyard, we might pretend that the swing is an ideal pendulum. In doing so, we pretend, among other things, that the rope from which the swing hangs is inextensible and weightless, that the swing is not affected by frictional forces, and so on. This proposal seems to be quite plausible in those circumstances in which we use a model to investigate some specific concrete system or other. However, this is not always the case. In many cases, we talk and think about a certain model without having any specific concrete system in mind. In those cases, it seems to be questionable to assume that there is some actual concrete pendulum or other that we pretend is the ideal pendulum. For example, if a textbook exercise asks readers to determine when the tension of the rope in the ideal pendulum reaches its maximum, the readers will not typically think of some concrete actual pendulum or other and pretend that it is an ideal pendulum in order to answer the question. They simply need to think of a pendulum that fits the description of the ideal pendulum.

3.5 Possible Concrete Objects

The second proposal about the nature of fictional models is that 'the ideal pendulum' does not refer to any actual concrete system but rather to some merely possible but non-actual pendulum. On this proposal, even if no actual pendulum fits the description of the ideal pendulum, there could have been a pendulum that fitted that description—i.e. there could have been a pendulum whose string is massless and inextensible and whose bob is a point mass, and so on. It is this merely possible pendulum that is the referent of the expression 'the ideal pendulum'.

Many, I suspect, are likely to shrug off this proposal as unpalatable because of its reliance on merely possible objects. However, this, in and of itself, does not seem to be a good reason for dismissing it. Even if talk of merely possible objects is far from being uncontroversial or philosophically unproblematic, most of us seem to believe that there could have existed things that do not actually exist. For example, we seem to believe that, even if Richard Nixon did not actually have a son, he could have had one; or that, even if there is no solid gold sphere 20 miles in diameter, there could be one. Admittedly, making philosophical sense of these beliefs is not an easy task. However, with the possible exception of the most radical actualists, everyone seems to agree that we need to have some account of talk of merely possible objects. If this is correct, the thesis that fictional models are merely possible systems does not seem to pose any novel philosophical challenge.

One, however, might worry that fictional models could not have existed in the same sense in which Nixon's son or a huge golden sphere could have existed. The existence of Nixon's son and that of the huge golden sphere seem both metaphysically and nomically possible; the existence of a pendulum that fits the description of the ideal pendulum, on the other hand, does not seem to be nomically (or perhaps even metaphysically) possible. For one, it is not clear if it is nomically possible for there to be a pendulum whose bob is a point mass because all nomically possible massive objects are extended in space.

It is beyond the scope of this book to discuss whether inextended massive objects are nomically or metaphysically possible. Whatever the case may be, however, the existence of a pendulum that fits the description of the ideal pendulum seems to be at least broadly logically possible. That is, there seems to be no straightforward contradiction in conceiving of a world in which there is a pendulum whose bob a point mass and whose string is massless and inextensible, which is in a uniform gravitational field. And this sense of possibility might be all we need for this proposal to be viable.

This proposal, however, is still not satisfactory. First, this proposal faces a few problems analogous to those that beset the first proposal—i.e. the description of the ideal pendulum does not fit just one possible pendulum but many possible pendula. For example, although the description of the ideal pendulum tells us that its bob has a certain mass m and its string has a certain length l, it does not specify what the values of m and l are. We usually take this to mean that the bob of the ideal pendulum can have any mass and that its string can have any length. That description fits uncountably many possible pendula and, since all of them equally satisfy the description of the ideal pendulum, none of them has a better claim than the others to being the referent of 'the ideal pendulum'.

To avoid this difficulty, it might be tempting to suggest that 'the ideal pendulum' does not refer to the same possible pendulum in all contexts but to different possible pendula in different contexts. For example, if we are considering a tire-swing whose rope is 1.5m long and whose seat, together with the child sitting on it, has a mass of 15kg, we will set m and l to be respectively 15 and 150. In that context, the ideal pendulum would refer to (one of) the possible pendulum(s) whose string is 150 cm long and whose bob has a mass of 15 kg.

As we have already seen, however, we often think of the ideal pendulum without thinking of its string as having any specific length or its bob as having any specific mass. For example, we know that, no matter what the specific length of the string is, the period of the pendulum is given by $2\pi(l/g)^{1/2}$. One could suggest that in this case we are just making a general claim about all of the possible pendula that fit the ideal pendulum description. In this case, the ideal pendulum refers to an arbitrary possible pendulum that fits the description of the ideal pendulum. This is somewhat analogous to the way in which we can use an arbitrary triangle to demonstrate some fact about triangles in general, disregarding the specific characteristics of the triangle we are using (such as the length of its sides or the width of the angles between them).

The suggestion that 'the ideal pendulum' refers to different possible individual pendula in different contexts may successfully avoid the above difficulties. However, it does not square particularly well with the linguistic practice of talking about the ideal pendulum as if it was a unique object—not different objects in different contexts.

The view that fictional models are possible concrete objects, however, faces a more serious general objection, which is analogous to Kripke's objection to the view that 'the Gruffalo' refers to a possible but not actual creature. According to this objection, fitting the description of, say, the Rutherford model of the atom is neither a necessary nor a sufficient condition for something to be the Rutherford model of the atom. Consider sufficiency first. Suppose that it were possible for atoms to be exactly as Rutherford's model represents them to be. In spite of that, the objection goes, none of those possible atoms could possibly be the referent of the expression 'the Rutherford model of the atom'. 'The Rutherford model of the atom' refers to a *model* of the atom proposed by Ernest Rutherford in the 1910s in order to account for certain atomic phenomena and not to a possible concrete object (let alone some possible atom satisfying Rutherford's description of an atom). If Rutherford created something by creating his model of the atom, he certainly did not create a concrete object (let alone a concrete object that is an atom.) So, it is not sufficient that something fit the description of the Rutherford model of the atom in order for it to be the Rutherford model of the atom.

If fitting the description of the Rutherford model of the atom is not sufficient to be the Rutherford model of the atom, it does not seem to be necessary either. Although most of us agree that it is *in some sense* true that the electron in the Rutherford model of the atom circles the nucleus in well-defined orbit, few take it to be literally true of the model, which, if anything, is an object created by Rutherford sometime in the 1910s and not a possible atom.

3.6 Actual Abstract Objects

The third proposal about the nature of fictional models is that, while there is an obvious sense in which the Rutherford model of the atom does not actually exist (i.e. it is not an actual physical system), there is another sense in which the Rutherford model of the atom does actually exist (i.e. it is one of the best-known scientific models in the history of physics, which Ernest Rutherford devised in the 1910s to account for the phenomenon known as Rutherford scattering). According to this proposal, insofar as 'the Rutherford model of the atom' refers to anything, it refers to an actual abstract entity, not to a possible concrete one. It is, therefore, neither necessary nor sufficient for something to fit the description of the Rutherford model of the atom in order to be the Rutherford model of the atom. In fact, on this view, the description of the Rutherford model of the atom is literally false of it, for the model is an abstract object that does not literally have any of the concrete properties attributed to it in its description.

If the view that fictional models are possible concrete systems seems to take the descriptions of models too seriously, however, the view that fictional models are actual abstract objects does not seem to take the descriptions of models seriously enough. Even if we might not think that it is literally true that, say, the ideal pendulum has a bob that swings back and forth (for abstract objects don't have a mass, are not subject to forces, and don't move), we still seem inclined to believe that that description is, in some sense, true of it. It is because the pendulum, in some sense, "has" a bob that behaves in a certain characteristic way that we can use it to predict or explain the behaviour of certain real systems. However, if a fictional model is *just* an abstract entity, it is not clear how to make sense of the intuition that those descriptions, although not literally true of the ideal pendulum, are "in some sense" true of it.

3.7 The Dualist account and the Dual Nature of Fictional Models

Both the view that fictional models are possible concrete systems and the view that they are actual abstract systems seem to capture some of our intuitions about fictional models. However, neither of them seems to be entirely satisfactory. What is interesting is that the two views seem to complement each other—one view seems to be successful where the other is not. One way to see this is to consider the distinction between external and internal sentences. Sentences about fictional models seem to fall into one of two categories, which, adopting a distinction sometimes used in the literature on fictional entities, I call 'internal sentences' and 'external sentences'. *External sentences,* such as (1) 'The Rutherford model of the atom was created by Ernest Rutherford at the turn of the 20th century', talk of fictional models as models. Whether a sentence such as (1) is true or false seems to depend crucially on how the actual world is, and evidence for its truth or falsity is largely empirical evidence. An historian of physics, for example, might argue that (1) is false. Ernest Rutherford, she might argue, did not originally propose the model of the atom we usually refer to as 'the Rutherford model of the atom'. It had already been proposed and investigated by others, including the Japanese physicist Nagaoka, whose work Rutherford was familiar with and explicitly acknowledged in (Rutherford 1911).

Internal sentences, on the other hand, talk of the model as if it were a concrete physical system. Internal sentences include, for example, (2) 'In the Rutherford model of the atom, electrons orbit around the nucleus in well-defined orbits.' We seem to take sentences such as (2) to be true only *"in some sense"* (if true at all). If, for example, a physics student who takes a true-or-false physics test answers that (2) is true her answer will be correct; if she answers that it is false, her answer will be incorrect. Yet, we would probably tend to maintain that (2) is not *literally* true. For (2) is not about any actual concrete physical system—the "electrons" and "the nucleus" mentioned in (2), for example, are not the electrons and the nucleus of any actual atom, because, as we now believe, actual electrons do not move in well-defined orbits around the nucleus. So, whereas (2) may not be literally true, there is a sense in which we would be inclined to say that it is true.

Now, the view that fictional models are actual abstract systems seems to be successful in accounting for our intuitions that some internal sentences are literally true and that some external sentences are literally false. However, in and of itself, it does not seem to be able to accommodate the intuitions that some internal sentences are nevertheless *"in some sense"* true. The view that fictional models are possible concrete systems, on the other hand, seems to be partially successful in accounting for the fact that some internal sentences are "in some sense" true. However, it seems to take those sentences too seriously for, on that view, those internal sentences that are true are *literally* true not just "in some sense" true. Moreover, on that view, it is not clear how to vindicate the intuition that some external sentences are literally true.

This, I think, is to be ascribed to the fact that both views refuse to acknowledge that fictional models (as well as other fictional entities) have a dual nature. I will now sketch an account that by acknowledging the dual nature of fictional model combines the advantages of both the abstract object and the possible object views. I call this account 'the *dualist account*'.[19]

3.8 The Dualist Account

According to the dualist account, a fictional model is an abstract object that stands for one or another of a set of possible concrete systems.[20] External sentences such as (1) are apt to be literally true or false because they say something about the abstract object to which 'the Rutherford model of the atom' refers. Internal sentences such as (2), on the other hand, are literally false because the Rutherford model of the atom is an abstract object, which has neither an electron nor a nucleus as its proper parts. Nevertheless, in some circumstances, the abstract object referred to as 'the Rutherford model of the atom' acts as a stand-in for one of the possible systems that satisfy the description of the Rutherford model (i.e. possible systems in which an electron orbits a nucleus in a well-defined orbit and…) and, so, (2) can be considered to be true of the Rutherford model "in some sense" even if it is literally false of it. We could say that a sentence like (2) is true "by proxy". This is not unlike when we talk of an actor as if she were actually her character. Although our assertions are typically literally false of the actor, we take them to be in some sense true because they are true of the character the actor is playing. If we are watching a play and I say that one of the characters is really cruel and you ask me which one, I might point to one of the actors and say 'That one'. In this case, I am treating the actor as a proxy for the character she plays and I seem to be asserting something about her (i.e. that she's cruel) when, in fact, I'm only asserting something about the character she plays.

---

[19] Elsewhere I argue that the dualist account is equally successful in dealing with more typical examples of fictional entities such as Sherlock Holmes (see (Contessa 2009)).

[20] Here, I ignore the possibility of some models describing impossible systems.

Three remarks are in order here. The first two concern ontological economy. Admittedly, the dualist account is ontologically inflationary, for it requires that we include both abstract and possible objects in our ontology. However, in and of itself, this is not a reason to reject the dualist account. We are not supposed to accept the dualist account in virtue of its ontological austerity, but rather in virtue of its descriptive adequacy—it vindicates a large number of intuitions that underlie the way we think and talk about fictional models. Ockham's razor urges us not to postulate entities *unless they are indispensable*. So, according to Ockham's razor, *if* there were an account of fictional models that was as descriptively adequate as the dualist account but more ontologically austere, *then* we should prefer it to the dualist account. In the absence of such an account, however, Ockham's razor does not prevent us from accepting the dualist account with all its ontological baggage.

The second remark that is order is that the ontological baggage of the dualist account may not be as weighty as it first appears. The dualist account does not commit us to any specific view about abstract and possible objects and most philosophers agree that, since we often speak as if there were abstract and possible objects, we need some philosophical account of that way of speaking. For example, it might be possible to combine the above account with a fictionalist account of abstract objects and possible worlds. If this were the case, one might be able to reap some of the benefits of the dualist account without paying any of the associated ontological costs.

The third remark concerns the *standing-for* relation. Here, I only want to note that the relation that holds between the abstract object that is the model and the possible systems for which it stands is not an especially mysterious relation. Most philosophers accept that some objects stand for other objects. For example, a blue area on a map stands for an expanse of water, and, when we count five objects on fingers, each finger stands for one of the objects. The relation that holds between the abstract and the possible object according to the dualist account seems to be just another instance of the *standing-for* relation. So a satisfactory philosophical account of that relation that applies to those other ordinary cases should also apply to our case.

3.9 Generative Descriptions

So far, I have maintained that, in some sense, scientific models have the characteristics we attribute to them. But where do these characteristics come from? In this section, I sketch an answer to this question. On the dualist account, a scientist creates a scientific model by publicly describing a possible system in an appropriate context and manner and proposing it as a model of a certain kind of actual system. For example, Rutherford created his model of the atom by describing a certain possible physical system in his 1911 paper and by proposing it as a model for the atom. (However, it is important to note that, according to the dualist account I defend here, the model is not the possible system described by Rutherford but the abstract object that stands for it, and which was actually generated by Rutherford's speech act.)

In what follows I call the original description by means of which a model is created '*the generative description* (of that model)'. The generative description, I think, is important but not sacred—subsequent users of the model can modify the model by re-describing it. One important way to modify the model is its *specification*. The specification of the model occurs whenever one of its users substitutes some indefinite values of some characteristics of the model with definite values or specifies some boundary conditions. For example, one can set the length of the string of the ideal pendulum to some specific length or the initial conditions of the kinetic model of a gas as having low entropy.

Another important modification of the model is its *alteration*. The alteration of the model occurs whenever a user explicitly attributes to the model some characteristic that was not present in its original description, or a characteristic that slightly differs from the one in the original description. For example, in the kinetic model of gases, the container in which a gas is enclosed can have different designs. It can be a completely energetically isolated or it can be in contact with a constant heat source. It can have a removable partition or a piston at one end.

Modifications of a model are particularly important both for the application of the model to specific situations, and for the investigation of the model. In principle, it is possible to regard each description of a modified version of the model as the generative description of a new model. This proliferation of models does not seem problematic insofar as the "family relations" among the models are clear. Different versions of a model are all related to each other and, as such, they have clear "family resemblances": they share the most relevant characteristics of the original version of the model (where the user can decide which characteristics of the basic model are most relevant for the purposes at hand). For example, we usually tend to see Bohr model of the atom and Rutherford model of the atom as distinct models, while we tend to see the so-called Sommerfeld model of the atom (in which the orbit of the electrons is elliptical) as only a generalized version of the Bohr model of the atom.

Moreover, different versions of a model are related to each other in the sense that there are "family ties" between them: each version is a more or less explicitly acknowledged modification of an original model. For example, in his groundbreaking paper 'On the Constitution of Atoms and Molecules', after describing the Rutherford model of the atom, Niels Bohr proposes some crucial modifications to it. In the new model, electrons are confined to specific orbits and they do not radiate energy except when they "jump" from one orbit to another. This modified model has come to be known as the Bohr model of the atom, and it is a clear descendant of the Rutherford model of the atom.

The more an original model is modified, the more likely it is that the resulting model will be regarded as a different model. Personally, I do not think there is a clear-cut answer as to whether, say, the damped pendulum is a modified version of the ideal pendulum or is a different model altogether. The two models are closely related, and whether one sees them as two versions of the same model or two different models depends only on how one assesses their family resemblances.

According to this story, the generative description is necessarily a *correct* description of the possible systems for which the model stands—by hypothesis, the model "has" all the characteristics the description ascribes to it. But is the generative description also a *complete* description of the model? In other words, even if we assume that the model "has" *all* the characteristics that the generative description explicitly ascribes to it, does it "have" *only* those characteristics?

In his seminal work on scientific models (on whose insights the account I defend so heavily relies), Ronald Giere seems to answer this question affirmatively. Giere repeatedly maintains that "[a model] has *all* and *only* those characteristics *explicitly specified*" (Giere 1985, p.78; emphasis mine). I am not sure whether this is Giere's considered view on the subject, but it seems to me that it does not do justice to what we might call 'the openness of scientific models'. Scientific models have more characteristics than those that are *explicitly* attributed to them by their generative descriptions—and this "openness" is one of the reasons that models are so interesting for us.

The most obvious (though least interesting) examples of "surplus" characteristics are those that are *implicitly attributed* to the model. These are those characteristics that the model *must* have, in order to have the characteristics that are explicitly attributed to it. For example, an object cannot be 3cm long without being less than 10cm long. Nor can an object have mass and momentum without also having a velocity. Once certain characteristics are explicitly attributed to a scientific model by its generative description, other characteristics are attributed implicitly, as any system that has the former characteristics necessarily has the latter characteristics as well (where the necessity here is of the logical or metaphysical kind).

However, models are open also in a more interesting sense. Often models turn out to have characteristics that were neither explicitly nor implicitly attributed to the model by its generative description. For example, in the Rutherford model, electrons orbit around a small but massive positive nucleus.[21] But is such a system stable? In his 1911 article, Rutherford sets this question aside (Rutherford 1911, p.671) but seems to believe that the system would be stable (at least from a mechanical point of view). However, as Niels Bohr pointed out, the system described by Rutherford is not stable. According to classical electrodynamics, the orbiting electron, being both accelerated and charged, would radiate energy in the form of light, rapidly spiral towards the nucleus, and ultimately collapse into it. Irrespective of whether Rutherford foresaw this characteristic of the model, it is clearly not a characteristic that was *explicitly* attributed to it by him and yet it seems to be a characteristic that no one (not even Rutherford) denied the model had after Bohr had pointed it out.

Cases like this, I think, are the rule and not the exception in the history of science. Almost every model of scientific interest turns out to have characteristics other than those explicitly (or even implicitly) attributed by its generative description. Upon investigation, scientific models often turn out to "have" characteristics that were unforeseen to their authors. In some cases, like the one we have considered, these characteristics are undesirable. However, this need not be so. In many cases models turn out to have unforeseen characteristics which make them even more interesting (Poisson's famous bright spot is one such case). Scientific models thus are open in the sense that they are capable of having more characteristics than the ones explicitly attributed to them by their generative descriptions or foreseen by their authors and it is partly because of this openness that it is worth creating and investigating models.

But if fictional models have characteristics that are not explicitly attributed to them by their generative descriptions, where do these additional characteristics come from? Here, I only sketch an answer to this question. As we have seen, some of the additional characteristics simply come from the characteristics of the model that are explicitly attributed to it as a matter of logical or metaphysical necessity. However, other additional characteristics stem from the ones explicitly attributed to the model and the laws that govern the behaviour of the objects in the model. For example, the possible systems for which the ideal pendulum stands all obey the laws of classical mechanics. So, for example, the bob of the pendulum could not have, say, a (non-zero) acceleration explicitly attributed to it without a force acting upon it. Analogously, the Rutherford model of the atom is governed by the laws of classical electrodynamics and therefore, since according to those laws a negatively charged particle cannot accelerate without radiating energy, the electron in the model will ultimately collapse into the nucleus. This sketch of an answer will, I hope, prove a fruitful first step towards a better understanding of what I have called 'the openness of models'. Still, much work remains to be done before this can be developed into a truly satisfactory answer (for example, my answer does not account for those models in which some objects obey classical laws and others obey quantum mechanical laws).

3.10 Conclusion

In this chapter, I have developed and defended an account according to which one important class of models, which I have called fictional models, belongs to the same ontological genus as fictional characters. According to this account, a model is an actual abstract object that stands for

---

[21] It is not clear whether Rutherford was completely unaware of the instability of his model of the atom (as might be suggested by Rutherford 1911, p.688) or whether he just thought that at that stage the question of stability was premature (Rutherford 1911, p.671). In a later paper, Rutherford attributed to Bohr the realisation that "[…] the stable positions of the external electrons cannot be deducted from the classical mechanics" (Rutherford 1914, p.498). Whatever the case may be, even if Rutherford foresaw that his model of the atom might have been unstable, its instability was certainly not among the characteristics that he explicitly attributed to it.

one of the many possible concrete objects that fit the generative description of the model. My hope is that a better understanding of what models are (which I hope this account will be able to provide) will lead to a better understanding of what is it that we do when we use models in science.

Part II: Epistemic Representation

4 The Denotational Account and the Inferential Account

4.1 Introduction

In Part I, I distinguished between two questions—the question of what makes a vehicle an *epistemic representation* of a certain target and the question of what makes it a *more or less faithful* epistemic representation of that target. Part II is devoted to answering the first of these questions and Part III the second. Part II comprises two chapters. In the first, I consider two possible accounts of epistemic representation, which I call, respectively, 'the *denotational account* (of epistemic representation)' and 'the *inferential account* (of epistemic representation)', and explain why I do not find them satisfactory. In the second chapter (Chapter 5), I develop and defend an alternative account of epistemic representation, which I call 'the *interpretational account*'.

4.2 The Denotational Account

What I call 'the denotational account' is inspired by a proposal by Craig Callender and Jonathan Cohen (2006), who themselves rely explicitly on the work of Paul Grice. A similar and widely-discussed proposal in the philosophy of art is often attributed to Nelson Goodman (1968). However, since it is not completely clear if that is the correct interpretation of Goodman's views, in this section, I focus exclusively on Callender and Cohen's version of the proposal.

According to Callender and Cohen, there is a general strategy to reduce a variety of forms of representation to a single fundamental form of representation (most likely, mental representation, which is the representational relation that, supposedly, holds between mental states and what they represent). If successful, this strategy would reduce a number of distinct, though interrelated, problems to a single fundamental problem—i.e., the problem of providing an account of mental representation:

> […] the representational status of most [representational entities (cars, cakes, equations, etc.)] is derivative from the representational status of a privileged core of representations. The advertised benefit of this […] approach to representation is that we won't need separate theories to account for artistic, linguistic, representation, and culinary representation; instead, [those who adopt this general approach propose] that all these types of representation can be explained (in a unified way) as deriving from some more fundamental sorts of representations, which are typically taken to be mental states. (Callender and Cohen 2006, p.70)

According to Callender and Cohen, the representational relation between a vehicle and a target is ultimately a matter of convention or stipulation. If the appropriate conventions are in place, virtually anything can be used to represent virtually anything else.

> Can the saltshaker on the dinner table represent [Michigan]? Of course it can, so long as you stipulate that the former represents the latter. Then, when your dinner partner asks you what is your favorite [U.S. state], you can make the salt shaker salient with the reasonable intention that your doing so will activate in your audience the belief that [Michigan] is your favorite [U.S. state] (obviously, this works better if your audience is aware of your initial stipulation; otherwise your intentions with respect to your audience are likely to go unfulfilled). […] On the story we are

> telling, then, *virtually anything can be stipulated to be a representational vehicle for the representation of virtually anything* […]. (Callender and Cohen 2006, pp.73–74; emphasis added) [22]

From our vantage point, the problem with Callender and Cohen's proposal should already be obvious. Callender and Cohen fail to distinguish between what I called epistemic representation and denotation and, although their proposal may provide a satisfactory account of denotation, it completely misses the mark as an account of epistemic representation.

Although Callender and Cohen do not frame their proposal in these terms, the following would seem to be a fair formulation of their proposal in the terminology adopted here:

(A) *v* is an epistemic representation of *t* (for *u*) if and only if:

(A.1) *u* takes *v* to denote *t*.

(A) is what I refer to as *the denotational account of epistemic representation*. According to it, *denotation* (i.e. (A.1)) is a necessary and sufficient condition for *epistemic representation*. Since Callender and Cohen do not use this terminology and since nothing hangs on such terminological issues, however, I use as much as possible their terminology here and continue to refer to their proposal as 'Callender and Cohen's proposal' and leave it to the reader to determine whether (A) is a fair interpretation of their proposal.

There are few better illustrations of why Callender and Cohen's proposal is unsatisfactory than one of their own examples. Apparently, people who are familiar with the geography of Michigan often exploit the similarity between the shape of the state and the shape of an upturned right hand to convey information about the geography of Michigan. So, if you were to ask Ahmed who lives in Port Austin, MI, where the village is located, he might point to the tip of the thumb of his upturned hand and say 'here'. Of course, the village is not on the tip of his thumb—it's just that the village is in the area of Michigan that roughly corresponds to the point of his hand he indicated. In order to extract that piece of information from his seemingly mysterious speech act, you have to perform a surrogative inference from Ahmed's upturned hand to the state of Michigan.

It seems clear that your and Ahmed's agreement that his right hand is to be taken to "represent" (i.e. denotes) Michigan is not sufficient to explain your exchange of information. Had you and Ahmed agreed that the saltshaker "represented" (i.e. denoted) Michigan, it is hard to see how Ahmed could have answered your question simply by pointing to a spot on the surface of the saltshaker and saying 'here'. This is because, if all you and Ahmed did was stipulate that the saltshaker "represented" (i.e. denoted) the state of Michigan, there would be no obvious way for you to perform the relevant piece of surrogative reasoning from the saltshaker to the state of Michigan. In other words, while both the hand and the saltshaker might well be equally good candidates for *denoting* the state of Michigan, they do not seem to be equally good candidates for being *epistemic representations* of (the geography of) the state of Michigan.

The point I am making is not that it would not be possible for you and Ahmed to use the saltshaker as an epistemic representation of (the geography of) the state of Michigan. According to the account of epistemic representation I defend in Chapter 5, if the appropriate conditions were in place, nothing would prevent the saltshaker from being (the vehicle of) an epistemic representation of the geography of the state of Michigan. The point I am making is that the mere stipulation that the saltshaker "represents" (i.e. denotes) the state of Michigan would not turn the saltshaker into an epistemic representation of (the geography of) Michigan (or of anything else for that matter). This is because the mere fact that the saltshaker denotes the state of Michigan

[22] Here I took the liberty of combining two distinct examples used by Callender and Cohen and that, in order to do this, I had to slightly modify the saltshaker example by substituting 'Michigan' for 'Madagascar' and 'US state' for 'landmass' in the quotation to which this note is appended. As far as I can see, these modifications do not have any substantial philosophical consequences.

does not license any valid surrogative inferences from the saltshaker to the geography of Michigan.

Now, Callender and Cohen would probably try to dismiss our preferences as a matter of pragmatics (as opposed to, presumably, semantics):

> […] it should be clear that the constraints ruling out these choices of would-be representational vehicles are pragmatic in character: they are driven by the needs of the representation users, rather than by essential features of the artefacts themselves. (Callender and Cohen 2006, p. 76)

What Callender and Cohen are claiming is that, even if it is *in principle* possible to use anything to represent anything else, *in practice* pragmatic constraints make some vehicles more suitable than others. But what is meant here by 'pragmatic constraints'? I think it is highly plausible to maintain that pragmatic constraints play a decisive role in our preferring, say, a light, foldable paper map of the London Underground over one in which the same marks are reproduced on a very large and heavy slate of stone. Even if the two maps were equally faithful epistemic representations of the London Underground network, the paper map would still be preferable on the grounds that it is much easier to carry around and consult than the stone map. However, I do not think that the constraints that make us favour the hand over the saltshaker can be as easily dismissed as *merely* pragmatic. And, ultimately, Callender and Cohen don't seem to think so either. In a particularly revealing passage, they write:

> […] the geometric similarity between the upturned human right hands and the geography of Michigan make the former a particularly useful way of representing relative locations in Michigan, and it normally would be foolish (but not impossible!) to use [a saltshaker] for this purpose since a more easily interpreted representational vehicle is typically available. (Callender and Cohen 2006, p.76)

In conceding this much, however, Callender and Cohen have given away the game. As Dominic Lopes (1996 pp.132–133), following David Lewis' analysis of convention (Lewis 1969, p.76), notes against (what is usually taken to be) Nelson Goodman's view of artistic representation, a choice is entirely conventional (with respect to a certain set of alternatives) only if those who make it have no intrinsic reason to prefer one of the available alternatives to any of the other alternatives in that set. For example, the choice of using 'gatto' (as opposed to 'cane' or 'Katz') to refer to cats seems to be entirely conventional—there seems to be no intrinsic reason why it should be so and not, say, the other way around (i.e. there seems to be no intrinsic property of the word 'gatto' that makes it a more suitable candidate than, say, 'cane' or 'Katz' for the job designating cats). If, on the other hand, as it is sometimes suggested, the prevalence of certain means of transportation together with the fact that most people are right-handed originally affected the choice between driving on the right-hand or left-hand side of the road, then the choice between those two options is not as conventional as is usually assumed.

Going back to our example, if we want to show someone the location of the village of Port Austin, Michigan, our preferences seem to be entirely clear: we would prefer a map to an upturned right hand and an upturned right hand to a saltshaker (or any other random object, for that matter). Callender and Cohen not only acknowledge our preferences, but go so far as to speculate about the reasons behind them. The reasons they suggest seem to be *intrinsic* reasons: it is because the hand has a certain intrinsic property—its shape—that we prefer it to the saltshaker. To maintain, as Callender and Cohen do in the above passage, that our choice of a hand rather than a saltshaker as a representation of the geography of Michigan is merely conventional because, though foolish, it would not be *impossible* to choose the saltshaker seems to be as disingenuous as maintaining that, if we were offered the choice between a happy life and a miserable one (with all else being equal), our choice would be conventional because, though it might be foolish, it would not be impossible for us to choose the miserable life over the happy one.

But then what exactly underpins our preference of the hand over the saltshaker? I think that, as it happens, Callender and Cohen's explanation of our preference is, by and large, correct and it is similar to the explanation a supporter of the interpretational account (which I develop and defend in Chapter 5) would give—the saltshaker would be a poor choice as a vehicle exactly because 'a more easily interpreted representational vehicle [(i.e. an upturned right hand)] is typically available'. Callender and Cohen, however, take great pains to avoid acknowledging that we prefer the hand to the saltshaker because the former is more easily interpreted as an epistemic representation of Michigan than the latter, as this undermines their claim that all it takes for something to "represent" the geography of Michigan is for us to stipulate that it does. If the saltshaker is to serve as the vehicle of an epistemic representation of Michigan, we must also come up with an interpretation of the saltshaker in terms of the geography of the state of Michigan, an interpretation that is already readily available in the case of the hand but not in the case of the saltshaker. In Chapter 5, I argue that such an interpretation is the missing ingredient for turning the saltshaker into an epistemic representation of Michigan. Before turning to that, however, I examine another unsuccessful proposal: the inferential account of epistemic representation.

4.3 The Inferential Account

In the previous section, I argued that the main shortcoming of the denotational account of epistemic representation is that *denotation* by itself is not sufficient for *epistemic representation*, because it is not sufficient for a user to take a vehicle to denote a target in order for that user to be able to perform surrogative inferences from the vehicle to the target. The inferential account of epistemic representation tries to overcome this problem simply by adding the further requirement that the user be able to perform surrogative inferences from the vehicle to the target. According to *the inferential account of epistemic representation*:

(B) *v* is an epistemic representation of *t* (for *u*) *if and only if*:

(B.1) *u* takes *v* to denote *t*, *and*

(B.2) *u* is able to perform (valid though not necessarily sound) surrogative inferences from *v* to *t*.

So, for example, according to (B), the difference between the upturned right hand and the saltshaker would be that, even if both can be taken to denote the state of Michigan equally well, we are able to perform surrogative inferences about the geography of Michigan from one but not from the other.

Terminological issues aside, (B) is meant to convey the spirit of the account of representation developed and defended by Mauricio Suárez (see (Suárez 2004), (Suárez 2005), (Suárez and Solé 2006)). However, I should note that there are two substantial differences between (B) and Suárez's proposal. The first is that Suárez puts forward his proposal as an account of what he calls '*scientific* representation,' while (B) is meant to be an account of *epistemic* representation. Personally, I suspect that this is merely a terminological difference and that we should not take the label used by Suárez to suggest that non-scientific epistemic representations fall outside of the scope of his account. Indeed, Suárez often discusses what we would call non-scientific epistemic representations as examples of what he calls 'scientific representation'. Even if my interpretation of the inferential account does not reflect Suárez's actual views, it still seems plausible to interpret it as an account of epistemic representation. And, in any case, since my criticisms of it apply independently of what interpretation one adopts, I do not think that much hangs on this interpretive question.

The second substantial difference is that, according to (B), *denotation* (i.e. (B.1)) and *surrogative reasoning* (i.e. (B.2)) are both individually necessary and jointly sufficient for *epistemic representation*, while according to Suárez's original proposal, *denotation* and *surrogative reasoning* are necessary but not sufficient for *epistemic representation*. Part of the reason why I ignore this aspect of Suárez's proposal is that Suárez's position and his reasons for adopting it are neither entirely clear nor always consistent.[23] As such, addressing this aspect of Suárez's proposal would needlessly complicate our discussion, distracting us from its most significant and interesting features. In any case, regardless of whether Suárez would subscribe to (B) because of the sufficiency issue, (B) does not seem to be any less plausible to me for that reason[24], and my own issues with (B) have nothing to do with the fact that it takes *denotation* and *surrogative reasoning* to be both necessary and sufficient for *epistemic representation*. My criticism would apply as well to a version of the inferential account that takes denotation and surrogative reasoning to be necessary but not sufficient for epistemic representation.

For the sake of clarity and consistency, in what follows I use 'the inferential account (of epistemic representation)' to refer exclusively to (B) and *not* to Suárez's own proposal. However, I take it that almost everything I'll say about the former also applies to the latter (but I leave it to the reader to decide whether that is actually the case in each particular instance).

One of the most prominent features of the inferential account of epistemic representation is that it is meant to be a deflationary account. According to the inferential account, there is nothing deeper to epistemic representation than its surface features. My main problem with the inferential account is not that it is inadequate (how could I think so if I have defined epistemic representation in terms of surrogative reasoning?) but rather that it is uninformative. In other words, even if, in some cases, it is advisable to adopt a deflationary account of a philosophical notion, I doubt epistemic representation is one of those cases. My key reason for thinking so is that I think there is in fact something deeper to epistemic representation than its surface features. If there is something deeper to epistemic representation than surrogative reasoning, then deflationism about epistemic representation is unwarranted. Unfortunately, I cannot argue this point until I have specified what deeper feature of epistemic reasoning I have in mind, which I do in the next chapter. For the time being, I just mention two (somewhat related) considerations that should make us suspicious of the inferential account's deflationary approach to epistemic

---

[23] Suárez offers two possible interpretations of the non-substantiality of the inferential account. On the one hand, Suárez claims that one should not look for further conditions because there are '[…] no deeper features to scientific representation other than its surface features' (Suárez 2004, p.769). According to Suárez, these features are surface features in the sense that they are features of the concept of scientific representation (Suárez 2004, n. 4). On this interpretation, Suárez claims, the inferential account would be a deflationary account of scientific representation (Suárez 2004, pp.770–771). On the other hand, Suárez seems to think that there are further, more concrete conditions by virtue of which the concept of scientific representation applies to cases of scientific representation, but that these further conditions differ from case to case. For example, Suárez claims that: 'in every specific context of inquiry, given a putative target and source, some stronger conditions will typically be met; but which one specifically will vary from case to case. In some cases it will be isomorphism, in other cases it will be similarity, etc.' (Suárez 2004, p.776). According to this interpretation, Suárez would be claiming that an account of representation can spell out a set of necessary conditions for the concept of scientific representation but it cannot spell out a set of necessary and sufficient conditions for its application. On this interpretation, the inferential account would be, Suárez says, a *minimalist* account of scientific representation.

[24] In fact, I think it is more plausible (see (Contessa 2007) for an argument that *denotation* and *surrogative reasoning* are sufficient for epistemic representation). In a joint paper with Albert Solé (Suárez and Solé 2006), Suárez goes so far as to suggest that denotation and surrogative reasoning may well be both necessary and sufficient for the definition of the *concept* of scientific representation, but that this does not exclude that some further conditions (such as similarity or isomorphism) must be met in each concrete *application* of that concept. Even if this were the case, argue Suárez and Solé, the inferential account would not be a substantial account of scientific representation because, even if they were both necessary and sufficient, denotation and surrogative reasoning would only be surface features of scientific representation and a substantial account of scientific representation is one that identifies *non-trivial* necessary and sufficient conditions for scientific representation.

representation, considerations that, I think, help to point us in the direction of the deeper feature of epistemic representation that I have in mind.

The first source of suspicion about the inferential account is that it seems to get the relationship between epistemic representation and surrogative reasoning the wrong way around. If a certain vehicle is an epistemic representation of a certain target for a certain user, it does *not* seem to be one *in virtue of* that user's ability to perform surrogative inferences from the vehicle to the target. In fact, it seems like just the opposite is the case—it seems that a user is able to perform surrogative inferences from a vehicle to a target *in virtue of* the fact that that vehicle is an epistemic representation of that target for that user. For instance, it is in virtue of the fact that the map is an epistemic representation of the London Underground network (for me) that I can use it to perform surrogative inferences from it to the London Underground network, and not the reverse. If I did not already take this piece of glossy paper with coloured lines printed on it to be an epistemic representation of the London Underground network, it would never cross my mind to use it to find my way around the London Underground network in the first place. So, surrogative reasoning presupposes epistemic representation and cannot account for it on pain of circularity, as epistemic representation seems more fundamental than surrogative reasoning.

The second consideration is that the inferential account makes surrogative reasoning unnecessarily mysterious. On the inferential account, the user's ability to perform valid surrogative inferences from a vehicle to a target seems to be a brute fact that cannot be further explained in terms of more fundamental facts that ground it. This makes the connection between *epistemic representation* and *surrogative reasoning* needlessly obscure and the performance of valid surrogative inferences an activity as mysterious and incomprehensible as soothsaying or divination. Moreover, the inferential account does not seem to provide us with any principled way to distinguish between valid and non-valid surrogative inferences. What makes some surrogative inferences valid and others not? Why does it seem that under normal circumstances it would be valid for me to infer from the London Underground map that Central Line trains operate between Liverpool Street and Holborn but not that the Queen was born in 1926 or that there are no eggs left in my fridge?

The relationship between epistemic representation and surrogative reasoning is analogous to the relationship between measles and Koplik spots. Even assuming that all and only the people who contracted measles develops Koplik spots, one does not have measles *in virtue of* the fact that one has developed Koplik spots. One both has measles and develops a measles rash *in virtue of* some underlying fact—i.e. the fact that one has contracted the measles virus. Analogously with epistemic representation, although *surrogative reasoning* is a necessary and sufficient condition for *epistemic representation*, this should not be taken to mean that one obtains in virtue of the other. So, if there are any deeper conditions in virtue of which both *epistemic representation* and *surrogative reasoning* obtain, and in virtue of which *epistemic representation* and *surrogative reasoning* are so closely related, it behoves any suitable account of epistemic representation to try to identify what those conditions might be. It is only if no such deeper conditions can be found that a deflationary stance towards epistemic representation is warranted. Otherwise, a non-deflationary account of epistemic representation is preferable to a deflationary account simply on the grounds of being more informative and illuminating. In the next chapter, I develop and defend an account of epistemic representation that, I claim, identifies a deeper condition in virtue of which both *epistemic representation* and *surrogative reasoning* obtain. I think establishing the existence of such a condition is the best argument I can offer against the inferential account or against any other deflationary approach to epistemic representation.

4.4 CONCLUSIONS

In this chapter I have examined two possible accounts of epistemic representation and found them both wanting. The denotational account fails to identify the distinguishing property of epistemic representations, namely that they afford their user rules for drawing valid surrogative inferences. The inferential account avoids this problem, but it does so by taking the phenomenon of surrogative inference as basic and explaining epistemic models in terms of it. I believe this gets things the wrong way around. We want to understand *why* and *how* epistemic representations make valid surrogative inference possible—so it is inadequate for our purposes to simply stipulate *that* they do so. In the next chapter, I develop an account of epistemic representation, the interpretational account, which, unlike the denotational account and the inferential account, sheds light on the connection between epistemic representation and surrogative reasoning.

5 The Interpretational Account

5.1 INTRODUCTION

In Chapter 1, I defined what it is for a vehicle to be an epistemic representation of a target (for a user) in terms of the user's ability to perform valid surrogative inferences from the vehicle to the target. This definition tells us that *surrogative reasoning* is both necessary and sufficient for *epistemic representation*. However, in Chapter 4, I argued that an account of epistemic representation that stops at this level of analysis is not satisfactory, for it suggests that the user's ability to perform surrogative inferences is a brute fact that cannot be further explained in terms of more fundamental facts. A satisfactory account of epistemic representation should, if possible, identify the deeper conditions in virtue of which the vehicle is an epistemic representation of the target for a user and, as such, can be used by that user to perform valid surrogative inferences to the target. In this chapter, I develop and defend an account of epistemic representation, which I call 'the interpretational account of epistemic representation' and which, I claim, does just that.[25]

The interpretational account of epistemic representation maintains that:

(C) *v* is an epistemic representation of *t* (for *u*) *if and only if*:

(C.1) *u* takes *v* to denote *t*, *and*

(C.2) *u* adopts an interpretation of *v* in terms of *t*, $i(v \rightarrow t)$.

According to the interpretational account of epistemic representation, *denotation* (i.e. (C.1)) and *interpretation* (i.e. (C.2)) are individually necessary and jointly sufficient conditions for *epistemic representation*. More importantly, according to the inferential account it is *in virtue of* the fact that *denotation* and *interpretation* obtain that *epistemic representation* obtains, or so I argue here.

The next two sections (5.2 and 5.3) are devoted to clarifying what *denotation* and *interpretation* amount to. Section 5.2 explains informally how, when denotation and interpretation obtain, the vehicle is an epistemic representation of the target. Section 5.3 gives a more formal definition of what it takes for a user to adopt a specific but very common kind of interpretation of a vehicle in terms of a target.

5.2 DENOTATION AND INTERPRETATION: AN EXAMPLE

Suppose that you are commissioned to design a map of a subway network *ex novo*. One way to go about this would be as follows. First, before even knowing what the actual network looks like, you could identify what types of objects on the network and what types of properties of and relations among those objects are relevant to your map based on what aspects of the subway system the map is supposed to represent. Let me call these types of objects, properties and

[25] Although the interpretational account of epistemic representation somewhat resemble the DDI account of representation sketched by R.I.G. Hughes (1997), due to the lack of detail in Hughes' proposal, I am not sure if the interpretational account should be considered a development of the DDI account or a distinct account. I leave it to the reader to determine which of these two interpretations is correct.

relations, respectively, 'the *t*-relevant objects', 'the *t*-relevant properties', and 'the *t*-relevant relations' (i.e. the objects, properties, and relations in the target that are relevant to your representation). Given that, in our example, the map is supposed to represent which stations are connected by the network's train lines, the *t*-relevant types of objects will include the stations on the network, the *t*-relevant properties of those objects will include the names of those stations, and the *t*-relevant relations among them will include the train services operating between those stations. However, not all objects, properties, and relations are *t*-relevant. For example, if your map is not supposed to represent the distances between stations, the distances between the stations will not be among the *t*-relevant relations your map is intended to capture.

The second step in designing your map of the subway network consists in selecting which types of objects on the map and which types of properties and relations among them will denote the *t*-relevant objects, properties and relations. For example, you can decide that, on your map, stations will be denoted by small black circles, that a circle bearing a name will denote the station with that name, and that direct train services operating between two stations will be denoted by coloured lines connecting the corresponding circles, with each train line distinguished by a different colour. I call '*v-relevant*' the sorts of objects, properties, functions, and relations on the map that denote *t*-relevant types of objects, properties, and relations.

After completing these two steps, you will have developed what I call an 'interpretation'. An interpretation has two interesting features. First, in an interpretation, which types of objects, properties, and relations are *t*-relevant depends on which aspects of the target the epistemic representation is supposed to represent. In other words, this is contingent on what I have called the *scope* of the representation. Had you been interested in representing other aspects of the subway network, you would have selected other types of objects, properties, and relations as *t*-relevant. For example, had you been interested in designing a map of the subway system useful to train drivers rather than passengers, the *t*-relevant objects might have included tracks, interchanges, signals, and platforms, rather than stations and train lines.

Second, in an interpretation, it is to some extent arbitrary which objects, properties, and relations on the map are *v*-relevant. Within the limits of common sense, it would have been possible for you to use different types of objects, properties, and relations on the map to denote the *t*-relevant types of objects, properties, and relations in the world. For example, nothing would have prevented you from using small black squares instead of small black circles to denote stations on the map. Pragmatic constraints, however, do set some clear limits to the arbitrariness of your choices in these matters. For example, it would be highly impractical to use elephants and extremely expensive to use precious stones. Notwithstanding sheer impracticality and animal rights issues, however, nothing would prevent you from producing an epistemic representation of a subway network in which stations are denoted by elephants and direct train services operating between two stations are denoted by the corresponding elephants being tied together by a coloured ribbon.

Once you have developed a general interpretation like the one above, you can use it to design an actual map of the specific subway system in which you are interested. You first turn to the actual subway system, making note of all and only those objects that are tokens of the *t*-relevant types of objects, and of any of the objects' properties or relations that are tokens of the *t*-relevant types of properties and relations. In our case, you would compile a list of stations and their names, and make note of the train lines directly connecting them to one another.

Then, you would draw on your map one and only one small black circle for each station on your list and draw coloured lines between any two stations that are connected by a direct train service (using a different colour for each line). As a result of this process, you will have designed a map of the subway system in question. Here, I call the interpretation on the basis of which you designed the map 'the *intended interpretation* of the map'.

Now consider the map you have just designed from the perspective of one of its users. In and of itself, the map is not an epistemic representation of anything—it is just a piece of paper with small circles and coloured lines printed on it. If the map is to become an epistemic representation for a user, the user must take some of the objects, properties, and relations on that piece of paper to denote something else—i.e. the user has to adopt some interpretation or other of the map.

If the user is familiar with interpreting other maps of subway systems, it will probably be easy for her to realize that that piece of paper she is holding in her hands is meant to be a map of a subway system, and furthermore that the small black circles on the map denote the stations of that system, and that the coloured lines between the circles denote direct train connections between stations. In other words, a user who is familiar with general interpretations of other subway maps is likely to swiftly grasp the standard interpretation of your map.

However, the user need not adopt the standard interpretation of the map in order for it to be an epistemic representation for her. For example, the user can take the circles to denote cities and towns, the coloured lines to denote highways, and the relative positions of the circles to denote the relative positions of the corresponding cities. Under a non-standard interpretation such as this one, the map would still be an epistemic representation for its user—although it would be an epistemic representation of a highway system rather than a subway system.

Suppose, however, that the user in question does in fact adopt the standard interpretation of your map. According to the interpretational account, the map is now an epistemic representation of a subway system for that user. So it might seem that *interpretation* is not only necessary but also sufficient for *epistemic representation*. Therefore, one might wonder whether *denotation* is superfluous after all. The reason why *denotation* is needed in addition to *interpretation* is that, if the user only adopts an interpretation of the map without taking the map to denote any specific subway system, the map will only be a representation of some subway system or other and not an epistemic representation of any specific subway system. It is only when the user takes the map to stand for a specific subway system that the map becomes an epistemic representation of *that* subway system for that user. So for example, a user who adopts the standard interpretation of the map but does not know that the map is a map of the Toronto subway network (and is unfamiliar with the city of Toronto and its subway network) can still infer that one of the stations on the subway network represented by the map is called 'Bathurst' even if, from this, she cannot infer that that station is part of the Toronto subway network.

Somewhat schematically, we could say that, when a user adopts an interpretation of a certain vehicle in terms of *a generic target*, the vehicle becomes an epistemic representation of some target or other for that user. However, it is only when the user takes that epistemic representation to denote *a specific target*, that the epistemic representation becomes an epistemic representation *of that target* for that user. In our example, whenever a user adopts an interpretation of this piece of paper, the piece of paper becomes a map of *a* subway system for that user; however, the map becomes a *map of the Toronto subway system* only when the user also takes the map (*qua* epistemic representation) to stand for the Toronto subway system.

A user's interpretation of a vehicle in terms of a target is not an interpretation of the vehicle in terms of any specific target, unless the vehicle is also taken by the user to denote a specific target (i.e. unless *denotation* also obtains). For the sake of simplicity, in what follows I will usually talk of a vehicle's interpretation in terms of the target, thus assuming that the user already takes the vehicle to denote some specific target. However, this is only a loose way of speaking, for, as I have argued in this section, *interpretation* does not entail or presuppose *denotation*.

5.3 Analytic Interpretations

In the previous section, I provided a specific example of what I mean by 'interpretation'. But what does it mean *in general* for a user to adopt an interpretation of a vehicle in terms of a target? According to a general, though somewhat loose, characterisation of the notion of interpretation, a user interprets a vehicle in terms of some target only if she takes facts about the vehicle to stand for (putative) facts about a target. One specific way to interpret a vehicle—though by no means the only way—is to adopt what I call 'an *analytic interpretation* of the vehicle'.

In order for a user to adopt an interpretation of a certain target, the user must identify a (non-empty) set of *v*-relevant objects in the vehicle ($\Omega^{V}=\{o_1^{V}, \ldots, o_n^{V}\}$), a (possibly empty) set of *v*-relevant properties of and relations among the *v*-relevant objects in the vehicle ($P^{V}=\{{}^{n}R_1^{V}, \ldots, {}^{M}R_r^{V}\}$, where ${}^{n}R$ denotes an *n*-ary relation and properties are construed as 1-ary relations), and a (possibly empty) set of *v*-relevant functions, $\Phi^{V}=\{{}^{n}F_1^{V}, \ldots, {}^{M}F_r^{V}\}$ (where ${}^{n}F$ denotes an *n*-ary function), either from $(\Omega^{V})^{n}$ to Ω^{V}. The user must also assume that, in the target, there is a (possibly empty) set of *t*-relevant objects in the vehicle ($\Omega^{T}=\{o_1^{T}, \ldots, o_n^{T}\}$), a (possibly empty) set of *v*-relevant properties of and relations among the *v*-relevant objects in the vehicle ($P^{T}=\{{}^{n}R_1^{T}, \ldots, {}^{M}R_r^{T}\}$) and a (possibly empty) set of *v*-relevant functions ($\Phi^{T}=\{{}^{n}F_1^{T}, \ldots, {}^{M}F_r^{T}\}$) either from $(\Omega^{T})^{n}$ to Ω^{T}.[26]

(16) *u* adopts *an (analytic) interpretation* of *v* in terms of *t*, $i°(v \longrightarrow t)$, if and only if:

[16.1] The user takes[27] every object in Ω^{V} to denote one and only one (putative) object in Ω^{T} and every (putative) object in Ω^{T} to be denoted by one and only one object in Ω^{V},

[16.2] The user takes every *n*-ary relation in P^{V} to denote one and only one *n*-ary relation in P^{T} and every *n*-ary relation in P^{T} to be denoted by one and only one *n*-ary relation in P^{V},

[16.3] The user takes every *n*-ary function in Φ^{V} to denote one and only one *n*-ary function in Φ^{T} and every *n*-ary function in Φ^{T} to be denoted by one and only one *n*-ary function in Φ^{V}.

What is special about analytic interpretations is that, according to them interpretation, *objects* denote *objects*, *properties* denote *properties*, *binary relations* denote *binary relations*, and so on. Note, however, that, usually, not all *prima facie* objects, properties, functions, and relations in the vehicle or the target are relevant to some user's interpretation of the vehicle in terms of the target. The only *v*-relevant objects, properties, and relations are those that denote objects, properties, and relations in the target, and the only *t*-relevant objects, properties, functions, and relations in the target are those that are denoted by, respectively, objects, properties, functions, and relations in the vehicle. For example, the relation *being connected by a light blue line* in the London Underground map is relevant (according to the standard interpretation of the map in terms of the network) because, on the standard interpretation of the map, it denotes a relation between stations on the network, but the relation *being two inches to the left of* is not relevant because, on the standard interpretation of the map, it does not denote any relation among stations in the network.

For the sake of simplicity, in what follows I sometimes call the object, property, or relation in the target that is denoted by an object, property, or relation in the vehicle according to a certain

---

[26] I use functions mainly as stand-ins for certain properties and relations that are determinates of determinables. So, for example, the binary relation *being ___km apart* will be represented by a (partial) binary function from $(\Omega^{T})^{2}$ to $\mathbb{R}$ that associates pairs of objects in Ω^{T} with the real number that represents their distance in kilometres. So, for example, if the relation *being 450km apart* holds between Toronto and Ottawa, then the relevant function will assign the real number 450 to <Toronto, Ottawa>. Analogously, the property *being ___m tall* will be represented as a unary function from Ω^{T} to $\mathbb{R}$ that associates each object in Ω^{T} with the real number that represents its height in meters.

[27] Let me note that 'takes' here does not mean 'believes.' A user can take one object to denote another even if she does not believe there is anything the object denotes. I say more about this below.

interpretation 'the corresponding object', 'the corresponding property', or 'the corresponding relation'.

5.4 ANALYTIC VS. NON-ANALYTIC INTERPRETATIONS

Most interpretations of vehicles in terms of targets that we ordinarily adopt can be construed as analytic interpretations. The standard interpretation of the London Underground map in terms of the London Underground network, for example, is an analytic interpretation. First, we take some objects on the map (i.e. small black circles and small coloured tabs) to denote objects on the network (i.e. stations) and, then, we take some of the properties of and relations among those objects on the map to stand for properties of and relations among stations on the network. For example, we take the relation *being connected by a light blue line* on the map to stand for the relation *being connected by Victoria Line trains* on the network. In what follows, I call any epistemic representation whose interpretation is analytic 'an *analytically interpreted epistemic representation*'.

In this book, my focus is exclusively on what I have called analytic interpretations. Unless otherwise specified, by saying that a user adopts an interpretation of the target, I always mean that the interpretation in question is analytic. However, I do not mean to imply that every interpretation of a vehicle in terms of its target is analytic. Epistemic representations whose standard interpretations are not analytic are at least conceivable. For example, suppose we took a mathematical model, such as the equation of motion for the simple pendulum, $(d^2\theta / dt^2)+(g/L)\theta=0$, to be an epistemic representation of some real pendulum. In the context of this epistemic representation, some objects (e.g. numbers) seem to stand for determinate properties of some of the objects that compose the pendulum (e.g. the length of the string from which the pendulum hangs, or the angle of displacement of the pendulum). In this case, our interpretation would not seem to be analytic. As I already mentioned in Chapter 3, however, in this book we assume that so-called mathematical models are not models but descriptions of the behaviour of objects that are part of a fictional model. So, for example, the equation of motion for the simple pendulum describes how the simple pendulum (which is a fictional model from classical mechanics) moves and is the fictional model that is used as an epistemic representation of some real-world system through an analytic representation of it.

In general, however, epistemic representations whose standard interpretations are non-analytic seem to be the exception rather than the rule. In the overwhelming majority of prototypical cases of epistemic representation (which include maps, diagrams, drawings, photographs, and, of course, models), with enough ingenuity, it seems possible to reconstruct the standard interpretation of the vehicle in terms of the target as an analytic one, like I have just done in the case of the simple pendulum.[28] If this is true, then restricting our attention to analytic interpretations will simplify our discussion without any significant loss of generality.

If one is willing to adopt a liberal account of what can count as relevant objects, properties and relations, however, it seems possible to reconstruct any non-analytic interpretation into an analytic one. Unfortunately, I do not have a general argument to prove that any non-analytic interpretation can be reconstructed as an analytic one in this manner (in fact, I am not even sure what such a proof would look like). Whether the accounts of epistemic representation and faithfulness that I develop in this book can be considered perfectly general crucially hinges on this question. If, as I am suggesting, it is possible to reconstruct any interpretation as an analytic interpretation, then the account offered in this book serves as a general account of epistemic representation. Even if this reconstruction is not always possible, however, I think that there are still reasons to hope that the account offered in this book is sufficient to deal with most

[28] I talk of 'reconstruction' because users are often unable to spell out how they interpret the vehicle in terms of the target and are sometimes even unaware that they do interpret the vehicle in terms of the target.

prototypical cases of epistemic representation, and that extending it to cases in which the interpretation adopted is irreducibly non-analytic will only turn out to be technically challenging but not conceptually so.

5.5 Interpretation and Epistemic Representation

In ordinary language, we often use the same word both to refer to an epistemic representation and to refer to the object that serves as the vehicle for that epistemic representation. 'The map', for instance, is sometimes used to refer to the material object that serves as a vehicle of a certain epistemic representation (as in 'The map was faded and tattered') and, at other times, to refer to the epistemic representation itself (as in 'The map was very accurate'). The fact that we use the same word to refer to both, however, should not mislead us into believing that the material object, in and of itself, is an epistemic representation of something else.

According to the interpretational account, the object (material or not) that serves as the vehicle of a certain representation (e.g. the piece of paper that serves as a map), in and of itself, is not an epistemic representation of anything. It is only when a user, more or less consciously, adopts an interpretation of that object in terms of some other object that the former becomes the vehicle of an epistemic representation of the latter for that user.

An epistemic representation can thus be schematically represented as an ordered pair $<v, i^{\circ}(v \rightarrow t)>$ whose first element, v, is the vehicle (i.e. the object that serves as a vehicle of the epistemic representation) and whose second element, $i^{\circ}(v \rightarrow t)$, is an interpretation of (the v-relevant objects, properties, and relations in) the vehicle in terms of (the t-relevant objects, properties, and relations in) the target.

Users are often unaware that they are adopting an interpretation of a vehicle in terms of the target. Even when they are aware that they are in some sense doing so, rarely are they in a position to spell out exactly what their interpretation of the vehicle in terms of the target involves. For example, people are not usually aware that they adopt an interpretation to draw inferences from, say, a photograph to what the photograph represents. They feel that they can just "see things" in the photograph. This intuition, however, does not support the view that they do *not* in fact adopt an interpretation of the photograph—it is simply that they have become so accustomed to adopting this kind of interpretation when interpreting a photograph that the process has become transparent to them. As such, they have come to believe that they can directly "see things" in photographs. It is only when we are less familiar with the standard interpretation of a certain form of epistemic representation (as when we are looking at an infrared photograph or an ultrasound scan), that our need for and our use of an interpretation becomes apparent to us.

One of the reasons why interpretations so often become transparent to us is that, in many cases, the same general interpretation (or family of closely related interpretations) can be used to interpret different vehicles as epistemic representations of different targets when these epistemic representations are representations of the same kind. So, we do not need to come up with or learn a new interpretation every time we come across a new epistemic representation, so long as we are already familiar with the general interpretation associated with that form of epistemic representation. For example, after learning how to interpret a geographic map, we are usually able to use the same (or a very closely related) interpretation for other geographic maps as well. I call such interpretations that are used as the standard interpretation of many representations of the same form '*general interpretations*'.

This feature of interpretation puts the interpretational account in a position to give some substance to the answer to the question raised in Chapter 4 concerning our preferences with respect to the map, the hand, and the saltshaker as candidate vehicles of an epistemic representation of the geography of Michigan. In the case of the map, we do not need to come up with an *ad hoc* interpretation of the vehicle in terms of the target because we are already familiar with a number of ready-made general interpretations that allow us to interpret geographical maps in terms of their targets, and these are likely to include one that will allow us to interpret the map of Michigan in terms of the geography of the state of Michigan. In the case of the upturned right hand, it is quite intuitive for those who are familiar with the geography of Michigan and with the standard interpretation of maps to come up with an interpretation of the upturned right hand in terms of the geography of Michigan (one according to which a point on the palm of the hand denotes a certain location in Michigan, the contour of the hand denotes the contour of the state, etc.). However, there seem to be no immediately intuitive interpretations of the saltshaker in terms of the geography of Michigan (let alone general ready-made interpretations) that would allow us to perform all the sound surrogative inferences about the geography of Michigan that can be performed from the map or the hand. According to the interpretational account of epistemic representation, our preferences can thus be explained in terms of how easily we can interpret the different vehicles in terms of the targets in such a way as to license as many sound surrogative inferences as possible.

5.6 How Does Interpretation Relate to Surrogative Reasoning?

One of my chief complaints against the denotational account and the inferential account (in Chapter 4) was that they do not seem to be capable of explaining (i) how epistemic representation is linked to surrogative reasoning and (ii) what determines which surrogative inferences are valid and which ones are not. The interpretational account, on the other hand, seems to provide excellent answers to both questions. According to the interpretational account, the fact that a user is able to perform surrogative inferences from a vehicle to a target and the fact that not all surrogative inferences from the vehicle to the target are valid stem from the fact that the user adopts an interpretation of the vehicle in terms of the target.

Let me explain. According to the interpretational account, when adopting an interpretation of a vehicle, a user thereby also implicitly adopts a set of rules that determine which surrogative inferences from the vehicle to the target are valid and which ones are not. The rules for an analytic interpretation are as follows:

(Rule 1) If, according to the analytic interpretation of v in terms of t, $i^{\circ}(v \longrightarrow t)$, adopted by u, $o_i^{\mathtt{V}}$ denotes $o_i^{\mathtt{T}}$, then it is valid for u to infer that $o_i^{\mathtt{T}}$ is in t if and only if $o_i^{\mathtt{V}}$ is in v,

(Rule 2) If, according to the analytic interpretation of v in terms of t, $i^{\circ}(v \longrightarrow t)$, adopted by u, $o_1^{\mathtt{V}}$ denotes $o_1^{\mathtt{T}}$, …, $o_n^{\mathtt{V}}$ denotes $o_n^{\mathtt{T}}$, and ${}^{n}R_k^{\mathtt{V}}$ denotes ${}^{n}R_k^{\mathtt{T}}$, then it is valid for u to infer that the relation ${}^{n}R_k^{\mathtt{T}}$ holds among $o_1^{\mathtt{T}}$, …, $o_n^{\mathtt{T}}$ if and only if ${}^{n}R_k^{\mathtt{V}}$ holds among $o_1^{\mathtt{V}}$, …, $o_n^{\mathtt{V}}$,

(Rule 3) If, according to the analytic interpretation of v in terms of t, $i^{\circ}(v \longrightarrow t)$, adopted by u, $o_i^{\mathtt{V}}$ denotes $o_i^{\mathtt{T}}$, $o_1^{\mathtt{V}}$ denotes $o^{\mathtt{T}}{}_1$, …, $o_n^{\mathtt{V}}$ denotes $o^{\mathtt{T}}{}_n$, and ${}^{n}F_k^{\mathtt{V}}$ denotes ${}^{n}F_k^{\mathtt{T}}$, then it is valid for the user to infer that the value of the function ${}^{n}F_k^{\mathtt{T}}$ for the arguments $o_1^{\mathtt{T}}$, …, $o_n^{\mathtt{T}}$ is $o_i^{\mathtt{T}}$ if and only if the value of the function ${}^{n}F_k^{\mathtt{V}}$ for the arguments $o_1^{\mathtt{V}}$, …, $o_n^{\mathtt{V}}$ is $o_i^{\mathtt{V}}$.

To illustrate how the first two rules apply to a concrete example, suppose that a user adopts the standard interpretation of the London Underground map in terms of the network and that she furthermore takes the map to stand for the network. According to *(Rule 1)*, from the fact that there is a circle labelled 'Holborn' on the map, it is valid for her to infer that there is a station called Holborn on the London Underground network and, from the fact that there is no circle or tab labelled 'Bathurst' on the map, it is valid for her infer that there is no station called Bathurst on the London Underground network. According to *(Rule 2)*, from the fact that a coloured line connects a circle labelled 'Holborn' to a tab labelled 'Bethnal Green', one can infer that a direct train service operates between Holborn and Bethnal Green stations. Conversely, from the fact that no coloured line connects the circle labelled 'Holborn' to the tab labelled 'Highbury & Islington', one can infer that no direct train service operates between Holborn and Highbury & Islington stations.

We are now finally in a position to give a definition of validity for epistemic representation whose interpretations are analytic.

(17) If *u* adopts an analytic interpretation of *v* in terms of *t*, $i°(v \rightarrow t)$, then a surrogative inference from *v* to *t* is *valid* (according to $i°(v \rightarrow t)$) if and only if it is in accordance with *(Rule 1)*, *(Rule 2)*, or *(Rule 3)*.

So, if a user is able to perform inferences from a vehicle to a target when the former is an analytically interpreted epistemic representation of the latter, it is because (a) an analytic interpretation of a vehicle in terms of a target subtends a set of rules to draw valid surrogative inferences from the vehicle to the target; and (b) a vehicle is an analytically interpreted epistemic representation of the target only when a user adopts an analytic interpretation of it in terms of the target.

Before concluding this section, let me note that an analytic interpretation subtends a set of rules in the sense that the adoption of a certain set of rules is part and parcel of the adoption of the underlying analytic interpretation—one cannot adopt an interpretation without, thereby, adopting the rules that determine what surrogative inferences from the vehicle to the target are valid given that interpretation. So, for example, it is part and parcel of what it means to take the circle labelled 'Holborn' on the London Underground map to denote Holborn station that, in doing so, one takes the map to be representing the London Underground network as containing a station named 'Holborn.'

5.7 Denotation and Epistemic Representation

The notion of denotation plays a crucial role in all of the accounts of epistemic representation we have examined. All of them agree that in order for a vehicle to be an epistemic representation of a target for a certain user, the user has to take the vehicle to denote (stand for) the target. On the interpretational account, however, denotation plays a second crucial role. According to the interpretational account, it is not sufficient for a user to take the vehicle as a whole to stand for the target; the user must also take some of the objects, properties, and relations that are part of the vehicle to denote some objects, properties, and relations that are part of the target. Only then does the user adopt an interpretation of the vehicle in terms of the target.

What is crucial to note, however, is that this claim is a very far cry from claiming that epistemic representation *amounts to* denotation, as the advocate of the denotational account would maintain. Consider for example the sentence token.

The Tube is the oldest subway system in the world

While we take the first two sets of ink marks (i.e. those that look like 'The Tube') to denote the London Underground network, that is not enough to make them an epistemic representation of the network. According to the denotational account, this is because, even if we take those ink marks to denote the London Underground network, we have not adopted any interpretation of those marks in terms of the London Underground network. If, in addition to taking those marks to stand for the London Underground network, we also took some of those marks, or some of their properties or relations, to stand for some of the objects, properties, or relations that are part of the network (if, for example, we took each letter to stand for a station in the network and the order of the letters to denote certain relations among the corresponding stations), however, then those marks would in fact constitute an epistemic representation of the network. So, in order to have an epistemic representation, it is not sufficient that we take the vehicle to stand for the target; we also need to take some of the components of the vehicle to stand for some of the components of the target.

Before concluding this section, let me note that, in order for a user to take an object in the vehicle to denote a (putative) object in the target, a user does not need to believe that the (supposedly) denoted object exists. For example, on old 1930s London Underground maps, one can find a tab labelled 'Dover Street'. According to the standard interpretation of the map in terms of the network, that tab denotes a station whose name is 'Dover Street'. However, since there is no longer a Dover Street station on today's London Underground network, that tab on the old map fails to denote any station today. Even if I know this to be the case, however, I can still take the tab labelled 'Dover Street' to denote the Dover Street station *according to the standard interpretation of the network that I am using*. In other words, if I adopt the standard interpretation of the map in terms of the network, I can take every circle and every tab on the map to denote a (putative) station even if I know that some circles and tabs fail to denote a station. (This is somewhat analogous to the case in which I take the name 'Santa Claus' to denote Santa Claus *according to a child who believes in the existence of Santa Claus* even if, personally, I do not take that name to denote any real person).

It is therefore important to distinguish between what the user believes about the target and what the epistemic representation itself "tells" her about the target. The latter is what I have previously called the representational content of the map (under its standard interpretation). In some cases, the user's beliefs about the target might conflict with what the epistemic representation "tells" her about the target. In such cases, the user has three options: (i) she can change her beliefs about the target; (ii) she can take the epistemic representation of the target to be only a partially faithful epistemic representation of the network; or (iii) she can adopt a different interpretation of the network. Which of these options the user chooses depends on a number of factors, including how confident she is in her beliefs about the target, in the reliability of the epistemic representation, and in the appropriateness of the interpretation she has adopted.

5.8 Models and Interpretation

According to the interpretational account of epistemic representation, a model is an epistemic representation of some real-world system for a user in virtue of the fact that (1) the user takes the model to denote the system and (2) the user adopts an interpretation of the model in terms of the system. We often go through the process of interpreting a model in terms of a certain system so effortlessly that we do not even realize that we are at all interpreting the model in terms of the system. For example, in the case of the inclined plane model and the tobogganing hill system, we take a simple model from classical mechanics and use it as an epistemic representation of a certain system by interpreting it in terms of that system. As I believe is usually the case with fictional models, our interpretation in this example is analytic—i.e. we take the inclined plane model to denote the system formed by my daughters, their toboggan, and the tobogganing hill; we take the box in the model to denote my daughters on their toboggan; we take the inclined plane to denote the hill; we take some relevant properties of the plane (e.g. its being frictionless or its exerting a force on the box) to stand for some properties of the hill; and, finally, we take some relevant properties of the box such as its position or its velocity to denote the position and velocity of my daughters on their toboggan.[29]

[29] Here I talk of the position of the box at a certain time and that of the racer as a property to keep in line with the language ordinarily used by philosophers. However, it is more convenient to regard certain properties such as the position and the velocity of the box at a certain time as functions rather than properties. So the velocity function is a function that associates with a certain object its velocity at a certain time.

Now, as we have already seen in the case of the map of the London Underground, an interpretation of a vehicle in terms of a target subtends a set of rules that license certain (valid) surrogative inferences from the vehicle to the target. So, whenever a user adopts an interpretation of a model in terms of a real-world system, the user will be able to perform (valid) surrogative inferences from the vehicle to the target. In the case of the inclined plane model, for example, according to the above interpretation, from the fact that final velocity of the box in the model is about 50 km/h, one can validly infer that, according to the model (on its standard interpretation), the final velocity of my daughters on their toboggan would be about 50 km/h. Of course, I do not need to believe this to be exactly the case in order to draw that (likely unsound) surrogative inference from the model to the system. In this case, as in other cases, we need to distinguish carefully between what the models "tells" its user (which is what I have called the representational content of the model) and what the user herself believes. From a model, I can draw all sorts of surrogative inferences that I take to be false or inaccurate, and yet these inferences are still valid to draw given how the model is interpreted in terms of the system.

Consider a more extreme case. According to the Ptolemaic model of the cosmos, the universe is a system of concentric crystal spheres. The Earth lies at the centre of the sublunary region, which is the innermost sphere. Outside the sublunary region are the heavens: eight tightly fit spherical shells, the outermost of which, the sphere of the fixed stars, hosts the stars. Each of the other spherical shells hosts one of the seven "planets," which, in the Ptolemaic model, include the Moon and the Sun. Each spherical shell rotates around its centre with uniform velocity. According to the standard interpretation of the model, the crystal spheres in the model denote the crystal spheres that supposedly host the planet in the system. However, one does not need to believe that there actually are crystal spheres that host the planets and stars in order to adopt an interpretation of the model according to which crystal spheres host the planets. A user of the model may firmly believe that planets are not hosted by crystal spheres and yet believe that what the model "tells" her is that the planets are hosted by crystal spheres. So, for example, an early 16th century user and a contemporary user of the Ptolemaic model do not need to adopt different interpretations of it (and consequently do not need to disagree as to which surrogative inferences from the model to the solar system are valid); rather, they may only disagree in their epistemic attitudes toward the conclusions of some of those inferences.

Analogous remarks apply also to scientific realists and anti-realists (or at least those scientific anti-realist who construe theories literally). Scientific realists and anti-realists do not need to adopt different interpretations of the same model, and consequently do not need to disagree as to which surrogative inferences from the model to the solar system are valid; rather, they may only disagree in their epistemic attitudes toward the conclusions of some of those inferences. However, some scientific anti-realists might be inclined to adopt non-standard interpretations of scientific models. For example, an instrumentalist might be tempted to adopt an interpretation according to which the only aspects of models that stand for something are those that stand for observable or measurable features of the target system.

5.9 Representation on the Cheap?

Before concluding this chapter, I would like to consider a potential worry about the interpretational account. The worry is that, on the interpretational account, epistemic representation comes too cheaply. After all, nothing seems to prevent me from adopting an interpretation of a ripe tomato in terms of the system formed by my daughters tobogganing down the hill, one according to which, say, the deeper the red of the tomato, the faster the toboggan will go. To some, this may seem to be a troubling conclusion. But what exactly is wrong with it?

The objection might be that from the tomato I would likely infer only false conclusions about the system. This may well be the case, but the interpretational account is meant to be an account of what makes a vehicle an epistemic representation of a certain target, not an account of what makes it a more or less *faithful* epistemic representation of the target. Further conditions need to be in place for the tomato to be a *faithful* epistemic representation of the system, conditions which the tomato would be unlikely to meet.

Maybe the objection is that, if *denotation* and *interpretation* are all that is needed for *epistemic representation*, then using models for prediction is much like using tarot cards. Of course, there is an enormous difference between using a good model and using tarot cards to discover whether my daughters will be safe on their toboggan—but the difference is *not* due to the fact that tarot cards are not an epistemic representation of the situation (after all, tarot cards are used to perform surrogative inferences about other things). Rather, the difference is likely that the good model provides me with what (one would hope) is a much more faithful representation of the situation than the tarot cards, and that (again, one would hope) I have good reason to think that this is so.

If epistemic representation seems to come too cheaply on the interpretational account, it may be because, after all, epistemic representation *is* cheap. It doesn't take much for someone to be able to perform surrogative inferences from one thing to another (although, at the same time, it does seem to take more than the denotational account and inferential account suggest). What does *not* come cheaply, however, is *faithful* epistemic representation. It is to accounts of faithfulness that I turn my attention in the last part of this book.

5.10 Conclusions

In this chapter, I have developed and defended an interpretational account of epistemic representation. According to it, in order for a vehicle to be an epistemic representation of a target (for a user), the user has to take the vehicle to denote the target and adopt an interpretation of the vehicle in terms of the target. That interpretation subtends a set of rules that determine which surrogative inferences from the vehicle to the target are valid and which ones are not. However, the fact that a certain inference is *valid*—i.e. that it is in accordance with the set of rules subtended by the interpretation—does not imply or guarantee that the inference is *sound*—i.e. that its conclusion is true (or even just approximately true). If a vehicle is to be a somewhat faithful epistemic representation of a certain target some further conditions need to hold beside the ones that make the vehicle into an epistemic representation of the target. Part III of this book is devoted to developing an account of what makes an epistemic representation of a certain target more or less faithful.

Part III: Faithfulness

6 The Similarity Account

6.1 Introduction

One of the theses underlying this book is that the so-called "problem of scientific representation" is neither a single problem nor specifically a problem about *scientific* representation, and that the handful of solutions to the so-called problem that are often discussed in the literature are in fact attempts to answer two distinct, general questions, namely:

(a) In virtue of what is a vehicle an epistemic representation of a target (for some user)?
(b) In virtue of what is a vehicle a faithful epistemic representation of a target (for some user)?

In Part II, I have considered three possible answers to question (a) and defended one of them, which I called the interpretational account of epistemic representation. Now, in this chapter and the next, I consider two possible answers to question (b), which I call 'the similarity account (of faithful epistemic representation)' and 'the structural account (of faithful epistemic representation)'. Although both of these accounts seem to have many supporters among philosophers of science, one is hard pressed to find a clear formulation of either account in the literature. This has given rise to many misunderstandings as to what each account exactly claims. The thesis underlying this and the following chapter is that, contrary to what is often assumed, the similarity account and the structural account are best understood as accounts of faithful epistemic representation (as opposed to accounts of epistemic representation *simpliciter*) and that these two accounts do not offer incompatible answers to question (b) above. In fact, in Chapter 8, I argue that the best account of faithful epistemic representation combines aspects of both the similarity account and the structural account. Before doing so, however, in this chapter, I briefly discuss the similarity account and some of the standard objections against it. In particular, I argue that, although many of the objections that target the similarity account are mistargeted, there is some truth to the accusation that similarity might be too vague a notion on which to ground the notion of faithfulness.

6.2 Similarity and Representation

One of our most deep-seated intuitions about the notion of representation is that representation is closely related to similarity. Pre-theoretically, we seem inclined to think that there is a direct relation between representation and similarity, and that, for example, a portrait represents its subject in virtue of the similarity between the two. As tempting as this view may at first seem, however, it does not withstand scrutiny. Since the arguments against this view are well-known, I only briefly mention a few here.[30] First of all, similarity and representation do not share all of the same "logical" properties—similarity is both a reflexive and symmetric relation, but, in most ordinary cases, representation is neither reflexive nor symmetric. Although a map may be similar to itself, it does not seem to represent itself. Rather, it represents some city. If the map is similar to the city, then the city is also similar to the map. However, although the map is a

[30] (Goodman 1968) is perhaps the best-known source of arguments against the naïve account of representation that links it directly with similarity.

representation of the city, the city is not usually taken to be a representation of the map.[31] Furthermore, similarity seems to be neither necessary nor sufficient for representation. If similarity were sufficient for representation, then how could one take a picture of one of a pair of identical twins that is not equally a representation of the other twin? If similarity were necessary for representation, how would it be possible for an unskilled artist to paint a portrait of a person that doesn't look at all like that person?

While the above arguments are far from conclusive, taken together they seem to suggest that the connection between similarity and representation is not as direct as we might initially assume. In particular, once we distinguish between epistemic representation and faithful epistemic representation, it becomes apparent that what the above arguments show is that similarity is neither necessary nor sufficient for epistemic representation. However, those arguments do not preclude similarity for being necessary for *faithful* epistemic representation. On this interpretation, the supporters of the similarity account maintain that, *if* a certain vehicle is an epistemic representation of a certain target, *then* the vehicle is a faithful epistemic representation of that target only insofar as it is similar to the target. Or, to put it more precisely, *the similarity account of faithful epistemic representation* states that:

(D) *v* is a *completely faithful epistemic representation* of *t* (for *u*) if and only if:

(D.1) *v* is an epistemic representations of *t* (for *u*), *and*

(D.2) *v* is perfectly similar to *t* in all relevant respects.

(E) *v* is (overall) a (strictly) *more faithful epistemic representation* of *t* than *v** if and only if:

(E.1) *v* and *v** are epistemic representations of *t* (for *u*),

(E.2) *v* is more similar to *t* than *v** in some relevant respect(s) and to a relevant degree, *and*

(E.3) *v** is no more similar to *t* than *v* in any (relevant) respect to any relevant degree.

Note that in claiming that the similarity account is an account of account of faithfulness (and not an account of epistemic representation, as is often assumed (see, e.g., (Suárez 2003)) I am not claiming that the supporters of the similarity account *intend* the similarity account as an account of what I call 'faithfulness'. To my knowledge, the supporters of the similarity account do not even distinguish between epistemic representation and faithful epistemic representation. Rather, my claim is that, once we draw the distinction between epistemic representation and faithful epistemic representation, it is natural to interpret the similarity account as an account of *faithful* epistemic representation and not as an account of epistemic representation *simpliciter*.

One of the consequences of interpreting the similarity account in this way is that the account does not constitute an *alternative* to the accounts of epistemic representation that I discussed in Part II. To the contrary, the similarity account is *complementary* to those accounts. This is because, according to the similarity account of faithful epistemic representation, the fact that a vehicle is an epistemic representation of a certain target is a necessary but not sufficient condition for it to be a *faithful* epistemic representation of that target and, therefore, the supporters of the similarity account need to supplement their account of faithful epistemic representation with an account of epistemic representation *simpliciter*.

As I interpret it, the similarity account attempts to identify which further condition(s) (i.e. which condition(s) beside the ones that make the vehicle an epistemic representation of the target) need to obtain in order for the vehicle to be a (more or less) *faithful* epistemic

[31] This of course does not mean that a city could not represent a map, but only that similarity does not seem to be sufficient for representation.

representation of the target. The similarity account of faithful epistemic representation, therefore, *does not* claim that, say, the inclined plane model is an epistemic representation of my daughters going downhill on their toboggan in virtue of the fact that the model is similar to the system (in certain respects and to a certain degree). What it *does* claim is that, *if* the inclined plane model is an epistemic representation of that system, *then* it is a *faithful* epistemic representation of that system to the extent to which it is similar to the system (in the relevant respects and to the relevant degree).

Once we adopt this interpretation of the similarity account, it becomes apparent that many of the standard objections that have been directed against it are wide of the mark. Consider, for example, the arguments Suárez (2004) adduces against the account of representation he labels '[sim].' According to [sim], a vehicle is an epistemic representation of a target if and only if it is similar to the target. The first argument against [sim], which Suárez dubs '*the logical argument*', is one we are already familiar with: while similarity is reflexive and symmetric, representation is usually neither reflexive nor symmetric. Unlike [sim], however the similarity account of faithful epistemic representation does not claim that, say, a map is an epistemic representation of a city in virtue of the similarity between the two. What it claims is that, *if* the map is an epistemic representation of the city, *then* the map is a *faithful* epistemic representation of the city to the extent to which the two are similar (in the relevant respects and to the relevant degree). And, since epistemic representation is not necessarily reflexive or symmetric, the fact that the further condition for faithful epistemic representation is both reflexive and symmetric does not make faithful epistemic representation reflexive or symmetric.

The similarity account of faithful epistemic representation also avoids Suárez's second argument against [sim]—*the argument from mistargeting*. According to that argument, [sim] cannot account for those cases in which a user mistakenly believes one object to be the target of a certain representation, when, in fact, the actual target is a different object. I may take a very poorly executed portrait of Queen Elizabeth II to portray Elton John, but, no matter how closely the portrait resembles Elton John, it is still a portrait of the Queen. According to the similarity account of faithful epistemic representation, however, the similarity between Elton John and the portrait is neither necessary nor sufficient for the portrait to be an epistemic representation of Elton John. Rather, the similarity account of faithful epistemic representation claims that, *if the portrait were an epistemic representation of Elton John*, it would be a faithful epistemic representation of him to the extent to which it was similar to him (in certain respects and to a certain degree). However, since the portrait is not, in fact, a portrait *of* Elton John, no amount of similarity between it and Elton John can turn it into a representation of Elton John according to the similarity account of faithful epistemic representation.

Suárez's third and the fourth arguments against [sim], *the non-necessity argument* and *the non-sufficiency argument*, claim that similarity is neither a necessary nor a sufficient condition for representation. As we have already seen, something can be a representation of something else without being similar to it and something can be similar to something else without being a representation of it. However, the similarity account of faithful epistemic representation does not claim that similarity is either necessary or sufficient for epistemic representation. According to the similarity account of faithful epistemic representation, similarity is a further condition that needs to obtain in addition to the conditions that make a vehicle an epistemic representation of the target if the vehicle is to be a *faithful* epistemic representation of it.

I think that the above arguments show that, properly interpreted, the similarity account avoids many of the standard objections directed against it. In the next two sections, however, I consider two further objections to the similarity account, which I take to be more serious. The first is that the notion of similarity is too vague; the second is that, in some cases, it is not clear in what sense the vehicle and the target are similar. While I do not think that either objection is decisive on its own, taken together they suggest that the intuitions that underlie the similarity account might need to be formulated more clearly.

6.3 Similarity and Vagueness

Philosophers are often suspicious of the notion of similarity. Their suspicion stems from the widespread belief that the notion of similarity is too vague to perform any genuine philosophical work. Many of the critics of the similarity account maintain that its appeal to similarity renders the account vacuous because anything is similar to anything else (see, e.g., Suárez 2003), or that similarity is little more than a blank to be filled in on a case-by-case basis (see, e.g., Frigg 2002). If the similarity account is to be taken seriously, its supporters should, therefore, try to dispel the impression that its reliance on the notion of similarity renders it vacuous or uninformative.

I suggest that supporters of the similarity account adopt a two-step strategy. The first step is to note that the similarity account does not make use of the notion of *overall similarity* but only of the notion of *aspectual similarity*. According to the similarity account, the vehicle and the target do not need to be similar *overall*; they only need to be similar *in certain respects* and *to a certain degree*. The second step is to claim that, in most ordinary contexts, we seem to be fairly good at making aspectual similarity judgements (such as 'the two sweaters are very similar in colour' or 'the two players are similar in height') and that, while we often have minor disagreements about aspectual similarity judgements in specific contexts, major disagreements seem to be rare.

If fully developed, this strategy might be able resolve both *the vacuity worry* and *the informativeness worry* in one fell swoop. Consider the vacuity worry first. Even if it were true that virtually everything is similar to anything else *in some respect* and *to some degree* (and I doubt that's the case), clearly it is not the case that everything is similar to anything else *in a specific respect* and *to a specific degree*. Even if there is a sense in which this book is similar to my laptop in some respects and to some degree (e.g. they are both mid-sized material objects, they both have a roughly flat parallelepipedal shape when closed, and they both sit often on my office desk), they are not similar with respect to, say, their colour (one is red and the other one is silver) or the materials they are made of (pulp and adhesive versus various metals and plastics, respectively).

Consider now the informativeness worry. Even if it were true that relevant respects and degrees of similarity have to be determined on a case-by-case basis, this does not seem to make an appeal to the notion of similarity necessarily uninformative, as, in most ordinary cases, the context determines the relevant aspects and degrees of similarity. For example, if I am about to buy a sweater and you advise me not to buy it because I already own many similar sweaters, what you likely mean is that I own many sweaters that are similar to it in colour, warmth, or cut, not that I own many sweaters that are similar to it in being located on planet Earth, being poor conductors of electricity, or being self-identical. While context might not fully and unequivocally determine what the relevant respects and the degrees of similarity are, it usually narrows the range of possibilities down significantly. The supporters of the similarity account can therefore concede that the term 'similar' is to some extent vague and context-dependent while maintaining that its vagueness and context-dependence are no more serious than those that characterize many other terms in ordinary language. Given a certain context, there are some clear cases of objects that are similar to each other (in certain respects and to a certain degree) and clear cases of objects that are dissimilar (in certain respects and to a certain degree) just like, given a certain context, there are clear cases of say, people who are tall and clear cases of people who are not tall.

While the above only sketches a response, and more needs to be said in order to adequately address worries about similarity, this approach seems promising. In the next section, however, I argue that, while these general worries about similarity push us towards an emphasis on aspectual similarity, there is a more specific kind of worry that pushes us in a different direction altogether.

6.4 Elusive Similarities

Even if the strategy sketched in the previous section, once fully implemented, were to lay to rest worries about the vagueness of the notion of similarity, one might still harbour another more specific worry. In many cases it is not entirely clear in what respects the vehicle of an epistemic representation is similar to its target. I call this '*the elusiveness worry*'. In this section, I first consider the elusiveness worry as it applies to models and sketch a possible way to deal with this specific manifestation of it; I then turn to what I consider a more serious manifestation of the same sort of worry. It will turn out that there is some tension between the interpretation of the notion of similarity that is needed to placate the elusiveness worry and the one that might put to rest the vacuity worry and the informativeness worry.

Consider first the elusiveness worry as it applies to models. R.I.G. Hughes, for example, claims:

> […] we may model an actual pendulum, a weight hanging by a cord, as an ideal pendulum. We may be even tempted to say that in both cases the relation between the pendulum's length and its periodic time is approximately the same, and that they are in that respect similar to each other. But *the ideal pendulum has no length, and there is no time in which it completes an oscillation* (Hughes 1999, p.S330; emphasis mine).

I take it that, according to Hughes, only material objects—i.e. objects that are both actual and concrete—can have concrete properties such as having a certain length or completing an oscillation in a certain time; and, since, whatever the ideal pendulum is, it is not a material object, it cannot have a length and cannot oscillate. Therefore, it is not clear in what sense the ideal pendulum could possibly be similar to an actual pendulum, which oscillates and has a certain length.[32]

The advocates of the similarity account, however, do not need to maintain that the ideal pendulum *literally* has the concrete properties it is ordinarily said to have in order to claim that it is similar to some real-world pendulum. Consider the argument that I raised in Chapter 3 from the analogy to fictional entities. Whatever fictional entities may be, they are not concrete actual objects. Yet they are often said to have properties that only concrete objects have, such as having a wart on the nose. Obviously, when people say things like 'The Gruffalo has a wart on his nose' in a meta-fictional context (i.e. when talking about a work of fiction), they do not mean that the Gurffalo *literally* has a wart on his nose. Rather, they mean that, according to the book *The Gruffalo*, the Gruffalo has a wart on his nose.

The fact that fictional characters do not literally have the properties that are ordinarily ascribed to them does not seem to prevent us from comparing them to concrete actual objects. When discussing literary works, for example, people often compare fictional entities to actual concrete ones. Readers, for example, may discuss how closely the London of Dickens' novels resembles the actual Victorian London, or whether the historical Richard III was as ruthless as the homonymous character in the Shakespearian play. These similarity comparisons seem perfectly legitimate and do not seem to presuppose that the fictional entities whose characteristics are being compared with those of actual concrete ones are themselves actual concrete objects that literally have the concrete properties ascribed to them. It would miss the point of the exercise entirely to object that the personalities of the historical Richard III and that of its fictional counterpart cannot be compared because the latter is not an actual concrete person, and as such does not have a personality. Those who draw the comparison do not seem to assume (nor need they assume) that the latter *literally* has a personality.

In saying this, I do not mean to play down the difficulty of explaining how it is that an object that is neither actual nor concrete can, in some sense, be truly said to have concrete properties;

[32] Similar objections can also be found in (Suárez 2002) and (Callender and Cohen 2006).

as we have seen in Chapter 3, this is no trivial philosophical problem. I only mean to show that the problem of explaining how non-concrete models can be similar to concrete systems is not a novel philosophical problem and that any adequate account of the nature of fictional entities should provide us with a solution to it. If we construe non-concrete models (such as the ideal pendulum or the inclined plane) as fictional entities (as I have argued in Chapter 3 that we might), then there seems to be a sense in which models can be "truly" said to have the concrete properties that we ascribe to them (such as having a certain length or period of oscillation) – even if, in fact, they do not literally have any of these properties. As far as I can see, this is all that we need to ground our similarity judgements between models and real-world systems.[33]

Although this specific manifestation of the elusiveness worry might be resolved in the manner I have just suggested, other manifestations of that same worry seem more recalcitrant to treatment. Consider, for example, the London Underground map and the London Underground network. According to the similarity account, if the map is a (completely) faithful epistemic representation of the network, then the map and the network should be perfectly similar in all relevant respects. But in what respects are the map and the network similar? The map is a piece of glossy paper with coloured lines printed on it, while the network is a complex system of trains, tracks, tunnels, platforms, and escalators. On the face of it, the two objects could hardly be more dissimilar from one other. So, in what respects can they possibly be perfectly similar, according to the similarity account?

The advocate of the similarity account might be tempted to answer that the map is similar to the network in that the circles and tabs on the map are connected by a coloured line and the stations denoted by those circles and tabs are connected by a direct train service. However, this line of defence trades on the fact that we happen to use the same abstract verb ('connect') to refer to two relations that do not seem to actually be similar at all. When we say that a train line connects two stations, what we mean is, roughly, that certain trains stop at both of those stations. But this is clearly not what we mean when we say that a line connects two points.

As far as I can see, the most (and possibly the only) plausible answer to the above question is that, if there is any similarity between the map and the network, it is a very global and abstract sort of similarity—it is what in Chapter 8 I call 'a structural similarity'. The problem with this kind of answer, however, is that, unless it is coupled with a clear definition of the notion of structural similarity, it seems to revitalise the vacuity worry and informativeness worry. However structural similarity might be defined, it does not seem to be the sort of similarity we appeal to in ordinary contexts when we appeal to aspectual similarity. In Chapter 8, I provide a clear definition of the notion of structural similarity. In order to do so, however, I first need to discuss other structural accounts of faithful epistemic representation and explain why they are not entirely satisfactory either.

---

[33] The critics of the similarity account, however, may think that this is not sufficient to solve the problem. Whereas real-world pendula have *definite* oscillation times and real strings have *definite* lengths, the ideal pendulum has an *indefinite* oscillation time T, and its string has an *indefinite* length L. So, how can the length of the string in a real-world pendulum be similar to the length of the string of the ideal pendulum, if the string of the ideal pendulum has no definite length? To ease this worry, it should be noted that, first of all, even if the ideal pendulum has no definite length and no definite oscillation time, it has *a* length and *an* oscillation time and the relation between them is a *definite* relation. Secondly, nothing prevents us from setting the values of any of the parameters of the ideal pendulum that have an indefinite value to definite ones. This is what someone might do in order to use the ideal pendulum model to represent some specific pendulum. She can, for example, set the parameter L so the string in the ideal pendulum is exactly 30 centimetres long. In the specified ideal pendulum, the string would thus have a *definite* length, like any real string. It is important to note that, usually, by setting some of the parameters of a model, one also thereby fixes other parameters. In the case of the ideal pendulum, for example, by setting the values of the length of the string, L, and the gravitational acceleration on the bob, g, one also indirectly fixes the period of the pendulum, which is equal to $2\pi(L/g)^{1/2}$.

6.5 Conclusion

In this chapter, I sketched what I have called the similarity account of faithful epistemic representation and I argued that it avoids many of the objections that are usually levelled at naïve similarity accounts of representation. According to the similarity account of faithful epistemic representation, the fact that the vehicle and the target are similar (in certain respects and to certain degree) is a necessary but not sufficient condition for the vehicle to be a faithful epistemic representation of the target. So, for example, the similarity account of faithful epistemic representation does not maintain that the portrait is an epistemic representation of its subject in virtue of their similarity. What it does claim is that, *if* the portrait is an epistemic representation of the subject, *then* it represents her faithfully to the extent to which the portrait and the subject are similar to each other in specified respects and to a specified degree.

This account of faithful epistemic representation seems plausible when applied to certain prototypical cases of epistemic representation, such as the aforementioned case of the portrait. First, in that case, our intuitions about the similarity between the portrait and its subject seem reasonably clear. If we show someone a well-executed portrait of someone they know, they will usually be able to tell us whether or not it "resembles" that person. Second, in those cases, the advocate of the similarity account is able to tell a plausible story about why, if the portrait is a faithful epistemic representation of its subject, its users are able to perform sound inferences from the portrait to its subject. The story will go somewhat as follows. A portrait is a faithful epistemic representation of a certain subject only if it is similar to the subject (in certain respects and to a certain degree). For example, the portrait represents its target faithfully only if the colour of the patches of paint that stand for, say, the irises of the subject is similar to a certain degree to the actual colour of the subject's irises. Then, the user can infer (in accordance with the standard interpretation of conventional portraits) that the irises of the subject are of the same colour as the patches of paint that denote them. If the colour of the patches of paint is indeed similar to the colour of the subject's eyes, the user's conclusion will be true (or at least approximately true).

The similarity account, however, does not seem to do equally well in other cases of epistemic representation. For example, as I argued in the last section, it is far from obvious in which respects and to what degree, if any, the London Underground map is similar to the London Underground network. This is not to say that there is no sense in which they are similar. Rather, it is to say that the two objects do not seem to be similar in the obvious, ordinary sense of 'similar'. This seems to create a dilemma for the supporters of the similarity account. If, as I have argued, the best strategy available to placate the vacuity worry and the informativeness worry is to appeal to our ordinary intuitions about aspectual similarities between pairs of objects, then it is difficult to see how that sort of similarity can subsist between aspects of the London Underground map and aspects of the London Underground network. If, on the other hand, supporters of the similarity account appeal to the presence of a more global and abstract kind of similarity between the map and the network, then our ordinary intuitions about similarity seem simply to fail us. This abstract and global sort of similarity is what I call 'structural similarity'.

However, as far as I can see, there are no conclusive objections against the similarity account of faithful epistemic representation. In fact, I believe that there is more than a grain of truth to it and I take it that the structural account, which I consider in Chapter 7, and the structural similarity account, which I develop in Chapter 8, are, in a sense, just a subspecies of the similarity account of faithful epistemic representation—they are similarity accounts that focus on the similarity between "the structures" of the vehicle and the target. Unlike the standard similarity account, however, the structural account provides us with a much more solid and well-defined general framework for understanding faithful epistemic representation and allows us to articulate clearly and precisely the intuition that similarity is related to (faithful) epistemic representation. It is to this account that I turn in the next chapter.

7 The Structural Account

7.1 Introduction

In this chapter, I examine what I call 'the structural account of faithful epistemic representation'. While the structural account seems to have many sympathizers among philosophers of science (including Patrick Suppes, Bas van Fraassen, Steven French, and James Ladyman), one is hard pressed to find a clear and detailed formulation of what such an account holds. By a clear and detailed formulation, I mean the set of conditions under which a vehicle counts as an epistemic representation (or a faithful epistemic representation) of a target. One is similarly hard pressed to find in the literature any worked out example of how a structural account of representation might be applied to concrete cases of epistemic representation. My goal in this chapter is to clarify what a structural account of faithful epistemic representation would look like. Rather than trying to interpret what the sympathizers of the structural account might mean when talking about representation, I try to develop what I take to be the strongest version of the structural account of faithful epistemic representation and illustrate how it applies to concrete cases of epistemic representation.

As I understand it, the structural account of faithful epistemic representation aims at accounting for the faithfulness of a representation in terms of a formal relation, a *morphism*, between two set-theoretic structures.

(18) A *set-theoretic structure* is an triple $\mathsf{S} = <\Omega^{\mathsf{S}}, \mathbf{P}^{\mathsf{S}}, \Phi^{\mathsf{S}}>$, where:

[18.1] Ω^{S} is a non-empty set of objects (also called the *universe* of S and sometimes denoted as $|\mathsf{S}|$),

[18.2] $\mathbf{P}^{\mathsf{S}}=\{{}^{M}R_1^{\mathsf{S}}, \ldots, {}^{v}R_j^{\mathsf{S}}\}$, where ${}^{M}R_1^{\mathsf{S}}, \ldots$, and ${}^{v}R_j^{\mathsf{S}}$ are *relations* on Ω^{S} (where the left-hand superscript indicates the number of places of the relation with properties are construed as unary relations), *and*

[18.3] $\Phi^{\mathsf{S}}=\{{}^{p}F_1^{\mathsf{S}}, \ldots, {}^{r}F_k^{\mathsf{S}}\}$, where ${}^{p}F_1^{\mathsf{S}}, \ldots$, and ${}^{r}F_k^{\mathsf{S}}$ are *functions* from *n*-tuples of members of Ω^{S} to elements of Ω^{S} ($(\Omega^{\mathsf{S}})^n \rightarrow \Omega^{\mathsf{S}}$) (where the left-hand superscript indicates the number of arguments of the function).

(For the sake of simplicity, I drop all subscripts and superscripts whenever the context allows).

Two of the best-known morphisms are isomorphism and homomorphism.

(19) A function, *f*, from the universe of A to the universe of B (f: $\Omega^{\mathsf{A}} \rightarrow \Omega^{\mathsf{B}}$) is a *homomorphism* if and only if:

[19.1] for every $o_i^{\mathsf{A}} \in \Omega^{\mathsf{A}}$, there is an $o_i^{\mathsf{B}} \in \Omega^{\mathsf{B}}$ such that $f(o_i^{\mathsf{A}})=o_i^{\mathsf{B}}$,

[19.2] for every $o_i^{\mathsf{A}} \in \Omega^{\mathsf{A}}$, if $f(o_i^{\mathsf{A}})=o_i^{\mathsf{B}}$ and $f(o_i^{\mathsf{A}})=o_k^{\mathsf{B}}$, then $o_i^{\mathsf{B}}=o_k^{\mathsf{B}}$,

[19.3] for every $R_i^{\mathsf{A}} \in \mathbf{P}^{\mathsf{A}}$, if $<o_1^{\mathsf{A}}, \ldots, o_k^{\mathsf{A}}> \in R_i^{\mathsf{A}}$, then $<f(o_1^{\mathsf{A}}), \ldots, f(o_k^{\mathsf{A}})> \in R_i^{\mathsf{B}}$,

[19.4] for every $F_j^{\mathsf{A}} \in \Phi^{\mathsf{A}}$, $f(F_j^{\mathsf{A}}(o_1^{\mathsf{A}}, \ldots, o_p^{\mathsf{A}}))=F_j^{\mathsf{B}}(f(o_1^{\mathsf{A}}), \ldots, f(o_p^{\mathsf{A}}))$.

(20) A function, f, from the universe of $\mathbf{A}$ to the universe of B (f: $\Omega^{\mathbf{A}} \rightarrow \Omega^{\mathrm{B}}$) is an *isomorphism* if and only if:

[20.1] for every $o_i^{\mathbf{A}} \in \Omega^{\mathbf{A}}$, there is an $o_i^{\mathrm{B}} \in \Omega^{\mathrm{B}}$ such that $f(o_i^{\mathbf{A}})=o_i^{\mathrm{B}}$,

[20.2] for every $o_i^{\mathrm{B}} \in \Omega^{\mathrm{B}}$, there is an $o_i^{\mathbf{A}} \in \Omega^{\mathbf{A}}$ such that $f(o_i^{\mathbf{A}})=o_i^{\mathrm{B}}$,

[20.3] for every $o_i^{\mathbf{A}} \in \Omega^{\mathbf{A}}$, if $f(o_i^{\mathbf{A}})=o_i^{\mathrm{B}}$ and $f(o_i^{\mathbf{A}})=o_k^{\mathrm{B}}$, then $o_i^{\mathrm{B}}=o_k^{\mathrm{B}}$

[20.4] for every $o_i^{\mathrm{B}} \in \Omega^{\mathrm{B}}$, if $f(o_i^{\mathbf{A}})=o_i^{\mathrm{B}}$ and $f(o_k^{\mathbf{A}})=o_i^{\mathrm{B}}$, then $o_i^{\mathbf{A}}=o_k^{\mathbf{A}}$,

[20.5] for every $R_i^{\mathbf{A}} \in \mathbf{P}^{\mathbf{A}}$ and $R_i^{\mathbf{A}} \in \mathbf{P}^{\mathrm{B}}$, $<o_1^{\mathbf{A}}, \ldots, o_i^{\mathbf{A}}> \in R_i^{\mathbf{A}}$ if and only if $<f(o_1^{\mathbf{A}}), \ldots, f(o_k^{\mathbf{A}})> \in R_i^{\mathrm{B}}$, and

[20.6] For every $F_i^{\mathbf{A}} \in \Phi^{\mathbf{A}}$ and $F_i^{\mathrm{B}} \in \Phi^{\mathrm{B}}, f(F_j^{\mathbf{A}}(a_1^{\mathbf{A}}, \ldots, a_p^{\mathbf{A}}))=F_j^{\mathrm{B}}(f(a_1^{\mathbf{A}}), \ldots, f(a_p^{\mathbf{A}}))$.

A structure $\mathbf{A}$ is *homomorphic* to a structure B (or there is a homomorphism between $\mathbf{A}$ and B) if and only if there is a homomorphism from the universe of $\mathbf{A}$ to the universe of B; $\mathbf{A}$ and B are *isomorphic* (or there is an *isomorphism* between $\mathbf{A}$ and B) if and only if there is an isomorphism from the universe of $\mathbf{A}$ to the universe of B. Finally, $\mathbf{A}$ is x-morphic to B if and only if there is some specific but unspecified morphism (e.g. isomorphism or homomorphism) from the universe of $\mathbf{A}$ to the universe of B.

As I interpret it, the structural account, like the similarity account, is an account of *faithful* epistemic representation, not an account of epistemic representation—it is an attempt to identify which further conditions (besides the vehicle being an epistemic representation of the target) must obtain in order for the vehicle to be a (more or less) faithful epistemic representation of the target.

To a first approximation, the general template of the structural account of faithful epistemic representation is as follows:

(F) *v* is a faithful epistemic representation of *t* (for *u*) if and only if:

(F.1) *v* is an epistemic representation of *t* (for *u*), and

(F.2) *v* is *x*-morphic to *t*.

It is important to note that, as it stands, (F) is not an account of faithful epistemic representation but a general template from which specific versions of the structural account of faithful epistemic representation can be generated. First of all, (F) does not specify what kind of faithful epistemic representation (F.1) and (F.2) are the necessary and sufficient conditions for. Is it completely faithful, partially faithful, or specifically faithful epistemic representation? Second, according to (F.1), a vehicle is a faithful epistemic representation of a certain target for a user only if it is an epistemic representation of that target for that user. Therefore, any version of the structural account of faithful epistemic representation needs to rely on some account of epistemic representation or other. I believe that the interpretational account of epistemic representation that I developed in Chapter 4.3 is the one that best suits a structural account of faithful epistemic representation. Third, (F.2) leaves two crucial questions unanswered. The first is what morphism needs to hold between the vehicle and the target in order for the vehicle to be a faithful epistemic representation of that target. The second is how a morphism, which is a relation that holds only between set-theoretic structures, can hold between the majority of ordinary vehicles and targets, which do not themselves seem to be set-theoretic structures.

Interpreted as an account of faithful epistemic representation, the structural account avoids many of the objections that are often aimed at the structural account (see, for example, Suárez (2003)'s objections against the account that he labels [iso]). Since these objections are very similar to the objections to the similarity account that I have discussed in Chapter 6 (§6.2) and my replies to those objections are, *mutatis mutandis*, the same, I only discuss one of them here. The objection assumes that the *x*-morphism mentioned in (F.2) is isomorphism (which seems to be a plausible assumption in the case in which the vehicle is a completely faithful epistemic representation of the target) and states that isomorphism is reflexive, symmetric, and transitive, but representation is non-reflexive (the map represents the city not itself), non-symmetric (the map represents the city but the city does not represent the map), and non-transitive (a photograph of the map on a website that sells maps represents the map, not the city). However, just like the similarity account of faithful epistemic representation, the structural account of faithful epistemic representation is immune to this sort of objection. As I interpret it, the structural account does not claim that, if some vehicle is *x*-morphic to something, then it is an epistemic representation of that thing; what it does claim is that, *if a vehicle is an epistemic representation of a target (for some user)*, then it is a faithful epistemic representation of that target only if some specific morphism holds between the vehicle and the target. So, while, for example, the map might be isomorphic to itself (more about this below), it does not follow that, on the structural account, it is an epistemic representation of itself, for, according to (F), the existence of a morphism between a vehicle and a target is neither necessary nor sufficient for epistemic representation.

In this chapter, I consider a number of versions of the structural account of faithful epistemic representation. After addressing a couple of pressing preliminary questions in Sections 7.2 and 7.3, in Section 7.4 I develop *the isomorphism account of completely faithful epistemic representation*. This account is useful in illustrating how the structural account applies to the simplest cases of faithful epistemic representation—those in which the vehicle is a completely faithful epistemic representation of the target. Unfortunately, however, most cases of epistemic representation are *not* cases of completely faithful epistemic representation. If the structural account is to be taken seriously, it must also account for partially faithful epistemic representation and, in those cases, the relevant isomorphism fails to hold between the vehicle and the target. This is a well-known problem among the supporters of the structural account. The commonly accepted solution to this problem is to opt for a morphism weaker than isomorphism to account for partially faithful epistemic representation. The disagreement is mostly about which weaker morphism is the best candidate for the job. In Section 7.5, I distinguish three kinds of unfaithfulness—which I call 'incorrectness', 'incompleteness', and 'inexactness'—and I argue that any combination of them can characterize a partially faithful epistemic representation. In Sections 7.7 and 7.8, I consider some candidate weaker morphisms and argue that they are all inadequate as general accounts of partially faithful epistemic representation, for none of them can adequately account for epistemic representations that are both incorrect and incomplete.

In Section 7.9, I then develop what I call 'the partial isomorphism account of partially faithful epistemic representation'. This account is a refinement of the account proposed by Steven French and his collaborators. There I show how the account, as developed here, is successful in dealing with epistemic representations that are both incorrect and incomplete. In Section 7.10, however, I argue that the account is not equally successful in dealing with partially faithful epistemic representations that are inexact. I also argue that, more seriously, like all of the other structural accounts available, the partial isomorphism account of partially faithful epistemic representation fails to accommodate the fact that faithfulness is a gradable notion.

7.2 Structure Instantiation

One of the first problems that confront the structural account of faithful epistemic representation is that morphisms are only defined as relations between set-theoretic structures, and most vehicles and targets do not seem to be set-theoretic structures.[34] For example, the London Underground map and the London Underground network, or the inclined plane model and my daughters tobogganing downhill do not seem to be set-theoretic structures. So, how would an account of representation that relies on the formal notion of a morphism apply to vehicles and targets that are not structures?

The most promising answer to this question, I believe, is to maintain that, even if those vehicles and targets are not set-theoretic structures, they can *instantiate* set-theoretic structures. Two concrete objects or systems, then, could be informally said to be *x*-morphic if and only if they instantiate *x*-morphic structures. Roman Frigg (2006) has suggested that to say that a concrete object (or system of objects) instantiates a certain structure amounts to giving an abstract description of it, one that applies if and only if some specific, more concrete description of it applies. Following Nancy Cartwright (1999, Ch. 2), Frigg suggests a necessary condition that must obtain in order for one description to be more abstract than another. A description such as 'Anita is playing a game' is more abstract than a set of descriptions—such as one that includes 'Anita is playing chess', 'Anita is playing football', 'Anita is playing poker'—only if (a) 'Anita is playing a game' cannot apply to a certain situation unless one or other of the more concrete descriptions in that set applies *and* (b) the fact that the more concrete description is satisfied is what the fact that the more abstract description is satisfied consists in on that instance. Frigg then claims:

> […] for it to be the case that *possessing a structure* applies to a system, *being an individual* must apply to some of its parts and *standing in a relation* to some of these. The crucial thing to realise at this point is that *being an individual* and *being in a relation* are abstract on the model of *playing a game*. (Frigg 2006, p.55)

So, for example, the London Underground network instantiates a structure $\mathbb{N}$=<{Acton Town, Aldgate, …, Woodside Park, Woolwich Arsenal}, …,$^{1}R_i^{\mathbb{N}}$ {<Blackhorse Road>, <Brixton>, <Euston>, <Finsbury Park>, <Green Park>, <Highbury & Islington>, <Oxford Circus>, <Pimlico>, <Seven Sisters>, <Stockwell>, <Tottenham Hale>, <Vauxhall>, <Victoria>, <Walthamstow Central>, <Warren Street>},…> if and only if an abstract description that includes, among other things, 'there is a property that Blackhorse Road, Brixton, …, Walthamstow Central, and Warren Street stations have (and no other station on the network has)' is true of the London Underground network. I call this description 'a *structural description* of the London Underground network'. This structural description is true only if a suitable, more concrete description is true of the network (in this case, the more concrete description would be something along the lines of 'Blackhorse Road, Brixton, …, Walthamstow Central, and Warren Street stations are on the Victoria Line and no other station is on the Victoria Line').

The structural description is more abstract because it tells us nothing about the nature of the objects that make up the vehicle and the target or about the nature of their properties and relations. All it tells us is that such-and-such objects are in the universe of the structure (and that some of them have some property or other or stand in some relation or other) while others are not. In the example above, the property is that of *being on the Victoria Line*, but any other more concrete property (say, having four escalators) will do insofar as it is shared by the abovementioned stations and by no other stations on the network.

On this account, the same object or system can instantiate a number of different structures depending on which more concrete description of the system the structural description is based on—i.e. depending on which of its parts we take to be the relevant objects and which of their concrete properties and relations have their abstract counterparts included in the structure. The

[34] Some advocates of the semantic account of theories seem to believe that scientific models *are* set-theoretic structures (see e.g. van Fraassen 1997). However, this would solve only half of the problem because the systems that models represent are not set-theoretic structures but concrete systems.

same vehicle and the same target thus are likely to instantiate not one but many structures under different (true) descriptions.

The fact that a concrete vehicle or target potentially instantiates a variety of structures under different descriptions, however, is not a problem for the structural account insofar as we have a principled way to single out one structure for the vehicle and one structure for the target as the relevant ones. The interpretational account of epistemic representation that I defend in Chapter 4 allows us to do just that. According to the interpretational account, a vehicle is an epistemic representation of a certain target if and only if a user adopts an interpretation of that vehicle in terms of the target. According to that account, one way for a user to interpret a vehicle in terms of a target is to take some objects, properties, and relations in the vehicle to denote (putative) objects, properties, and relations in the target. I called these objects, properties, and relations in the vehicle and the target, respectively, the *v*-relevant objects, properties, and relations and the *t*-relevant ones (according to that interpretation).

On the standard interpretation of the London Underground map in terms of the London Underground network, for example, the circles and tabs on the map are *v*-relevant objects (according to the standard interpretation of the map) because they stand for stations on the network and the relation *being connected by a light blue line* in the map is a *v*-relevant relation (according to the standard interpretation of the map) because it stands for a relation between stations on the network. The relation *being two inches left of*, on the other hand, is not *v*-relevant (according to the standard interpretation of the map) because, according to that interpretation, it does not stand for any relation among stations on the network. Analogously, the relation *being connected by Victoria Line trains* in the network is *t*-relevant (according to the standard interpretation of the map) because there is a relation in the map that stands for that relation, while the relation *being three miles away from* is not *t*-relevant (on the standard interpretation of the map) because, according to that interpretation, no relation among objects on the map stands for that relation.

Now suppose that there is a true description of the vehicle in terms of all and only the *v*-relevant objects and their *v*-relevant properties and relations (according to the user's interpretation of the vehicle in terms of the target). Suppose further that there is a true description of the target in terms of all and only the *t*-relevant objects and their *t*-relevant properties and relations (according to the user's interpretation of the vehicle in terms of the target). Under those ideal descriptions, both the vehicle and the target instantiate structures, which I call, respectively, 'the *relevant structure of the vehicle*, $\mathbf{V}$, (*relative to u's interpretation of v in terms of t*, $i^{\circ}(v \rightarrow t)$)' and '*the relevant structure of the target*, $\mathbf{T}$, (*relative to u's interpretation of v in terms of t*, $i^{\circ}(v \rightarrow t)$)'.

In general, 'relevant structure' can be defined as follows:

(21) If v is an epistemic representation of t (for u), then $\mathbf{V}$ is the relevant structure of v (relative to u's interpretation of v in terms of t, $i^{\circ}(v \rightarrow t)$) if and only if:

[21.1] $A^{\mathbf{V}}=\{o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}\}$ is the universe of the structure $\mathbf{V}$ if and only if $A^{\mathbf{V}}$ is the set of v-relevant objects according to $i^{\circ}(v \rightarrow t)$.

[21.2] ${}^{n}R_k^{\mathbf{V}}=\{<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}>, \ldots, <o_i^{\mathbf{V}}, \ldots, o_j^{\mathbf{V}}>\}$ if and only if , according to $i^{\circ}(v \rightarrow t)$, some n-ary v-relevant relation holds among $o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}, \ldots$, and among $o_i^{\mathbf{V}}, \ldots, o_j^{\mathbf{V}}$, but does not hold among any other n-tuple of objects in $A^{\mathbf{V}}$.

[21.3] ${}^{n}F_k^{\mathbf{V}}\{f(o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}})=o_k^{\mathbf{V}}, \ldots, f(o_i^{\mathbf{V}}, \ldots, o_j^{\mathbf{V}})=o_z^{\mathbf{V}}\}$ if and only if, according to $i^{\circ}(v \rightarrow t)$, some n-ary v-relevant function takes $o_k^{\mathbf{V}}$ as its value if its arguments are $o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}, \ldots$, and takes $o_z^{\mathbf{V}}$ as its value if its arguments are $o_i^{\mathbf{V}}, \ldots, o_j^{\mathbf{V}}$.

(22) If v is an epistemic representation of t (for u), then $\mathbf{T}$ is the relevant structure of t (relative to u's interpretation of v in terms of t, $i^{\circ}(v \rightarrow t)$) if and only if:

[22.1] $A^{\mathbf{T}}=\{o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}}\}$ is the universe of the structure $\mathbf{T}$ if and only if, according to $i^{\circ}(v \longrightarrow t)$, $A^{\mathbf{T}}$ is the set of *t*-relevant objects.

[22.2] ${}^{n}R_k^{\mathbf{T}}=\{<o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}}>, \ldots, <o_i^{\mathbf{T}}, \ldots, o_j^{\mathbf{T}}>\}$ if and only if, according to $i^{\circ}(v \longrightarrow t)$, some *n*-ary *t*-relevant relation holds among $o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}}$, …, among $o_i^{\mathbf{T}}, \ldots, o_j^{\mathbf{T}}$, but does not hold among any other *n*-tuple of objects in $A^{\mathbf{T}}$.

[22.3] ${}^{n}F_k^{\mathbf{T}}\{f(o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}})=o_k^{\mathbf{T}}, \ldots, f(o_i^{\mathbf{T}}, \ldots, o_j^{\mathbf{T}})=o_z^{\mathbf{T}}\}$ if and only if, according to $i^{\circ}(v \longrightarrow t)$, some *n*-ary *t*-relevant function takes $o_k^{\mathbf{T}}$ as its value for the arguments $o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}}$, …., takes $o_z^{\mathbf{T}}$ as a value for the arguments $o_i^{\mathbf{T}}, \ldots, o_j^{\mathbf{T}}$.

By singling out some objects, properties, and relations in the vehicle and in the target as relevant to the epistemic representation, an analytic interpretation thus allows us to specify, in a principled and natural way, two structures among the many that the vehicle and the target can plausibly be taken to instantiate. As far as I can see, given the independent plausibility of the interpretational account of epistemic representation and the structural account of completely faithful epistemic representation, the fact that they cohere with and support each other so well is a further reason to accept them both.

Even if, strictly speaking, a morphism can only hold between the relevant structure of a vehicle and that of a target, for the sake of simplicity, in what follows, I often talk informally of morphisms holding between a vehicle and a target or of a vehicle and a target being *x*-morphic. This, however, should always be interpreted as a loose way of saying that a morphism holds between the vehicle and the target relative to the interpretation of the vehicle in terms of the target adopted by the user.

Before concluding this section, I should note that, while set-theoretic structures are usually construed *extensionally*, for our purposes the relevant structures of a vehicle and a target (relative to a certain interpretation) are better construed *intensionally*. This is to avoid the possibility of conflating distinct but coextensive relevant properties and relations. Suppose, for example, that *being an interchange station* and *having elevators* are both *t*-relevant properties of the London Underground network (according to an interpretation of a map in terms of the network) and that all and only interchange stations have elevators. In this case, the two properties are co-extensional. However, since the two properties are distinct properties of the stations that have them, they should be kept distinct (because, for example, the map might be right about which stations are interchange stations but wrong about which stations have elevators).

Since construing properties and relations as sets of *n*-tuples of objects is very convenient in this context, I shall continue to take properties and relations to be *n*-tuples of elements of the domain, and introduce an *n*-tuple of positive integers to distinguish distinct co-extensional properties and relations. So, for example, the fact that, say, ${}^{2}R_{78}^{\mathbf{A}}$ and ${}^{2}R_{143}^{\mathbf{A}}$ stand for two distinct co-extensional binary relations can be reflected by adding the tuple <78, 78> to the elements of ${}^{2}R_{78}^{\mathbf{A}}$ and the tuple <143, 143> to the elements of ${}^{2}R_{143}^{\mathbf{A}}$, so that ${}^{2}R_{78}^{\mathbf{A}} \neq {}^{2}R_{143}^{\mathbf{A}}$. In what follows, by 'the relevant structure' I usually mean 'the relevant intensional structure'.

7.3 Intended Morphisms

In the previous section, I introduced the notions of the relevant structure of the vehicle, $\mathbf{V}$, and the relevant structure of the target, $\mathbf{T}$. It might be tempting to think that we now have all we need to formulate a structural account of completely faithful epistemic representation. The account would state that

(G) *v* is a completely faithful epistemic representation of *t* (for *u*) if and only if:

(G.1) :

(G.1.1) u takes v to denote t,

(G.1.2) u adopts an interpretation of v in terms of t, $i^{\circ}(v \rightarrow t)$, and

(G.2) an isomorphism holds between $\mathbf{V}$ (i.e. the relevant structure of v relative to $i^{\circ}(v \rightarrow t)$) and $\mathbf{T}$ (i.e. the relevant structure of t relative to $i^{\circ}(v \rightarrow t)$).

As it turns out, however, (G) is not yet an adequate account of completely faithful epistemic representation. To see why, suppose, for example, that the printers have mistakenly inverted the colours of the light blue line and the red lines on the London Underground map (but not its legend), so that the line that is red on the regular map is light blue on the defective map and *vice versa*. If the relevant structures of the ordinary map and the network are isomorphic, so are the ones of the defective map and the network, for the relevant structures of the ordinary map and the defective map are themselves isomorphic. The defective map, however, is not a completely faithful epistemic representation of the network, for, from it, one could falsely infer, for example, that Central Line trains stop at Highbury & Islington. However, if (G) were an adequate account of completely faithful epistemic representation, then if the ordinary map were a completely faithful epistemic representation of the network, the defective map would be as well, for the relevant structures of the two maps are isomorphic to each other.

This consequence, I think, can be avoided by introducing the notion of an *intended morphism*. An intended morphism is one that associates an object, a property, or a relation in the vehicle with an object, a property, or a relation in the target only if the user's interpretation of the vehicle in terms of the target takes the first to denote the second. Unlike the isomorphism between the regular map and the network, the isomorphism between the defective map and the network is not an *intended* isomorphism, for, among other things, it incorrectly associates circles connected by a red line with stations connected by Victoria Line trains even if, on the standard interpretation of the map, a red line connecting two stations is supposed to represent the fact that Central Line trains operate between those two stations.

The notion of an intended morphism is more precisely defined as follows.

(23) A morphism, f, between the relevant structure of v and that of t is *intended* (*relative to a analytic interpretation* $i^{\circ}(v \rightarrow t)$) if and only if:

[23.1] for all $o_i^{\mathbf{V}} \in \mathbf{\Omega}^{\mathbf{V}}$ and all $o_i^{\mathbf{T}} \in \mathbf{\Omega}^{\mathbf{T}}$, if $f(o_i^{\mathbf{V}}) = o_i^{\mathbf{T}}$, then, according to $i^{\circ}(v \rightarrow t)$, $o_i^{\mathbf{V}}$ denotes $o_i^{\mathbf{T}}$,

[23.2] for all ${}^{n}R_k^{\mathbf{V}} \in \mathbf{P}^{\mathbf{V}}$ and ${}^{n}R_k^{\mathbf{T}} \in \mathbf{P}^{\mathbf{T}}$ and for all $<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}> \in (\mathbf{\Omega}^{\mathbf{V}})^n$, if $<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}> \in {}^{n}R_k^{\mathbf{V}}$ and $<f(o_1^{\mathbf{V}}), \ldots, f(o_n^{\mathbf{V}})> \in {}^{n}R_k^{\mathbf{T}}$, then, according to $i^{\circ}(v \rightarrow t)$, ${}^{n}R_k^{\mathbf{V}}$ denotes ${}^{n}R_k^{\mathbf{T}}$, *and*

[23.3] for all ${}^{n}F_k^{\mathbf{V}} \in (\mathbf{\Phi}^{\mathbf{V}})$ and ${}^{n}F_k^{\mathbf{T}} \in (\mathbf{\Phi}^{\mathbf{T}})$ and all $<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}> \in (\mathbf{\Omega}^{\mathbf{V}})^n$ and all $o_i^{\mathbf{V}} \in (\mathbf{\Omega}^{\mathbf{V}})^n$, if ${}^{n}F_k^{\mathbf{V}}(o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}) = o_i^{\mathbf{V}}$ and ${}^{n}F_k^{\mathbf{T}}(f(o_1^{\mathbf{V}}), \ldots, f(o_n^{\mathbf{V}})) = f(o_i^{\mathbf{V}})$, then according to $i^{\circ}(v \rightarrow t)$, ${}^{n}F_k^{\mathbf{V}}$ denotes ${}^{n}F_k^{\mathbf{T}}$.

I think that no version of the structural account of faithful epistemic representation can be successful unless it employs the notion of an intended morphism (or some similar notion). If the isomorphism between the vehicle and the target is not an intended one, then not only will the vehicle not be a completely faithful epistemic representation of the target, it might be an extremely unfaithful epistemic representation of it. To see why this is the case, we can imagine a defective map of the London Underground that is exactly like a regular map except for the fact that each line is randomly assigned a different colour from the one it has on the regular map and each tab or circle is randomly assigned the name of a different station from the one it has on the regular map. Such a map would be as close as possible to being a completely *un*faithful epistemic representation of the network (if there is any such thing (see below)) and yet, since it is isomorphic to the standard map, if the standard map is isomorphic to the network, then so is this defective map. However, whereas (at least) one of the isomorphisms that hold between the standard map and the network is an intended isomorphism, none of the isomorphisms that holds between the defective map and the network will be an intended isomorphism.

Although supporters of the structural account have sometimes expressed the need to focus only on the intended morphisms between a vehicle and a target (see, for example, van Fraassen 1997), no account of what conditions a morphism must meet in order to be intended has ever been specified. The notion of an interpretation, however, allows us to clearly define what conditions a morphism needs to meet in order to be an intended one. I am now in a position to formulate what I call 'the isomorphism account of completely faithful epistemic representation', which, I argue, is successful in all cases in which a vehicle is a completely faithful epistemic representation of its target.

7.4 Isomorphism

The isomorphism account of completely faithful epistemic representation claims that:

(H) *v* is a completely faithful epistemic representation of *t* (for *u*) if and only if:

(H.1) :

(H.1.1) *u* takes *v* to denote *t*,

(H.1.2) *u* adopts an (analytic) interpretation of *v* in terms of *t*, $i°(v \longrightarrow t)$, and

(H.2) an intended isomorphism holds between the relevant structure of *v* relative to $i°(v \longrightarrow t)$, $\mathbf{T}$, and the relevant structure of *t* relative to $i°(v \longrightarrow t)$, $\mathbf{V}$.

The isomorphism account of completely faithful epistemic representation is successful in all cases in which the vehicle is a completely faithful epistemic representation of a certain target (relative to some analytic interpretation adopted by its user). In order to show this, I need to show that:

(a) if the vehicle is an (analytically interpreted) epistemic representation of the target for a certain user and an intended isomorphism holds between the vehicle and the target, then all valid surrogative inferences from the vehicle to the target are sound and
(b) if the vehicle is an (analytically interpreted) epistemic representation of the vehicle in terms of the target and all valid surrogative inferences from the vehicle to the target are sound, then an intended isomorphism holds between the vehicle and the target.

Consider (a) first. If a user adopts an analytic interpretation of the vehicle in terms of the target, an inference from the vehicle to the target will be valid if and only if it is in accordance with the rules outlined in Section 5.6. I now show that, if an intended isomorphism holds between the vehicle and the target, then all valid inferences—i.e. all inferences that are in accordance with *(Rule 1)*, *(Rule 2)*, or *(Rule 3)*—are sound.

(Rule 1). Assume that there is an object $o_i^{\mathbf{V}}$ in the vehicle that denotes a (putative) object $o_i^{\mathbf{T}}$ in the target (according to the interpretation adopted by the user). According to *(Rule 1)*, it is then valid to infer that there is an object $o_i^{\mathbf{T}}$ in the target. Now, if an isomorphism, *f*, holds between the vehicle and the target, then there must be an object, $o_x^{\mathbf{T}}$, in the universe of the relevant structure of the target, $A^{\mathbf{T}}$, such that $f(o_i^{\mathbf{V}})=o_x^{\mathbf{T}}$. If *f* is an intended isomorphism, it must be the case that $o_x^{\mathbf{T}}=o_i^{\mathbf{T}}$, because $o_i^{\mathbf{T}}$ is the object that is denoted by $o_i^{\mathbf{V}}$ according to the interpretation adopted by the user.

Now, assume that there is no object $o_i^{\mathbf{V}}$ in the vehicle that denotes a (putative) object $o_i^{\mathbf{T}}$ in the target (according to the interpretation adopted by the user). According to *(Rule 1)*, it is then valid to infer that there is no object $o_i^{\mathbf{T}}$ in the target. If an isomorphism *f* holds between the vehicle and the target then every object in the universe of the relevant structure of the target must be in one-to-one correspondence with some object different from o_i^V and, if this isomorphism is intended, then o_i^T cannot be among the objects in the universe of the structure of the target because an intended isomorphism would associate o_i^T only with o_i^V, which is the object that denotes it according to the interpretation adopted by the user. Therefore, it is sound to infer that the object o_i^T is not in the target. So, if an intended isomorphism holds between the vehicle and the target, any inference that is in accordance with *(Rule 1)* will be a sound inference.

(Rule 2). Assume that certain *v*-relevant objects in the vehicle, o_1^V, …, o_n^V, are in a certain *v*-relevant *n*-ary relation, $''R_k^V$ and that, according to the interpretation adopted by the user, o_1^V denotes o_1^T, …, o_n^V denotes o_n^T, and $''R_k^V$ denotes $''R_k^T$. According to *(Rule 2)*, it is therefore valid to infer that a relation $''R_k^T$ holds among o_1^T, …, o_n^T. If an isomorphism *f* holds between the vehicle and the target, then a relation $''R_x^T$ will hold among the objects $f(o_1^V)$, …, $f(o_n^V)$. If the isomorphism is intended, then, since o_1^V denotes o_1^T, …, and o_n^V denotes o_n^V, it must be the case that $f(o_1^V)= o_1^T$, …, $f(o_n^V)= o_n^T$ and, since $''R_k^V$ denotes $''R_k^T$, the relation $''R_k^T$ must be the relation $''R_k^V$.

Assume now that certain relevant objects in the vehicle, o_1^V, …, o_n^V, are not in a certain relevant *n*-ary relation, $''R_k^V$ and that, according to the interpretation adopted by the user, o_1^V denotes o_1^T, …, o_n^V denotes o_n^T, and $''R_k^V$ denotes $''R_k^T$. According to *(Rule 2)*, it is therefore sound to infer that the relation $''R_k^T$ does not hold among o_1^T, …, o_n^T. Here there are two cases to consider. Either a different relevant *n*-ary relation holds among o_1^V, …, o_n^V or no *n*-ary relation holds among them. If a different *n*-ary relation holds among o_1^V, …, o_n^V and an isomorphism holds between the vehicle and the target, then there will be a relation $''R_x^T$ that holds among the objects $f(o_1^V)$, …, $f(o_n^V)$. If the isomorphism is intended, then, since o_1^V denotes o_1^T, …, and o_n^V denotes o_n^T, it must be the case that $f(o_1^V)= o_1^T$, …, $f(o_n^V)= o_n^V$ but $''R_x^T$ cannot be $''R_k^V$ because, if the isomorphism is intended, $<o_1^V, \ldots, o_n^V> \in R_k^V$, and $<f(o_1^V), \ldots, f(o_n^V)> \in R_k^T$ only if, according to the interpretation adopted by the user, R_k^V denotes R_k^T. If no *n*-ary relation holds among o_1^V, …, o_n^V and an isomorphism holds between the vehicle and the target, then no relation holds among $f(o_1^V)$, …, $f(o_n^V)$. If the isomorphism is intended, then, since o_1^V denotes o_1^T, …, and o_n^V denotes o_n^T, it must be the case that $f(o_1^V)= o_1^T$, …, $f(o_n^V)=o_1^T$ and therefore no relation holds among those objects. So, if an intended isomorphism holds between the vehicle and the target, any inference that is in accordance with *(Rule 2)* is sound.

(Rule 3). Assume that the function $''F_k^V$ has o_i^V as its value when its arguments are o_1^V, …, o_n^V, and, according to the interpretation adopted by the user, o_i^V denotes o_i^T, o_1^V denotes o_1^T, …, o_n^V denotes o_n^T, and $''F_k^V$ denotes $''F_k^T$. According to *(Rule 3)*, it is therefore valid to infer that the value of the function $''F_k^T$ for the arguments o_1^T, …, o_n^T is o_i^T. If an isomorphism *f* holds between **V** and **T**, then $f(''F_k^V\ (o_1^V, \ldots, o_n^V))=f(o_i^V)=''F_k^T\ (f(o_1^V), \ldots, f(o_n^V))$. However, since, if the isomorphism is intended, then $f(o_i^V)=o_i^T$ and $''F_k^T\ (f(o_1^V), \ldots, f(o_n^V))=''F_k^T\ (o_1^T, \ldots, o_n^T)$, it must be the case that $''F_k^T\ (o_1^T, \ldots, o_n^T)= o_i^T$. So, if an intended isomorphism holds between the vehicle and the target, any inference that is in accordance with *(Rule 3)* is sound.

The above argument shows that, if an intended isomorphism holds between the vehicle and the target, all valid inferences from the vehicle to the target will be sound. I now turn to the converse claim, (b)—i.e. the claim that, if all valid surrogative inferences from the vehicle to the target are sound, then an intended isomorphism holds between the vehicle and the target. The argument for that claim goes as follows.

If all inferences that are in accordance with *(Rule 1)* are sound, then, it must be the case that, for every object o_i^V that is in the vehicle, the object denoted by o_i^V, o_i^T, is in the target and that, for every object o_i^V that is not in the vehicle, the object denoted by o_i^V, o_i^T, is not in the target.

If all inferences that are in accordance with *(Rule 2)* are sound, then it must be the case that, for every *n*-tuple of objects, o_1^V, …, o_n^V, that are in a *n*-ary relation $''R_k^V$, the objects denoted by o_1^V, …, o_n^V, o_1^T, …, o_n^T, are in the relation denoted by $''R_k^T$ and that, for every *n*-tuple of objects,

$o_1^{\mathrm{V}}, \ldots, o_n^{\mathrm{V}}$, that are not in a n-ary relation $''R_k^{\mathrm{V}}$, the objects denoted by $o_1^{\mathrm{V}}, \ldots, o_n^{\mathrm{V}}, o_1^{\mathrm{T}}, \ldots, o_n^{\mathrm{T}}$, are not in the relation denoted by $''R_k^{\mathrm{T}}$. So, an n-tuple of objects $o_1^{\mathrm{V}}, \ldots, o_n^{\mathrm{V}}$ is in a certain relation $''R_k^{\mathrm{V}}$ if and only if the objects denoted by $o_1^{\mathrm{V}}, \ldots, o_n^{\mathrm{V}}$ are in the relation $''R_k^{\mathrm{T}}$ denoted by $''R_k^{\mathrm{V}}$.

Finally, if all inferences in accordance with *(Rule 3)* are sound, then it must be the case that for every n-ary function $''F_k^{\mathrm{V}}$ whose value for the arguments $o_1^{\mathrm{V}}, \ldots, o_n^{\mathrm{V}}$ is o_i^{V}, the value of function denoted by $''F_k^{\mathrm{V}}$, $''F_k^{\mathrm{T}}$, is the object denoted by o_i^{V} when the arguments are the objects denoted by $o_1^{\mathrm{V}}, \ldots, o_n^{\mathrm{V}}$.

From this, it follows that, if all inferences in accordance with *(Rule 1)*, *(Rule 2)*, or *(Rule 3)* are sound, then it is possible to construct a function, f, from the relevant structure of the vehicle A^{V} to the relevant structure of the target A^{T} such that:

a) for every o_i^{V} and o_i^{T}, $f(o_i^{\mathrm{V}})=o_i^{\mathrm{T}}$ if and only if o_i^{V} denotes o_i^{T} according to the interpretation adopted by the user.
b) for all $''R_k^{\mathrm{V}}$ and $''R_k^{\mathrm{T}}$, $<o_1^{\mathrm{V}}, \ldots, o_n^{\mathrm{V}}> \in {''R_k^{\mathrm{V}}}$ if and only if $<f(o_1^{\mathrm{V}}), \ldots, f(o_n^{\mathrm{V}})> \in {''R_k^{\mathrm{T}}}$ and $''R_k^{\mathrm{V}}$ denotes $''R_k^{\mathrm{T}}$ according to the interpretation adopted by the user.
c) for all $''F_k^{\mathrm{V}}$ and $''F_k^{\mathrm{T}}$, $f(''F_k^{\mathrm{V}}(o_1^{\mathrm{V}}, \ldots, o_n^{\mathrm{V}}))={''F_k^{\mathrm{T}}}(f(o_1^{\mathrm{V}}), \ldots, f(o_n^{\mathrm{V}}))$ and $''F_k^{\mathrm{V}}$ denotes $''F_k^{\mathrm{T}}$ according to the interpretation adopted by the user.

Since any function that meets these conditions is an intended isomorphism, it follows that, if all inferences in accordance with *(Rule 1)*, *(Rule 2)* and *(Rule 3)* are sound, then an intended isomorphism holds between the vehicle and the target.

Let me briefly illustrate how the isomorphism account applies to the case of the London Underground map and the London Underground network. Suppose that an intended isomorphism holds between the relevant structure of the London Underground map and that of the London Underground network (relative to the standard interpretation of the map in terms of the network). This means that there is a bijective function that associates the circles and tabs on the map with the stations on the network in such a way that each circle and tab is associated with the corresponding station, and that each circle or tab has a certain (relevant) property (or is in a certain relation) if and only if the station it denotes has the corresponding (relevant) property (or is in the corresponding relation). For example, if an intended isomorphism holds between the relevant structure of the London Underground map and that of the London Underground network, then the circle labelled 'Holborn' and the tab labelled 'Bethnal Green' are connected by a red line if and only if Holborn and Bethnal Green stations are connected by Central Line trains because, according to the standard interpretation of the map, the circle labelled 'Holborn' and the tab labelled 'Bethnal Green' denote, respectively, Holborn and Bethnal Green stations and the relation *being connected by a red line* denotes the relation *being connected by Central Line trains*. Therefore, it follows that, if the user were to infer from the map that Holborn and Bethnal Green stations are connected by Central Line trains in accordance with *(Rule 2)*, her inference would be sound.

It is important to emphasize that the above holds only in the case that the isomorphism between the relevant structure of the map and that of the network is an *intended* isomorphism. If no intended isomorphism holds between the relevant structure of the map and that of the network, then some valid inferences will not be sound. Consider the examples discussed in Section 7.3. If the standard map is isomorphic to the network, then so is the defective map, for the standard map and the defective one are isomorphic to each other. The defective map, however, is not a completely faithful epistemic representation of the network, as some of the inferences that one can validly perform from the map to the network are unsound. For example, from the defective map, it is valid to infer that Highbury and Islington station is on the Circle Line, when in fact Holborn is on that line but Highbury and Islington station is not. So, even if the defective map and the network are isomorphic, the defective map is not a completely faithful epistemic representation of the network because none of the isomorphisms between the relevant structure of the map and that of the network is an intended one.

7.5 INCOMPLETENESS, INCORRECTNESS, AND INEXACTNESS

So far, I have argued that the isomorphism account of completely faithful epistemic representation is successful in all those cases of completely successful representation in which the user adopts an analytic interpretation of a vehicle in terms of a target, as in the case of the London Underground map. However, not all epistemic representations are completely faithful. This is particularly true of scientific models, which are our main focus here. Scientific models are usually idealized and approximate representations of their targets and, as such, they are usually far from being completely faithful epistemic representations of the systems they are used to represent. In this section, I distinguish three kinds of unfaithfulness that characterise partially faithful epistemic representations—*incompleteness*, *incorrectness*, and *inexactness*—and show that the isomorphism account cannot account for partially faithful epistemic representation.

(24) v is an *incorrect* epistemic representation of t (for u) if and only if:

[24.1] :

[24.1.1] u takes v to denote t,

[24.1.2] u adopts an interpretation of v in terms of t, $i^{\circ}(v \rightarrow t)$, and

[24.2] :

[24.2.1] for some $o_i^{V} \in A^{V}$, there is no $o_i^{T} \in A^{T}$ such that o_i^{V} denotes o_i^{T} (according to $i^{\circ}(v \rightarrow t)$), or

[24.2.2] for some ${}^{n}R_k^{V}$, o_1^{V}, …, and o_k^{V}, (according to $i^{\circ}(v \rightarrow t)$) o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, ${}^{n}R_k^{V}$ denotes ${}^{n}R_k^{T}$ and $\langle o_1^{V}, \ldots, o_k^{V} \rangle \in {}^{n}R_k^{V}$ but $\langle o_1^{T}, \ldots, o_n^{T} \rangle \notin {}^{n}R_k^{T}$.

(25) v is an *incomplete* epistemic representation of t for u if and only if:

[25.1] :

[25.1.1] u takes v to denote t,

[25.1.2] u adopts an interpretation of v in terms of t, $i^{\circ}(v \rightarrow t)$, and

[25.2] :

[25.2.1] for some $o^{T}_i \in A^{T}$, there is no $o_i^{V} \in A^{V}$ that denotes o_i^{T} (according to $i^{\circ}(v \rightarrow t)$), *or*

[25.2.2] for some ${}^{n}R_k^{V}$, o_1^{V}, …, and o_k^{V}, (according to $i^{\circ}(v \rightarrow t)$) o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, ${}^{n}R_k^{V}$ denotes ${}^{n}R_k^{T}$, and $\langle o_1^{T}, \ldots, o_k^{T} \rangle \in {}^{n}R_k^{T}$ but $\langle o_1^{V}, \ldots, o_n^{V} \rangle \notin {}^{n}R_k^{V}$.

(26) v is an *inexact* epistemic representation of t for u if and only if:

[26.1] :

[26.1.1] u takes v to denote t,

[26.1.2] u adopts an interpretation of v in terms of t, $i^{\circ}(v \rightarrow t)$, *and*

[26.2] for some ${}^{n}F^{V}_k$, o_1^{V}, …, and o_k^{V}, (according to $i^{\circ}(v \rightarrow t)$) o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, ${}^{n}F_k^{V}$ denotes ${}^{n}F_k^{T}$ but ${}^{n}F_k^{V}(o_1^{V}, \ldots, o_k^{V}) \neq {}^{n}F_k^{T}(o_1^{T}, \ldots, o_n^{T})$.

As we have seen in the previous section, if an intended isomorphism holds between a vehicle and a target, then the vehicle is a completely faithful epistemic representation of that target. Whenever at least one of the conditions for an intended isomorphism between the vehicle and the target fails to obtain, the target is a less-than-completely faithful epistemic representation of the target. In those cases, the vehicle is what I called a partially faithful epistemic representation of the target. The above definitions distinguish between three kinds of unfaithfulness depending on which condition(s) for intended isomorphism fail(s) to obtain. In this section, I focus on incorrectness and incompleteness and leave inexactness aside until the next section. Informally, incorrectness obtains whenever a relevant object in the vehicle has no counterpart in the target or whenever some relevant object in the vehicle have a relevant property or is in a relevant relation but the corresponding object in the target does not have the corresponding property or is not in the corresponding relation; incompleteness obtains whenever a relevant object in the target has no counterpart in the vehicle or whenever some relevant object in the target has a relevant property or is in a relevant relation but the corresponding object in the vehicle does not have the corresponding property or is not in the corresponding relation.

To illustrate the difference between incorrectness and incompleteness, consider again the example of the new and the old maps of the London Underground. According to the isomorphism account of completely successful representation, the old London Underground map is no longer a completely faithful epistemic representation of the London Underground network, for no intended isomorphism holds between the relevant structure of the map and that of today's network. There are at least two ways in which the intended isomorphism between the relevant structure of the old map and that of the network fails to hold and, as a consequence, the old map is both an incorrect and an incomplete epistemic representation of today's network.

The old map is an *incomplete* epistemic representation of today's network because, among other things, some of the stations and train lines on today's network have no counterpart on the map. For example, on today's network, Victoria Line trains operate between Highbury & Islington station and Victoria station. In the structure instantiated by today's network (relative to the standard interpretation of the old map), the ordered pair <Highbury & Islington, Victoria> is an element of the set of all pairs of stations connected by Victoria Line trains. However, since the circles that denote those stations are not connected by any coloured line on the old map, the ordered pair <tab labelled 'Highbury & Islington', circle labelled 'Victoria'> is not an element of any set of pairs of circles or tabs connected by a coloured line in the structure of the map (relative to its standard interpretation). As a consequence, an intended isomorphism fails to hold between the structure of the old map and the structure of the network.

The old map is also an *incorrect* epistemic representation of today's network because (among other things) some of the circles, tabs, and coloured lines on the map have no counterpart in today's network. For example, on the old map there is a tab labelled 'Dover Street' connected by a dark blue line to the circles labelled 'Piccadilly Circus' and 'Green Park'. In the structure instantiated by the old map, the tab labelled 'Dover Street' is an element of the set of circles connected by a blue line. Since there is no station called Dover Street on today's network, however, the tab labelled 'Dover Street' fails to denote any station on today's network. Therefore, there is no intended isomorphism between the map and the network, as an intended isomorphism would associate every circle on the map with some station on the network bearing the name printed beside the circle.

Despite the old map being an incomplete and incorrect epistemic representation of today's network, however, a large number of inferences from the map to today's network that are valid (according to the standard interpretation of the map in terms of the network) are still sound. Users could therefore draw a number of sound inferences from the map to the network without adopting a non-standard interpretation of the map in terms of the network.

Cases like this are far from being exceptional. In fact, they seem to be the norm when it comes to scientific models. Take, for example, Ptolemaic models of the universe. Despite being extremely unfaithful epistemic representations of the Universe as a whole (or even of the Solar System), some Ptolemaic models can be used (and are still sometimes used) to draw (approximately) true conclusions about the apparent motions and positions of a number of celestial bodies including the "fixed" stars, the Sun, the Moon, and the planets.

I take it that a satisfactory account of faithful epistemic representation should be able to explain how partially faithful epistemic representations, such as the old map and the Ptolemaic model of the universe, can be to some extent faithful even if no intended isomorphism holds between the vehicle and the target, as this is often considered one of the main problems faced by structural accounts (see, e.g., (Suárez 2003)). An alternative approach would be to assume that, for every epistemic representation that is partially faithful under its standard interpretation, there is some *ad hoc* interpretation of the vehicle in terms of the target under which the vehicle is a completely faithful epistemic representation of the target. I take this proposal to be dissatisfactory for a number of reasons. The main reason is that it is not descriptively adequate. In most cases, users of partially faithful epistemic representations of a certain target seem to adopt the standard interpretations of those epistemic representations and not some alleged *ad hoc* interpretation of them (on which only the sound inferences are valid). In fact, in many cases, the users would be unable to adopt such an interpretation, as they are unaware of which valid inferences are sound and which ones are not (pre-Copernican users of the Ptolemaic model, for example, used it to perform both sound and unsound inferences from the model, and this seems to suggest that they adopted its standard interpretation).

7.6 Models and Inexactness

It is widely acknowledged that scientific models are usually far from being completely faithful epistemic representations of their target systems. Consider again, for example, the inclined plane and the tobogganing hill. In that example, I intended to use the model to determine whether my daughters will go faster than I think it is safe. Suppose that, once we have plugged into the model numerical values for h and g, the final velocity of the box turns out to be sufficiently low. In virtue of what is this conclusion true (if it is indeed true)? The structural account of faithful epistemic representation would maintain that it is true in virtue of the fact that a certain morphism holds between the structure instantiated by the model and that instantiated by the system under the description of them that underlies the interpretation of the model in terms of the target. However, the intended morphism cannot be isomorphism.

Since most (if not all) scientific models are idealized and approximated epistemic representations of their target systems, isomorphism cannot be the morphism that holds between the relevant structure of a model and that of its target system. In fact, to my knowledge, not one of the sympathizers of the structural account thinks that isomorphism can be the morphism that holds between most scientific models and the systems they are used to represent. However, it is instructive to see exactly how isomorphism fails to obtain, because the exercise will assist us in identifying the characteristics that a morphism should have in order to play this role. Furthermore, it illustrates the third kind of unfaithfulness I introduced in the previous section—inexactness. Before being able to turn to this matter, however, we must first discuss the notion of relevant structure as it applies to a model such as the inclined plane model.

Following the work of Patrick Suppes and his collaborators (see, e.g., Suppes 2002) and of Wolfgang Balzer, Ulises Moulines and Joseph Sneed (Balzer, Moulines and Sneed 1987) on the set-theoretic structure of the models of classical particle mechanics, it is plausible to maintain that, on its standard interpretation, the inclined plane model instantiates a structure $\mathbb{M}=\langle\Omega^{\mathbb{M}}=\{\mathbf{O}^{\mathbb{M}}, T^{\mathbb{M}}, \mathbb{R}^3, \mathbb{R}, \mathbb{Z}^+\}, \mathbf{P}^{\mathbb{M}}=\varnothing, \mathbf{\Phi}^{\mathbb{M}}=\{\mathbf{r}^{\mathbb{M}}, m^{\mathbb{M}}, \mathbf{f}^{\mathbb{M}}, \mathbf{g}^{\mathbb{M}}\}\rangle$, where $\mathbf{O}^{\mathbb{M}}$ is a (non-empty) set of objects in the model (which in the case of the inclined plane contains only the box), $T^{\mathbb{M}}$ is an interval of real numbers (which is used to represent time in the model), $\mathbb{R}^3$ is a three-dimensional vector space over real numbers (which is used to represent all sort of three-dimensional vector quantities in the model), $\mathbb{R}$ is the set of real numbers (which are used to represent scalar quantities in the model), $\mathbb{Z}^+$ is the set of positive integers (which are used to label things in the model). Informally, the functions $\mathbf{r}^{\mathbb{M}}$ and $m^{\mathbb{M}}$ are used to represent, respectively, the positions and the masses of the objects in the model, and the functions $\mathbf{f}^{\mathbb{M}}$ and $\mathbf{g}^{\mathbb{M}}$ are used to represent the "internal" and "external" forces acting on those objects. More specifically $\mathbf{r}^{\mathbb{M}}$ is a function whose domain is $(\mathbf{O}^{\mathbb{M}}\times T^{\mathbb{M}})$ (i.e. the Cartesian product of $\mathbf{O}^{\mathbb{M}}$ and $T^{\mathbb{M}}$) and whose co-domain is $\mathbb{R}^3$; informally, $\mathbf{r}^{\mathbb{M}}$ assigns to each object and time pair a three-dimensional vector representing the position of that object at that time. For every $o_i \in \mathbf{O}^{\mathbb{M}}$ and every $t_i \in T^{\mathbb{M}}$, $\mathbf{r}^{\mathbb{M}}$ is twice differentiable at t_i. For the sake of clarity, I shall call '$\mathbf{v}^{\mathbb{M}}(o_i, t_j)$' and '$\mathbf{a}^{\mathbb{M}}(o_i, t_j)$' the first and the second derivative of $\mathbf{r}^{\mathbb{M}}(o_i, t_j)$ with respect to t ($d\mathbf{r}^{\mathbb{M}}(o_i, t_j)/dt=\mathbf{v}^{\mathbb{M}}(o_i, t_j)$ and $d^2\mathbf{r}^{\mathbb{M}}(a_i, t_j)/dt^2=\mathbf{a}^{\mathbb{M}}(o_i, t_j)$), so that $\mathbf{v}^{\mathbb{M}}(o_i, t_j)$ and $\mathbf{a}^{\mathbb{M}}(o_i, t_j)$ associate each object and time pair with the three-dimensional vectors that represent, respectively the velocity and the acceleration of that object at that time. The function $m^{\mathbb{M}}$ has $\mathbf{O}^{\mathbb{M}}$ as it s domain and the set of positive real numbers, $\mathbb{R}^+$ as its co-domain and, informally, it associates each object in $\mathbf{O}^{\mathbb{M}}$ with the real number representing the magnitude of its mass in grams. The function $\mathbf{f}^{\mathbb{M}}$ has the product of $((\mathbf{O}^{\mathbb{M}})^2\times T^{\mathbb{M}}\times\mathbb{Z}^+)$ and $\mathbb{R}^3$ as its co-domain and, informally, it associates every pair of objects in the domain and every time with the three-dimensional vector that represents one of the forces that the first object exerts on the second one at that time (the positive integer is just a convenient way to label different forces that one object may exert on the other). These forces can be construed as "internal" forces—forces exerted by objects that are within the system in question). The function $\mathbf{g}^{\mathbb{M}}$ has $(\mathbf{O}^{\mathbb{M}}\times T^{\mathbb{M}}\times\mathbb{Z}^+)$ as its domain and $\mathbb{R}^3$ as its co-domain and, informally, it associates each object and time pair with the three-dimensional vector that represents one of the "external" forces acting on it at that time (where external forces can be informally construed as forces that are exerted by objects in the model that are not included in $\mathbf{O}^{\mathbb{M}}$ or by source-less force fields).

There are different sort of constraints on the values these functions can have given a certain set of arguments. The most general set of constraints, which I call 'the *general constraints*', stem from the fact that the inclined plane model is a model of classical mechanics and, as such, the objects in it are subject to the general laws of classical mechanics, which constrain the values the functions can take in various ways. For example, according to Newton's Second Law, for every $o_i \in \mathbf{O}^{\mathbb{M}}$ and every $t_j \in T^{\mathbb{M}}$,

$$\mathbf{f}^{\mathbb{M}}(o_i, o_1, t)+\ldots+\mathbf{f}^{\mathbb{M}}(o_i, o_{i-1}, t)+\mathbf{f}^{\mathbb{M}}(o_i, o_{i+1}, t)+\ldots+\sum_{k=1}^{\infty}\mathbf{g}^{\mathbb{M}}(o_i, t, k)=m^{\mathbb{M}}(o_i)\mathbf{a}^{\mathbb{M}}(o_i, t).$$

According to Newton's Third Law, for every pair of objects, o_i and $o_k \in \mathbf{O}^{\mathbb{M}}$ and every $t_j \in T^{\mathbb{M}}$,

$$\mathbf{f}^{\mathbb{M}}(o_i, o_k, t_j)=-\mathbf{f}^{\mathbb{M}}(o_k, o_i, t_j).$$

A more specific set of constraints, which I call 'the *specific constraints*', stems from the specific features of the inclined plane model. In this particular case, the specific constraints concern the values of the function $\mathbf{g}^{\mathbb{M}}$. In the model, two forces act on the box at all times. The first one is an external gravitational force, $\mathbf{g}^{\mathbb{M}}(b, t_i, 1)$, whose magnitude is constant and equal to $m^{\mathbb{M}}(b)g$ (where g is gravitational acceleration). The second is the normal force $\mathbf{g}^{\mathbb{M}}(b, t_i, 2)$ that the plane exerts on the box whose direction is perpendicular to the plane and whose magnitude is constant and equal to $m^{\mathbb{M}}(b)g\sin\theta$ (where θ is the angle of inclination of the plane).[35] Since there are no other forces acting on the box, $\mathbf{g}^{\mathbb{M}}(b, t_i, k)=\mathbf{0}$ for all $k>2$.

[35] Note that here I consider the box the only object in the universe of the structure of the inclined plane model $\mathbf{O}^{\mathbb{M}}$ and the normal force of the plane on the box an external force. This is because considering the plane an object itself would give rise to certain counterintuitive consequences. For example, if the plane was one of the objects in the universe of the structure of the model, the function $m^{\mathbb{M}}$ would associate a certain mass with it and there is no obvious sense in which the plane in the model has a mass.

The general and specific constraints are all the constraints on the values of the functions in the structure of the inclined plane model as such. Despite these constraints, however, the functions in the structure of the model still do not have definite values for all arguments unless some additional constraints are specified. I call this further set of constraints 'the *inputs of the model*'. In the case of the inclined plane model, one such set of constraints consists in specifying the mass of the box ($m^{\mathbb{M}}(b)$), its initial position and velocity ($\mathbf{r}^{\mathbb{M}}(b, t_0)$ and $\mathbf{v}^{\mathbb{M}}(b, t_0)$), the gravitational acceleration (the values of g), and the angle of inclination of the plane (the value of θ). By specifying g and θ, the function $\mathbf{g}^{\mathbb{M}}$ will have a definite value for all arguments. Once we specify ($m^{\mathbb{M}}(b)$), $\mathbf{a}^{\mathbb{M}}(b, t_i)$ will have a definite value for all arguments as well $\mathbf{a}^{\mathbb{M}}(b, t_i)=(\mathbf{g}^{\mathbb{M}}(b, t_i, 1)+\mathbf{g}^{\mathbb{M}}(b, t_i, 2))/m^{\mathbb{M}}(b)$. Now we need only to specify the position and velocity of the box at some t_i such as t_0 in order for $\mathbf{r}^{\mathbb{M}}(b, t_0)$ to have definite values at all other times as well (($\mathbf{v}^{\mathbb{M}}(b, t_i)= \mathbf{a}^{\mathbb{M}}(b, t_i)t_i+\mathbf{v}^{\mathbb{M}}(b, t_0)$ and $\mathbf{r}^{\mathbb{M}}(b, t_0)=½\ \mathbf{a}^{\mathbb{M}}(b, t_i)t_i^2+\mathbf{v}^{\mathbb{M}}(b, t_0)t_i+\mathbf{r}^{\mathbb{M}}(b, t_0)$). Until the inputs of the model are specified, the inclined plane model, therefore, does not instantiate a structure. Rather, it instantiates what, following van Fraassen (1980, p.44), we can call 'a *structure-type*'. I call '*structure-tokens*' those structures that are instances of a certain structure-type. So, the inclined plane model instantiates a structure-token (i.e. a specific structure) only when a specific set of inputs of the model is specified.

Consider now the system composed by my daughters tobogganing down the hill. The system can be seen as instantiating a structure-type as well. Each token of that structure type is a structure of the form $\mathbb{S}=\langle\Omega^{\mathbb{S}}=\{O^{\mathbb{S}}, T^{\mathbb{S}}, \mathbb{R}^3, \mathbb{R}, \mathbb{Z}^+\}, P^{\mathbb{S}}=\varnothing, \Phi^{\mathbb{S}}=\{\mathbf{r}^{\mathbb{S}}, m^{\mathbb{S}}, \mathbf{f}^{\mathbb{S}}, \mathbf{g}^{\mathbb{S}}\}\rangle$, where $O^{\mathbb{S}}$ is a non-empty set of objects (which in this case contains (the mereological sum of) my daughters and their toboggan) and all other elements are defined analogously to the way they were defined for $\mathbb{M}$, so that, for example, the functions $\mathbf{r}^{\mathbb{M}}$ and $m^{\mathbb{M}}$ are used to represent, respectively, the position and the mass of my daughters on the toboggan and the functions $\mathbf{f}^{\mathbb{M}}$ and $\mathbf{g}^{\mathbb{M}}$ are used to represent the "internal" and "external" forces acting on my daughters and their toboggan. As in the case of the structure of the inclined plane model, we can also think of the structure of the tobogganing hill system as a structure-type. However, our reasons for doing so are different in the case of the system. In the case of the system, the reason to think of the structure as a structure-type is that we can "run the system" a number of times and, in all likelihood, the position and velocity of the toboggan and the forces acting on it will be different at different times. We can therefore think of the value of the functions in each token of the structure-type of the system as representing the value of that quantity at a certain time on a specific (possible) ride down the hill. This means that in each token of the structure-type of the system the function $\mathbf{r}^{\mathbb{S}}(f(b), f(t_i))$ associates the toboggan with its exact position at the time denoted by t_i on its n-th run ($\mathbf{r}^{\mathbb{S}}(f(b), f(t_i))=\mathbf{r}^{\mathbb{S}}(f(b), f(t_i))_n$). I call this token of the structure-type $\mathbb{S}$ '$\mathbb{S}_n$'.

Since morphisms are only defined for structure-tokens, not structure-types, if we want to apply the notion of a morphism to a case like that of the inclined plane model and the tobogganing hill system, we have to make sense of the notion of a morphism between structure-types. Here, I call 'the *inputs of the system*' those aspects of the system that correspond to the aspects of the model that I have called the inputs of the model. For example, since the inputs of the inclined plane model include the mass, initial position, and initial velocity of the box, the inputs of the tobogganing hill system will include the mass, initial position, and initial velocity of the toboggan. On each ride down the hill, these inputs will have definite values. Therefore, for each structure-token $\mathbb{S}_k$, the functions, $m^{\mathbb{S}}(f(b))_k$, the initial position, $\mathbf{r}^{\mathbb{S}}(f(b), f(t_0))_k$, the initial velocity, $\mathbf{v}^{\mathbb{S}}(f(b), f(t_0))_k$ etc. will have certain specific values. The token of the structure-type of the inclined plane model that corresponds to the structure-token $\mathbb{S}_k$, $\mathbb{M}_k$, is the one in which the value of all the inputs of the model is set equal to the value of the inputs of the system in structure-token $\mathbb{S}_k$ (e.g. $f(m^{\mathbb{M}}(b))=m^{\mathbb{S}}(f(b))_k$, $f(\mathbf{r}^{\mathbb{M}}(b, t_0))=\mathbf{r}^{\mathbb{S}}(f(b), f(t_0))_k$, $\mathbf{v}^{\mathbb{M}}(b, t_0)=\mathbf{v}^{\mathbb{M}}(f(b), f(t_0))_k$). The structure-type of the inclined plane model, $\mathbb{M}$, and that of the tobogganing hill system are thus x-morphic if and only if, for all $\mathbb{M}_k$ and $\mathbb{S}_k$ such that $\mathbb{M}_k$ is the token of the structure-type $\mathbb{M}$ that corresponds to the token $\mathbf{S}_k$ of the structure-type $\mathbf{S}$, $\mathbf{M}_k$ is *x*-morphic to $\mathbf{S}_k$. This allows us to continue to talk of the relevant structure of the inclined plane model and that of the tobogganing hill system and of the intended morphisms among them even if these structures are actually structure-types and not structure tokens.

As is probably already clear, no intended isomorphism holds between the relevant structure-type of the inclined plane model and the tobogganing hill system. Since in the standard interpretation of the model in terms of the system, the box denotes the toboggan, its position denotes the velocity of the toboggan, the external forces acting on it denote the "external" forces acting on the toboggan and so on, an intended morphism between the two structures would be one that associates the box with the toboggan and the elements of $T^{\mathbf{M}}$, $\mathbb{R}^3$, and $\mathbb{R}$ in the universe of the structure of the model with the same elements in the universe of the structure of the system (e.g. for every $r \in \mathbb{R}$, $f(r)=r$). The intended morphism between the two structures is an isomorphism only if, for every $t_i \in T^{\mathbf{M}}$ and every $k \in \mathbb{Z}^+$, $f(\mathbf{r}^{\mathbf{M}}(b, t_i))=\mathbf{r}^{\mathbf{S}}(f(b), f(t_i))$ and $f(\mathbf{g}^{\mathbf{M}}(b, t_i, k))=\mathbf{g}^{\mathbf{S}}(f(b), f(t_i), f(k))$ and $f(\mathbf{g}^{\mathbf{M}}(b, t_i, k))=\mathbf{g}^{\mathbf{S}}(f(b), f(t_i), f(k))$. In other words, the intended morphism is an isomorphism only if the position of the box and the forces acting on it at a certain time are identical to the position of the toboggan and the forces acting on it at the same time.

Now, we needn't analyse the situation in too much detail to realize that this cannot be the case. Consider, the "external" forces acting on the racer. First of all, unlike the gravitational force on the box in the model, the gravitational force that the Earth exerts on my daughters and their toboggan is not linear but increases as the square of the distance between the racer and the centre of mass of the earth decreases. Unlike the value of $f(\mathbf{g}^{\mathbf{M}}(b, t_i, 1))$, the value of $\mathbf{g}^{\mathbf{S}}(f(b), f(t_i), f(1))$, therefore, changes slightly as t_i increases and the toboggan goes downhill. Since, for all t, $f(\mathbf{g}^{\mathbf{M}}(b, t_i, 1)) \neq \mathbf{g}^{\mathbf{S}}(f(b), f(t_i), f(1))$, the intended morphism, f, cannot be an isomorphism as it does not meet condition [20.6] of definition (20).

Second, the normal force between the hill and the toboggan is likely to be different at different times. Unlike the inclined plane, the hill is not a perfectly straight slope and therefore the contact force the hill exerts on the toboggan is likely to change as a function of time. So, unlike the value of $f(\mathbf{g}^{\mathbf{M}}(b , t_i, 2))$, the value of $\mathbf{g}^{\mathbf{S}}(f(b), f(t_i), f(2))$ is likely to differ for different values of t_i. Since, for most t_i, it is likely that $f(\mathbf{g}^{\mathbf{M}}(b , t_i, 2)) \neq \mathbf{g}^{\mathbf{S}}(f(b), f(t_i), f(2))$, the intended morphism, f, cannot be an isomorphism as it does not meet condition [20.6] of definition (20).

Third, in the inclined plane model, there are only two "external" forces acting on the toboggan. The value of the function $\mathbf{g}^{\mathbf{M}}(b, t_i, k)$ for every $k>2$ is the zero vector. However, there are a number of other "external" forces acting on the toboggan that have no counterpart in the model, including air friction on my daughters and their toboggan, the friction between the snow on the hill and the toboggan, the gravitational force that any massive object in the universe (from the molecules of the air to distant galaxies) exerts on my daughters and their toboggan, and so on. So, for some $k>2$ and for every $t_i \in T^{\mathbf{M}}$, $f(\mathbf{g}^{\mathbf{M}}(b , t_i, k)) \neq \mathbf{g}^{\mathbf{S}}(f(b), f(t_i), f(k))$.

From the above considerations, it follows that the position of the box at each time $t_i \in T^{\mathbf{M}}$ after t_0 are different from the position of the toboggan at that time; or, in symbols, $f(\mathbf{r}^{\mathbf{M}}(b, t_i)) \neq \mathbf{r}^{\mathbf{S}}(f(b), f(t_i))$, for all $t_i > t_0$. We can thus conclude that the inclined plane model is an inexact epistemic representation of the tobogganing hill system, where an epistemic representation is *inexact* if and only if, for some ${}^nF_k^{\mathbf{V}}$, $o_1^{\mathbf{V}}$ denotes $o_1^{\mathbf{T}}$, …, $o_n^{\mathbf{V}}$ denotes $o_n^{\mathbf{T}}$, $o_i^{\mathbf{V}}$ denotes $o_i^{\mathbf{T}}$ and ${}^nF_k^{\mathbf{V}}$ denotes ${}^nF_k^{\mathbf{T}}$ and ${}^nF_k^{\mathbf{V}}(o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}})=o_i^{\mathbf{V}}$ but ${}^nF_k^{\mathbf{T}}(o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}}) \neq o_i^{\mathbf{T}}$.

This is far from being a peculiarity of this specific case. To a greater or lesser extent, approximation and idealization characterise most (if not all) scientific models. For example, no model of classical mechanics instantiates a structure that is isomorphic in the intended manner to that of any real system in the sense specified above, as usually classical models do not contain a counterpart for most of the forces that act on the objects in the system, and even those forces for which there is a counterpart are often only approximations of the forces in the system. As a result, most models in classical mechanics are inexact epistemic representations of their target systems.

7.7 Weaker Morphisms: Homomorphism

While most sympathizers of the structural account seem to agree that, in order to account for what I called partially faithful epistemic representation, we need to resort to morphisms weaker than isomorphism, there is little or no agreement as to which weaker morphism is the appropriate one. Homomorphism (see, e.g., Bartels 2006), Δ/Ψ-morphism (Swoyer 1989), partial isomorphism (see, e.g., French and Ladyman 1999 and da Costa and French 2003) are three candidates that have been defended in the literature. The rest of the chapter tries to determine whether weakening the isomorphism requirement solves the problem of accounting for partially faithful epistemic representation.

The first morphism I consider is homomorphism. This gives rise to what I call '*the homomorphism account of partially faithful epistemic representation*'. According to it:

(I) *v* is a partially faithful epistemic representation of *t* (for *u*) if and only if:

(I.1) :

(I.1.1) *u* takes *v* to denote *t*,

(I.1.2) *u* adopts an interpretation of *v* in terms of *t*, $i°(v\longrightarrow t)$, *and*

(I.2) an intended homomorphism holds between the relevant structure of *v* relative to $i°(v\longrightarrow t)$, $\mathbf{T}$, and the relevant structure of *t* relative to $i°(v\longrightarrow t)$, $\mathbf{V}$.

The problem with this proposal is that, although it can handle incomplete epistemic representations, it cannot deal with partially faithful epistemic representations that are incorrect or inexact. For example, since the old London Underground map is both incorrect and incomplete and the inclined plane model is inexact, the homomorphism account cannot account for the fact that both the old London Underground map and the inclined plane model are partially faithful epistemic representations of their respective targets. Therefore, I argue first that, if an intended homomorphism holds between the vehicle and the target, the vehicle may be an incomplete epistemic representation of the target and then that, if an intended homomorphism holds between the relevant structure of the vehicle and that of the system, the vehicle must be a correct and exact epistemic representation of the target. As usual, I assume throughout that the user adopts an analytic interpretation of the vehicle in terms of the target.

Incompleteness. According to (25), a vehicle is an incomplete epistemic representation of a target only if, for some $o_i^{\mathbf{T}} \in \Omega^{\mathbf{T}}$, $o_i^{\mathbf{V}}$ denotes $o_i^{\mathbf{T}}$ according to the interpretation of the vehicle in terms of the target, $i°(v\longrightarrow t)$, and $o_i^{\mathbf{V}} \notin \Omega^{\mathbf{V}}$ or, for some ${}^{n}R_k^{\mathbf{T}}$, $<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}> \notin {}^{n}R_k^{\mathbf{V}}$, even if $<o_1^{\mathbf{T}}, \ldots, o_k^{\mathbf{T}}> \in {}^{n}R_k^{\mathbf{T}}$ and $o_1^{\mathbf{V}}$ denotes $o_1^{\mathbf{T}}$, …, $o_n^{\mathbf{V}}$ denotes $o_n^{\mathbf{T}}$, ${}^{n}R_k^{\mathbf{V}}$ denotes ${}^{n}R_k^{\mathbf{T}}$ according to $i°(v\longrightarrow t)$. However, an intended homomorphism can hold between the vehicle and the target even if, for some $o^{\mathbf{T}}_i \in \Omega^{\mathbf{T}}$, $o_i^{\mathbf{V}}$ denotes $o_i^{\mathbf{T}}$ and $o_i^{\mathbf{V}} \notin \Omega^{\mathbf{V}}$ or, for some ${}^{n}R_k^{\mathbf{T}}$, $<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}> \notin {}^{n}R_k^{\mathbf{V}}$, even if $<o_1^{\mathbf{T}}$, …, $o_k^{\mathbf{T}}> \in {}^{n}R_k^{\mathbf{T}}$ and $o_1^{\mathbf{V}}$ denotes $o_1^{\mathbf{T}}$, …, $o_n^{\mathbf{V}}$ denotes $o_n^{\mathbf{T}}$, ${}^{n}R_k^{\mathbf{V}}$ denotes ${}^{n}R_k^{\mathbf{T}}$. For an intended homomorphism to hold between the vehicle and the target, it is necessary that, for every object $o_i^{\mathbf{V}}$ in the universe of the relevant structure of the vehicle, the object denoted by $o_i^{\mathbf{V}}$, $o_i^{\mathbf{T}}$, is in the universe of the relevant structure of the target. However, an intended homomorphism can hold even if, for some object $o_i^{\mathbf{T}}$ in the universe of the relevant structure of the target, there is no object in the universe of the relevant structure of the vehicle that denotes $o_i^{\mathbf{T}}$. Analogously, for an intended homomorphism to hold between the vehicle and the target, it is necessary that if ${}^{n}R_k^{\mathbf{V}}$

denotes $''R_k^{\mathtt{T}}$ and $<o_1^{\mathtt{V}}, \ldots, o_n^{\mathtt{V}}>\in''R_k^{\mathtt{V}}$, $<o_1^{\mathtt{T}}, \ldots, o_k^{\mathtt{T}}>\in''R_k^{\mathtt{T}}$. However, it is not necessary that, if $<o_1^{\mathtt{T}}, \ldots, o_n^{\mathtt{T}}>\in''R_k^{\mathtt{T}}$, $<o_1^{\mathtt{V}}, \ldots, o_k^{\mathtt{V}}>\in''R_k^{\mathtt{V}}$. So, if an intended homomorphism holds between the vehicle and the target, the vehicle may well be an incomplete epistemic representation of the target.

Correctness. If an intended homomorphism holds between the vehicle and the target, the vehicle cannot be an incorrect epistemic representation of the target. According to definition (24), if the vehicle was an incorrect epistemic representation of a target, it would be the case that, for some $o_i^{\mathtt{V}}\in\Omega^{\mathtt{V}}$, $o_i^{\mathtt{V}}$ denotes $o_i^{\mathtt{T}}$ and $o_i^{\mathtt{T}}\notin\Omega^{\mathtt{T}}$ or, for some $''R_k^{\mathtt{V}}$, $<o_1^{\mathtt{T}}, \ldots, o_n^{\mathtt{T}}>\notin''R_k^{\mathtt{T}}$, even if $<o_1^{\mathtt{V}}, \ldots, o_k^{\mathtt{V}}>\in''R_k^{\mathtt{V}}$ and $o_1^{\mathtt{V}}$ denotes $o_1^{\mathtt{T}}$, …, $o_n^{\mathtt{V}}$ denotes $o_n^{\mathtt{T}}$, $''R_k^{\mathtt{V}}$ denotes $''R_k^{\mathtt{T}}$. If this was the case, however, no intended homomorphism could hold between the vehicle and the target because, according to definitions (19) and (23), if an intended homomorphism held between the vehicle and the target, then every object $o_i^{\mathtt{V}}\in\Omega^{\mathtt{V}}$ would be associated with the object $o_i^{\mathtt{T}}\in\Omega^{\mathtt{T}}$ denoted by it and, for every $''R_k^{\mathtt{T}}$, if $<o_1^{\mathtt{V}}, \ldots, o_k^{\mathtt{V}}>\in''R_k^{\mathtt{V}}$, then $<o_1^{\mathtt{T}}, \ldots, o_k^{\mathtt{T}}>\in''R_k^{\mathtt{T}}$. So, if an intended homomorphism holds between the vehicle and the target, the vehicle must be a correct epistemic representation of the target.

Exactness. If an intended homomorphism holds between the vehicle and the target, the vehicle cannot be an inexact epistemic representation of the target. If the vehicle was an inexact representation of the target, it would be the case that, for some $''F_k^{\mathtt{V}}$, $o_1^{\mathtt{V}}$ denotes $o_1^{\mathtt{T}}$, …, $o_n^{\mathtt{V}}$ denotes $o_n^{\mathtt{T}}$, and $''F_k^{\mathtt{V}}$ denotes $''F_k^{\mathtt{T}}$ and $f(''F_k^{\mathtt{V}}(o_1^{\mathtt{V}}, \ldots, o_n^{\mathtt{V}})\neq''F_k^{\mathtt{T}}(f(o_1^{\mathtt{V}}), \ldots, f(o_n^{\mathtt{V}}))$. If this were the case, however, no intended homomorphism could hold between the vehicle and the target because, according to definitions (19) and (23), if an intended homomorphism holds between the vehicle and the target, then for every $''F_k^{\mathtt{V}}$, if $o_1^{\mathtt{V}}$ denotes $o_1^{\mathtt{T}}$, …, $o_n^{\mathtt{V}}$ denotes $o_n^{\mathtt{T}}$, and $''F_k^{\mathtt{V}}$ denotes $''F_k^{\mathtt{T}}$, $f(''F_k^{\mathtt{V}}(o_1^{\mathtt{V}}, \ldots, o_n^{\mathtt{V}})$ would have to be equal to $''F_k^{\mathtt{T}}(f(o_1^{\mathtt{V}}), \ldots, f(o_n^{\mathtt{V}}))$. So, if an intended homomorphism holds between the vehicle and the target, the vehicle must be an exact epistemic representation of the target.

The situation is reversed if, instead of a homomorphism between $\mathtt{V}$ and $\mathtt{T}$, the account required what we might call 'an inverse homomorphism' to hold between them (i.e. a homomorphism between $\mathtt{V}$ and $\mathtt{T}$). As it can be easily verified, this account would be able to handle epistemic representations that are incorrect, complete and exact, but not epistemic representations that are incomplete or inexact.

Each of the two accounts of partially faithful epistemic representation that I have examined in this subsection can only account for one sort of unfaithfulness. However, neither can account for partially faithful epistemic representations that are both incorrect and incomplete, like the old London Underground map, or for partially faithful epistemic representations that are inexact, like the inclined plane model. Both accounts are therefore inadequate.

7.8 Weaker Morphisms: Δ/Ψ-Morphism

Consider now a second candidate morphism—i.e. what Swoyer (1991) calls Δ/Ψ-morphism. For the sake of simplicity and consistency, here I modify and simplify Swoyer's proposal slightly. Most of what I say, however, applies *mutatis mutandis* to Swoyer's original proposal. The crucial notion of *Δ/Ψ-morphism* can be defined as follows.

(27) A function, *f*, from $\mathtt{A}$ to $\mathtt{B}$ is a *Δ/Ψ-morphism* if and only if:

[27.1] for all $''R_i^{\mathtt{A}}\in\Delta\subseteq\mathbf{P}^{\mathtt{A}}$ $(\Delta\neq\varnothing)$, if $<o_1^{\mathtt{A}}, \ldots, o_k^{\mathtt{A}}>\in''R_i^{\mathtt{A}}$, then $<f(o_1^{\mathtt{A}}), \ldots, f(o_k^{\mathtt{A}})>\in R_i^{\mathtt{B}}$ (*f preserves* all relations in Δ),

[27.2] for all $''R_i^{\mathtt{A}}\in\Psi\subseteq\mathbf{P}^{\mathtt{A}}$, if $<f(o_1^{\mathtt{A}}), \ldots, f(o_k^{\mathtt{A}})>\in R_i^{\mathtt{B}}$, then $<o_1^{\mathtt{A}}, \ldots, o_k^{\mathtt{A}}>\in''R_i^{\mathtt{A}}$ (*f counter-preserves* all relations in Ψ).

The *Δ/Ψ-morphism account of partially faithful epistemic representation* maintains that:

(J) *v* is a partially faithful epistemic representation of *t* (for *u*) if and only if:

(J.1) :

(J.1.1) *u* takes *v* to denote *t*,

(J.1.2) *u* adopts an interpretation of *v* in terms of *t*, $i^\circ(v \rightarrow t)$, *and*

(J.2) an intended Δ/Ψ-morphism holds between the relevant structure of *t* relative to $i^\circ(v \rightarrow t)$, $\mathbf{T}$, and the relevant structure of *v* relative to $i^\circ(v \rightarrow t)$, $\mathbf{V}$.

In this section, I argue that the intended Δ/Ψ-morphism does not successfully account for partially faithful epistemic representation. A first, minor problem is that Δ/Ψ-morphism does not allow a certain kind of incomplete representation to be partially faithful epistemic representation. A Δ/Ψ-morphism is a function from Ω^{T} to Ω^{V}. This means that an intended Δ/Ψ-morphism associates every object in Ω^{T} with the object that denotes it in Ω^{V}. However, an epistemic representation can still be partially faithful even if some objects in Ω^{T} are not denoted by any objects in Ω^{V}. For example, some stations on today's London Underground network have no counterpart on the 1930's map, but nevertheless we consider the latter a partially faithful epistemic representation of the former. This problem can easily be avoided by modifying the definition of Δ/Ψ-morphism so that a Δ/Ψ-morphism is a function, *f*, from *a non-empty subset of* the universe of the relevant structure of the target.

A second, more serious problem arises from the very notions of preservation and counter-preservation of a relation. A relation among objects in the universe of the relevant structure of the target (e.g. the relation *being connected by Metropolitan Line trains* in the London Underground network) is preserved if and only if, for any two stations, if those stations are connected by Metropolitan Line trains, then the circles or tabs that denote those stations are connected by a maroon line; it is counter-preserved if and only if, for any two circles or tabs, if those circles or tabs are connected by a maroon line, then the stations they denote are connected by Metropolitan Line trains. So, if the relation *being connected by Metropolitan Line trains* is preserved, it is sound to infer that two stations are *not* connected by Metropolitan Line trains if the circles or tabs that denote them are *not* connected by a maroon line; if, on the other hand, the relation *is* counter-preserved, it is sound to infer that two stations *are* connected by Metropolitan Line trains if the circles or tabs that denote them are connected by a maroon line.

According to the Δ/Ψ-morphism account, a vehicle is a partially faithful epistemic representation of a target only if some properties or relations in the target are counter-preserved. The rationale behind this requirement is that, if the Δ/Ψ-morphism holds, then it is possible to explain why some of the inferences performed by the user are sound—they are sound because if the relation holds among the objects in the vehicle, then the relation denoted by it also holds among the objects in the target that are denoted by those objects in the vehicle—and therefore why the epistemic representation in question is a partially faithful one.

However, the requirement is too strong—an epistemic representation can be partially faithful even if no property or relation in the target is counter-preserved. On the old London Underground map, for example, the circles labelled 'Aldgate' and 'Hammersmith' are connected by a maroon line, but today the corresponding stations are not connected by Metropolitan Line trains. Therefore the relation *being connected by Metropolitan Line trains* is not counter-preserved by any intended Δ/Ψ-morphism between the relevant structure of the old map and that of the network. Nevertheless, one can perform many sound inferences from the fact that circles or tabs are connected by a maroon line to the fact that the stations denoted by them are connected by Metropolitan Line trains. So, one can perform sound inferences from the fact that a certain relation holds among certain objects in the vehicle to the fact that the relation denoted by it holds among the objects they denote in the target, even if the relation is not counter-preserved. Even if no relation among objects in the target is counter-preserved, the vehicle may still be a partially faithful epistemic representation of the target if, for some objects in the vehicle, from the fact that a certain relation, *R*, holds among them, it is sound to infer that the relation denoted by *R* holds between the objects in the target that are denoted by those objects.

Here it is important to recall that an account of partially faithful epistemic representation is meant to identify in virtue of what an epistemic representation is a faithful one, not by virtue of what the user comes to know or believe that it is a faithful one. A user may not be able to determine which of the inferences from the fact that circles or tabs are connected by a maroon line to the fact that the stations denoted by them are connected by Metropolitan Line trains are sound and which are not, but the representation is still faithful to some degree if some of the inferences from it to the target are sound.

This intended Δ/Ψ-morphism account is not even successful in accounting for one of the cases that Swoyer seems to have in mind when introducing the notions of preservation and counter-preservation, namely the case of topographical maps. Swoyer correctly points out:

> […] it is a basic geometrical fact that a two-dimensional projection of a sphere cannot depict all its features without distortion, so when we use flat maps to represent the Earth, something has to give. For sixteenth-century mariners, concerned to convert lines of constant compass bearing (rhumb lines) into straight lines on their maps, Mercator's projection, which misrepresents scale, offered the best compromise; for other purposes equal area maps, which accurately represent scale but not shape, are preferable (Swoyer 1989, 470).

Swoyer is right in claiming that some maps use projections that counter-preserve some properties or relations of the geographical area represented. For example, the Polar Azimuthal projection counter-preserves the distance of every point from the North Pole, but it does so at the cost of distorting shapes and areas.[36] This is far from a unique case. Projections that counter-preserve one property or relation do so at the cost of distorting other properties or relations. However, it is exactly for this reason that many of the projections that are most commonly used, such as the Lambert Conformal Conic projection, do not counter-preserve any property or relation; instead, they try to minimize the distortion of as many properties or relations as possible (see, for example, (Fisher and Miller 1944)). Since maps based on these projections do not counter-preserve any property or relation, the Δ/Ψ-morphism account cannot account for the fact that they are partially faithful epistemic representation of their targets.

As the case of maps based on the Lambert Conformal Conic projection shows, a representation may be partially faithful even if none of the relations among objects in the universe of the target is counter-preserved by any intended Δ/Ψ-morphism. The Δ/Ψ-morphism account of partially faithful epistemic representation is therefore inadequate as an account of partially faithful epistemic representation.

7.9 Weaker Morphisms: Partial Isomorphism

In this section, I develop what I call 'the partial isomorphism account of partially faithful epistemic representation', which has been championed in some form or other by Newton Da Costa, Steven French, James Ladyman and their collaborators. While my formulation of the account differs from theirs in various respects, I hope it is an improvement on previous formulations.

[36] See (Fisher and Miller 1994), which is an excellent introduction to topographical map projections, or the more technical (Monmonier 1977) and (Maling 1992).

Here, I take the partial isomorphism account of partially faithful epistemic representation to claim that

(K) *v* is a partially faithful epistemic representation of *t* (for *u*) if and only if:

(K.1) :

(K.1.1) *u* takes *v* to denote *t*,

(K.1.2) *u* adopts an interpretation of *v* in terms of *t*, $i^{\circ}(v \rightarrow t)$, *and*

(K.2) an intended partial isomorphism holds between the relevant structure of *v* relative to $i^{\circ}(v \rightarrow t)$, $\mathbf{T}$, and the relevant structure of *t* relative to $i^{\circ}(v \rightarrow t)$, $\mathbf{V}$.

Two structures, $\mathbf{A}$ and $\mathbf{B}$, are *partially isomorphic* if and only if there are two partial substructures of $\mathbf{A}$ and $\mathbf{B}$ that are isomorphic. The best way to introduce the notion of partial substructure is, first, to introduce the notion of a *substructure* and that of a *partial* structure and, then, to combine the two together to get the notion of a *partial substructure*.

(28) A (total) structure $\mathbf{B}$ is *a substructure of* a (total) structure $\mathbf{A}$ if and only if:

[28.1] The universe of $\mathbf{B}$, $\Omega^{\mathbf{B}}$, is a subset of the universe of $\mathbf{A}$, $\Omega^{\mathbf{A}}$,

[28.2] For every ${}^{n}R_i^{\mathbf{A}} \in \mathbf{P}^{\mathbf{A}}$, there is ${}^{n}R_i^{\mathbf{B}} \in \mathbf{P}^{\mathbf{B}}$ such that ${}^{n}R_i^{\mathbf{B}} = {}^{n}R_i^{\mathbf{A}} \cap (\Omega^{\mathbf{B}})^{n}$. (More informally, for every relation in $\mathbf{A}$, there is a relation in $\mathbf{B}$ whose extension is the subset of the corresponding relation in $\mathbf{A}$ that only contains the *n*-tuples of elements that are in the universe of $\mathbf{B}$).

[28.3] For every ${}^{n}F_j^{\mathbf{A}} \in \Phi^{\mathbf{A}}$, there is a ${}^{n}F_j^{\mathbf{B}} \in \Phi^{\mathbf{B}}$, such that, if $dom({}^{n}F_j^{\mathbf{A}})$ is the domain of ${}^{n}F_j^{\mathbf{A}}$ and $codom({}^{n}F_j^{\mathbf{A}})$ is the codomain of ${}^{n}F_j^{\mathbf{A}}$, $dom({}^{n}F_j^{\mathbf{B}}) = dom({}^{n}F_j^{\mathbf{A}}) \cap (\Omega^{\mathbf{B}})^{n}$ and $codom({}^{n}F_j^{\mathbf{B}}) = codom({}^{n}F_j^{\mathbf{A}}) \cap (\Omega^{\mathbf{B}})^{n}$. (More informally, for every function in $\mathbf{A}$, there is a function in $\mathbf{B}$ whose domain is the subset of the universe of $\mathbf{A}$ that only contains the *n*-tuples of elements that are in the universe of $\mathbf{B}$).

A *partial structure* is an *n*-tuple $\mathbf{P} = \langle \Omega^{\mathbf{P}}, \mathbf{P}^{\mathbf{P}}, \Phi^{\mathbf{P}} \rangle$, which is defined analogously to a total structure except for the fact that the relations in $\mathbf{P}^{\mathbf{P}}$ are partial relations on $\Omega^{\mathbf{P}}$ and the functions in $\Phi^{\mathbf{P}}$ are partial functions from $(\Omega^{\mathbf{P}})^{n}$ to $\Omega^{\mathbf{P}}$. A *partial relation*, ${}^{n}R_i^{\mathbf{P}}$, is a triple $\langle sat({}^{n}R_i^{\mathbf{P}}), dissat({}^{n}R_i^{\mathbf{P}}), indet({}^{n}R_i^{\mathbf{P}}) \rangle$, where $sat({}^{n}R_i^{\mathbf{P}})$ is the set of *n*-tuples of elements of $\Omega^{\mathbf{P}}$ that satisfy the relation ${}^{M}R_i^{\mathbf{P}}$, $dissat({}^{n}R_i^{\mathbf{P}})$ is the set of *n*-tuples of elements of $\Omega^{\mathbf{P}}$ that do not satisfy ${}^{n}R_i^{\mathbf{P}}$, and $indet({}^{n}R_i^{\mathbf{P}})$ is the set of *n*-tuples of elements for which it is indeterminate whether or not they satisfy ${}^{n}R_i^{\mathbf{P}}$. A total relation can be thus seen as a limit case of partial relation—a partial relation whose third component is the empty set. A *partial function* is a function that may not assign any value to some arguments within its domain (or, if you prefer, assigns a "null" value '†').[37]

We can now introduce the notion of partial substructure.

(29) A partial structure $\mathbf{B}$ is *a partial substructure of* a total structure $\mathbf{A}$ if and only if:

[37] It is worth noting that my definition of a partial structure differs from the notion employed by da Costa, French, and their collaborators in one aspect, which will turn out to be crucial to my account. Da Costa and French restrict their analysis to structures that contain relations but not functions. Obviously, this is not a problem in itself. Any *n*-ary function can be construed as an (*n*+1)-ary relation whose first *n relata* are the arguments of the function and whose last *relatum* is the value of the function. In fact, a partial unary function, *f*(*x*), that does not assign a value to a certain argument, *a*, in its domain can be seen as a relation that is undefined for any ordered pair whose first component is *a* and whose second component is an element of the co-domain of the function.

[29.1] The universe of $\mathbf{B}$, $\Omega^{\mathbf{B}}$, is a (non-empty) subset of the universe of $\mathbf{A}$, $\Omega^{\mathbf{A}}$.

[29.2] For every *n*-ary relation, ${}^{n}R_i^{\mathbf{A}} \in \mathbf{P}^{\mathbf{A}}$, there is an *n*-ary relation, $<sat({}^{n}R_i^{\mathbf{B}}), indet({}^{n}R_i^{\mathbf{B}}), dissat({}^{n}R_i^{\mathbf{B}})> \in \mathbf{P}^{\mathbf{B}}$ such that $sat({}^{n}R_i^{\mathbf{B}}) \subseteq ({}^{n}R_i^{\mathbf{A}} \cap (\Omega^{\mathbf{B}})^{n})$ and $dissat({}^{n}R_i^{\mathbf{B}}) \cap {}^{n}R_i^{\mathbf{A}} = \varnothing$. (Informally, a certain *n*-tuple of elements of the universe of $\mathbf{B}$ belongs to the set of *n*-tuples that satisfy a certain relation in $\mathbf{B}$ *only if* it is in the extension of the corresponding relation in $\mathbf{A}$ and it belongs to the set of *n*-tuples that do not satisfy that relation *only if* it is not in the extension of the corresponding relation in $\mathbf{A}$.)[38]

[29.3] For every *n*-ary function, ${}^{n}F_j^{\mathbf{A}} \in \Phi^{\mathbf{A}}$, there is a partial *n*-ary function ${}^{n}F_j^{\mathbf{B}} \in \Phi^{\mathbf{B}}$ such, that, if $dom({}^{n}F_j^{\mathbf{A}})$ is the domain of ${}^{n}F_j^{\mathbf{A}}$ and $codom({}^{n}F_j^{\mathbf{A}})$ is the codomain of ${}^{n}F_j^{\mathbf{A}}$, $dom({}^{n}F^{\mathbf{B}}_i) = (dom({}^{n}F^{\mathbf{A}}_i) \cap (\Omega^{\mathbf{B}})^{n})$ and $codom({}^{n}F^{\mathbf{B}}_i) = codom({}^{n}F^{\mathbf{A}}_i) \cap (\Omega^{\mathbf{B}})^{n}$ and such that for every $<o_1^{\mathbf{B}}, \ldots, o_k^{\mathbf{B}}> \in (\Omega^{\mathbf{B}})^{n}$, either ${}^{n}F_j^{\mathbf{A}}(o_1^{\mathbf{B}}, \ldots, o_k^{\mathbf{B}}) = {}^{n}F_j^{\mathbf{B}}(o_1^{\mathbf{B}}, \ldots, o_k^{\mathbf{B}})$ or ${}^{n}F_j^{\mathbf{B}}(o_1^{\mathbf{B}}, \ldots, o_k^{\mathbf{B}})$ is indeterminate. (Informally, for every function in $\mathbf{A}$, there is a partial function in $\mathbf{B}$ whose domain is the subset of the domain of the function in $\mathbf{A}$ that only contains the *n*-tuples of elements that are in the universe of $\mathbf{B}$ and whose codomain is the subset of the domain of the function in $\mathbf{A}$ that only contains the *n*-tuples of elements that are in the universe of $\mathbf{B}$ and that either assigns to every n-tuple in the universe of $\mathbf{B}$ the same value the corresponding function in $\mathbf{A}$ assigns to that same n-tuple or it assigns an indeterminate value.)

Since the isomorphism between the partial substructures of the relevant structure of the vehicle, $\mathbf{V}$, and that of the target, $\mathbf{T}$, is an *intended* isomorphism, all of the intended isomorphic partial substructures—i.e. all of their partial substructures, $\mathbf{V}^*$ and $\mathbf{T}^*$, such that an intended isomorphism holds between them—must meet the following conditions:

(i) If $o_i^{\mathbf{V}}$ denotes $o_i^{\mathbf{T}}$, $o_i^{\mathbf{V}^*} \in \Omega^{\mathbf{V}^*}$ and $o_i^{\mathbf{T}^*} \in \Omega^{\mathbf{T}^*}$ only if $o_i^{\mathbf{V}} \in \Omega^{\mathbf{V}}$ and $o_i^{\mathbf{T}} \in \Omega^{\mathbf{T}}$.

(ii) If $o_1^{\mathbf{V}}$ denotes $o_1^{\mathbf{T}}$, …, $o_n^{\mathbf{V}}$ denotes $o_n^{\mathbf{T}}$, ${}^{n}R_k^{\mathbf{V}}$ denotes ${}^{n}R_k^{\mathbf{T}}$, $o_1^{\mathbf{V}^*} \in \Omega^{\mathbf{V}^*}$, …., $o_n^{\mathbf{V}^*} \in \Omega^{\mathbf{V}^*}$, then:

(ii.a) $<o_1^{\mathbf{V}^*}, \ldots, o_n^{\mathbf{V}^*}> \in sat({}^{n}R_k^{\mathbf{V}^*})$ and $<o_1^{\mathbf{T}^*}, \ldots, o_n^{\mathbf{T}^*}> \in sat({}^{n}R_k^{\mathbf{T}^*})$ only if $<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}> \in {}^{n}R_k^{\mathbf{V}}$ and $<o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}}> \in {}^{n}R_k^{\mathbf{T}}$,

(ii.b) $<o_1^{\mathbf{V}^*}, \ldots, o_n^{\mathbf{V}^*}> \in dissat({}^{n}R_k^{\mathbf{V}^*})$ and $<o_1^{\mathbf{T}^*}, \ldots, o_n^{\mathbf{T}^*}> \in sat({}^{n}R_k^{\mathbf{T}^*})$ only if $<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}> \notin {}^{n}R_k^{\mathbf{V}}$ and $<o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}}> \notin {}^{n}R_k^{\mathbf{T}}$, *and*

(ii.c) $<o_1^{\mathbf{V}^*}, \ldots, o_n^{\mathbf{V}^*}> \in indet({}^{n}R_k^{\mathbf{V}^*})$ and $<o_1^{\mathbf{V}^*}, \ldots, o_n^{\mathbf{V}^*}> \in dissat({}^{n}R_k^{\mathbf{V}^*})$ only if either $<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}> \in {}^{n}R_k^{\mathbf{V}}$ and $<o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}}> \notin {}^{n}R_k^{\mathbf{T}}$, or $<o_1^{\mathbf{V}}, \ldots, o_n^{\mathbf{V}}> \notin {}^{n}R_k^{\mathbf{V}}$ and $<o_1^{\mathbf{T}}, \ldots, o_n^{\mathbf{T}}> \in {}^{n}R_k^{\mathbf{T}}$.

---

[38] Let me note that this is markedly different from how French and his collaborators define these concepts. For example, French and Ladyman (1999) claim that a certain *n*-tuple of elements of the universe of $\mathbf{B}$ belongs to the set of *n*-tuples that satisfy a certain relation in $\mathbf{B}$ *if* and only if it is in the extension of the corresponding relation in $\mathbf{A}$; it belongs to the set of *n*-tuples that do not satisfy that relation *if* and only if it is not in the extension of the corresponding relation in $\mathbf{A}$. However, this cannot possibly be what French and Ladyman actually have in mind because, according to this definition, the set $indet({}^{n}R_i^{\mathbf{P}})$ would always be empty and therefore the set of partial substructures of $\mathbf{A}$ would be identical to that of the substructures of $\mathbf{A}$. In order for $\mathbf{B}$ to be a (genuine) *partial* substructure of $\mathbf{A}$ it must be the case that $indet({}^{n}R_i^{\mathbf{P}})$ is not empty. This is accomplished by the definition I have put forward here but not by the one that French and Ladyman propose.

(iii) If o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, o_i^{V} denotes o_i^{T}, $''F_k^{V}$ denotes $''F_k^{T}$, $o_1^{V*}\in\Omega^{V*}$, …., $o_n^{V*}\in A^{V*}$, $o_1^{T*}\in A^{T*}$, …., and $o_n^{T*}\in\Omega^{T*}$, then:

(iii.a) $''F_k^{V*}(o_1^{V*}, \ldots, o_n^{V*})=o_i^{V*}$ and $''F_k^{T*}(o_1^{T*}, \ldots, o_n^{T*})=o_i^{T*}$ only if $''F_k^{V}(o_1^{V}, \ldots, o_n^{V})=o_i^{V}$ and $''F_k^{T}(o_1^{T}, \ldots, o_n^{T})=o_i^{T}$, and

(iii.b) $''F_k^{V*}(o_1^{V*}, \ldots, o_n^{V*})$ and $''F_k^{T*}(o_1^{T*}, \ldots, o_n^{T*})$ are indeterminate only if $''F_k^{V}(o_1^{V}, \ldots, o_n^{V})\neq o_i^{V}$ or $''F_k^{T}(o_1^{T}, \ldots, o_n^{T})\neq o_i^{T}$.

Unlike the two other proposals I have considered, the partial isomorphism account of partially faithful epistemic representation can handle both incorrect and incomplete epistemic representations.

Incorrectness. A vehicle is an incorrect epistemic representation of a target if and only if, either for some $o_i^{V}\in\Omega^{V}$, o_i^{V} denotes o_i^{T} and $o_i^{T}\notin\Omega^{T}$; or else, for some $''R_k^{V}$, o_1^{V}, …, and o_n^{V}, o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, $''R_k^{V}$ denotes $''R_k^{T}$ and $<o_1^{V}, \ldots, o_n^{V}>\in{}''R_k^{V}$, but $<o_1^{T}, \ldots, o_n^{T}>\notin{}''R_k^{T}$. However, an intended partial isomorphism can hold between the vehicle and the target even if, for some $o_i^{V}\in\Omega^{V}$, o_i^{V} denotes o_i^{T} and $o_i^{T}\notin\Omega^{T}$ or, for some $''R_k^{V}$, o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, $''R_k^{V}$ denotes $''R_k^{T}$ and $<o_1^{V}, \ldots, o_n^{V}>\in{}''R_k^{V}$, but $<o_1^{T}, \ldots, o_n^{T}>\notin{}''R_k^{T}$. For an intended partial isomorphism to hold between the vehicle and the target, an object can be an element of the universe of the relevant structure of the vehicle even if the object that denotes it is not in the relevant structure of the target—the first object will simply not be in the universe of the isomorphic partial substructure of the vehicle. Analogously, for an intended partial isomorphism to hold between the vehicle and the target, it is not necessary that, if o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, $''R_k^{V}$ denotes $''R_k^{T}$, and $<o_1^{V}, \ldots, o_n^{V}>\in{}''R_k^{V}$, then $<o_1^{T}, \ldots, o_n^{T}>\in{}''R_k^{T}$. If $<o_1^{V}, \ldots, o_n^{V}>\in{}''R_k^{V}$ and $<o_1^{T}, \ldots, o_n^{T}>\notin{}''R_k^{T}$, it will simply be the case that $<o_1^{V*}, \ldots, o_n^{V*}>\in indet(''R_k^{V*})$ and $<o_1^{T*}, \ldots, o_n^{T*}>\in indet(''R_k^{T*})$. So, if an intended partial isomorphism holds between the vehicle and the target, the vehicle can be an incorrect representation of the target.

Incompleteness. A vehicle is an incomplete epistemic representation of a target if and only if, for some $o_i^{T}\in\Omega^{T}$, o_i^{V} denotes o_i^{T} and $o_i^{V}\notin\Omega^{V}$, or, for some $''R_k^{T}$, o_1^{T}, …, and o_n^{T}, o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, $''R_k^{V}$ denotes $''R_k^{T}$ and $<o_1^{T}, \ldots, o_n^{T}>\in{}''R_k^{T}$, but $<o_1^{V}, \ldots, o_n^{V}>\notin{}''R_k^{V}$. However, an intended partial isomorphism can hold between the vehicle and the target even if, for some $o_i^{T}\in\Omega^{T}$, o_i^{V} denotes o_i^{T} and $o_i^{V}\notin\Omega^{V}$ or, for some $''R_k^{T}$, o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, $''R_k^{V}$ denotes $''R_k^{T}$ and $<o_1^{T}, \ldots, o_n^{T}>\in{}''R_k^{T}$, but $<o_1^{V}, \ldots, o_n^{V}>\notin{}''R_k^{V}$. For an intended partial isomorphism to hold between the vehicle and the target, an object can be an element of the universe of the relevant structure of the target even if the object that denotes it is not in the relevant structure of the vehicle—the first object will simply not be in the universe of the isomorphic partial substructure of the target. Analogously, for an intended partial isomorphism to hold between the vehicle and the target, it is not necessary that, if o_1^{V} denotes o_1^{T}, …, o_n^{V} denotes o_n^{T}, $''R_k^{V}$ denotes $''R_k^{T}$ and $<o_1^{T}, \ldots, o_n^{T}>\in{}''R_k^{T}$, $<o_1^{V}, \ldots, o_n^{V}>\in{}''R_k^{V}$. If $<o_1^{T}, \ldots, o_n^{T}>\in{}''R_k^{T}$ and $<o_1^{V}, \ldots, o_n^{V}>\notin{}''R_k^{V}$, it will simply be the case that $<o_1^{V*}, \ldots, o_n^{V*}>\in indet(''R_k^{V*})$ and $<o_1^{T*}, \ldots, o_n^{T*}>\in indet(''R_k^{T*})$. So, if an intended partial isomorphism holds between the relevant structure of the target and that of the vehicle, the vehicle can be an incomplete representation of the target.

The partial isomorphism account is not only able to deal with cases of incorrect and incomplete representation such as that of the old London Underground and today's network, but it also allows us to explain why the other structural accounts were successful (insofar as they were successful). Isomorphism, homomorphism, and Δ/Ψ-morphism are all limit cases of partial isomorphism. Isomorphism is the case in which the isomorphic partial substructures of the relevant structures of the vehicle and the target are the full structures themselves (i.e. the universe of both structures are identical with the universe of their partial substructures and, for all relations, the set of *n*-tuples for which it is undetermined whether the *n*-tuples satisfy or do not satisfy the structure in question is empty). Homomorphism is the case in which the isomorphic partial substructure of the vehicle is the full relevant structure of the vehicle. **Δ/Ψ**-morphism is the case in which the isomorphic partial substructure of the vehicle and the target contain some relations that are indeterminate for all *n*-tuples of elements. The partial isomorphism account is thus able to account why the isomorphism account, homomorphism account, and **Δ/Ψ**-morphism account were successful, insofar as they were.

7.10 Partial Isomorphism and Inexactness

As we have seen, the partial isomorphism account can handle both *incorrectness* and *incompleteness*. In this section, however, I argue that it is not equally successful at dealing with *inexactness*. Inexactness is particularly important for our purposes because it seems to be a pervasive sort of unfaithfulness in models from mathematized sciences. We have seen this to be the case with the case of the inclined plane model and the tobogganing hill. In that case, we had reason to believe that the model was inexact in the sense that, for example, some of the external forces acting on the box are not the same as the corresponding forces acting on my daughters and their toboggan (one example of this is air friction, which is zero on the box, but significant on my daughters and their toboggan).

To be clear, the problem is not that the partial isomorphism account cannot somehow accommodate inexact epistemic representations, for the partial isomorphism account does have a strategy for dealing with inexact epistemic representations. The problem, I think, is that the strategy in question is inadequate. The strategy simply treats cases of inexactness as cases of incorrectness. One way to implement this strategy is to postulate that all functions with an indeterminate value have the same value, so that, if ${}^{n}F_{k}^{\mathtt{V}}$ denotes ${}^{n}F_{k}^{\mathtt{T}}$, $o_1^{\mathtt{V}}$ denotes $o_1^{\mathtt{T}}$, …, and $o_n^{\mathtt{V}}$ denotes $o_n^{\mathtt{T}}$, but there is not an intended isomorphism such that $f({}^{n}F_{k}^{\mathtt{V}*}(o_1^{\mathtt{V}*}, \ldots, o_n^{\mathtt{V}*})) = {}^{n}F_{k}^{\mathtt{T}*}(f(o_1^{\mathtt{V}*}), \ldots, f(o_n^{\mathtt{V}*}))$, then an intended isomorphism can still hold between partial substructures that assign indeterminate values to ${}^{n}F_{k}^{\mathtt{V}*}(o_1^{\mathtt{V}*}, \ldots, o_n^{\mathtt{V}*})$ and ${}^{n}F_{k}^{\mathtt{T}*}(o_1^{\mathtt{V}*}, \ldots, o_n^{\mathtt{V}*})$. The reason why this way of handling inexactness treats it as a kind of incorrectness is that an *n*-ary function ${}^{n}F_{k}^{\mathtt{A}}(o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}) = o_i^{\mathtt{A}}$ can be construed as an $(n+1)$-ary relation $< o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}, o_i^{\mathtt{A}} > \in {}^{(n+1)}R_{k}^{\mathtt{A}}$, and a partial function whose value is indeterminate for the arguments $o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}$ can be seen as a partial relation $< o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}, o_i^{\mathtt{A}} > \in indet({}^{(n+1)}R_{k}^{\mathtt{A}})$ for all $o_i^{\mathtt{A}}$.

However, while treating inexactness as a kind of incorrectness may be satisfactory in *some* cases of inexactness, it is not appropriate in *all* cases. In using scientific models, for example, it is convenient to distinguish between *approximations* and *idealizations*. While it is possible to think of idealization as a kind of approximation, the rationale behind approximations and idealizations seems to be slightly different. For example, the rationale behind not including in the inclined plane model a counterpart for forces such as the gravitational attraction of distant galaxies on the racer seems to be that their magnitude, while different from zero, is so close to zero that the effect of these forces on my daughters and their toboggan would likely be smaller than we are able to detect. Their influence, then, will be negligible, and these forces can be left out of the model. The rationale behind not including in the model a counterpart for forces such as the air friction, on the other hand, is that these forces, even if not negligible, are very complicated to model and, often, the gain in faithfulness that derives from including a counterpart for them in the model is not worth the effort it takes to do so.

Here I call cases of the first kind '*approximations*' and cases of the second kind '*idealizations*'.[39] My contention is that, while treating inexactness as incorrectness may be satisfactory in cases of idealization, it is not satisfactory in cases of approximation. When it comes to approximations, it *does* matter that, even if $f({}^{n}F_{k}^{\mathtt{V}}\ (o_{1}^{\mathtt{V}}, \ldots, o_{n}^{\mathtt{V}}) \neq {}^{n}F_{k}^{\mathtt{T}}(f(o_{1}^{\mathtt{V}}), \ldots, f(o_{n}^{\mathtt{V}}))$, $f({}^{n}F_{k}^{\mathtt{V}}(o_{1}^{\mathtt{V}}, \ldots, o_{n}^{\mathtt{V}})$ and ${}^{n}F_{k}^{\mathtt{T}}(f(o_{1}^{\mathtt{V}}), \ldots, f(o_{n}^{\mathtt{V}}))$ are nevertheless "close enough". Suppose, for example, that I am comparing the standard inclined plane model with an inclined plane model with air friction. While both models might be inexact when it comes to air friction, the inclined plane model with air friction, which makes the force due to air friction proportional to the cross-sectional area of my seated daughters and to the square of their velocity, seems to be more faithful than the standard inclined plane model, which sets the external force for air friction to zero, for the former model better approximates the force my daughter would experience due to air friction.

More generally, the reason for distinguishing between incorrectness and inexactness is that the latter comes in degrees while the former is an all-or-nothing matter. A user can draw conclusions that are strictly speaking false from inexact and incorrect epistemic representations alike. However, inexact epistemic representations, unlike incorrect epistemic representations, may allow their users to draw conclusions that are "closer to the truth" or "farther from the truth" (in the sense outlined in Section 2.2).

An account of partially faithful epistemic representation that construes inexactness as a sort of incorrectness, however, would classify all inexact representations as equally unfaithful. If inexactness is construed as a species of incorrectness, for example, a model that gets the velocity of the toboggan almost completely right and one that grossly underestimates it would count as equally faithful epistemic representations, and a geographical map that employs the Lambert Conformal Conic projection (and thus distorts slightly most features of the territory) would be classified as being as unfaithful as one that grossly distorts most features of the same territory. The problem of inexactness, which is so crucial in the case of models (as well as in many other cases of epistemic representation), can only be dealt with adequately by an account that acknowledges the specific nature of inexactness and its distinctness from incorrectness. I develop such an account in the next chapter. Before moving on to that, however, I am going to argue that there is a more profound reason to be dissatisfied with the structural account of faithful epistemic representation—the reason is that the structural account fails to accommodate the fact that faithfulness is a matter of degree.

7.11 How Little Faithfulness is Too Little Faithfulness?

Even if, as I have argued in the previous section, the partial isomorphism account cannot account satisfactorily for inexact epistemic representations, it can account for partially faithful epistemic representation that are *both* incorrect *and* incomplete. One, however, might worry that the partial isomorphism account is successful in handling incorrect and incomplete epistemic representations only because it sets the bar for partially faithful epistemic representation so low as to make it almost impossible for a vehicle not to be a partially epistemic representation of a certain target. After all, a partial isomorphism would seem to hold between the relevant structure of the vehicle and that of the target even if their intended isomorphic partial substructures are substructures whose universes contain only one element each and whose properties, relations, and functions are all indeterminate (let me call this 'a minimal partial isomorphism'). The worry is that, if the only intended partial isomorphism between the vehicle and the target is a minimal one, the vehicle would seem to be a completely *un*faithful epistemic representation of the target (i.e. one such that no valid surrogative inference from it to the target would be sound). While I

[39] Obviously, I do not mean to offer an account of approximation or idealization; nor do I intend to offer a criterion to demarcate idealizations and approximations. I use 'approximation' and 'idealization' simply as two convenient labels to distinguish between two kinds of unfaithfulness.

think this worry is ultimately misguided, I think that it reveals a deeper flaw with the structural account's approach to partially faithful epistemic representation.

Suppose that as part of an elaborate practical joke a friend gives you a fake subway map of the Tokyo subway network. Since you have never been to Tokyo and you have never seen a map of the Tokyo subway network, you can't tell that the map is a fake. But what would the fake map have to be like in order to be such that the only intended partial isomorphism that holds between it and the Tokyo subway network is a minimal one? Consider two possible answers. The first is that the map would have to be such that only one circle on the map corresponds to an actual station on the network (say, Shibuya station), while all other circles on the map stand for fictitious stations, none on the coloured lines on the map corresponds to a subway line on the network, etc. However, even if this map is *almost* a completely unfaithful epistemic representation of the network, it would still count as a partially faithful epistemic representation of the network, for one could still soundly infer from the map that one of the stations on the network is called Shibuya.

One might argue, however, that the case I have just described is not a case of minimal partial isomorphism, as the name of the station is a property of that station and, as such, it cannot be part of a minimal partial substructure like the one described above. If this view is correct (as I think it is), the only way for the partial isomorphism to be minimal would be for there to be only one circle on the map (for, if there were more circles on the map, then we could establish a partial isomorphism between the partial substructure containing those circles and some partial substructure on the network even if there were no correspondence between the properties of and relations among those circles and the properties of and relations among the stations on the network). Admittedly, the product would not look much like a subway map but, as far as I can see, it would still not be a *completely* unfaithful epistemic representation of the network. After all, one could still perform at least one sound surrogative inference from the minimal map to the network (i.e. the inference to the conclusion that there is at least one station on the network).

While some may feel that this is not enough for the fake map to be considered a partially faithful epistemic representation of the Tokyo subway network, I think that we should draw a different lesson from this sort of case. The lesson is that that the partial isomorphism account of partially faithful epistemic representation (like the other accounts of partially faithful epistemic representation that I have considered in this section) focuses on the wrong sort of question. As far as I can see, what cases such as the one of the fake map reveal is that the crucial question about partially faithful epistemic representation is not 'What makes a vehicle a partially faithful epistemic representation of a certain target (as opposed to a completely unfaithful epistemic representation of that target)?' As we have just seen, it is unclear whether there are any *completely* unfaithful epistemic representations (let alone whether it is possible to draw a clear-cut line between them and partially faithful epistemic representations that are extremely unfaithful). The crucial question an account of faithfulness should try to answer is rather 'What makes this vehicle a more faithful epistemic representation of a certain target than that other vehicle?' One of the advantages of this approach is that it allows us to sidestep the question of whether there are any completely unfaithful epistemic representations (which likely would lead to stalemate). This approach allows us to claim that, while there might be less and less faithful (and more and more unfaithful) epistemic representations of a certain target, there might be no such thing as a completely unfaithful epistemic representation. Even if, when the amount of information about the target we can validly infer from the vehicle becomes vanishingly small, we might be more reluctant to call the vehicle 'a *faithful* epistemic representation of the target'. Strictly speaking, however, a completely unfaithful epistemic representation of the target would have to be one from which no sound surrogative inferences can be validly drawn and, I suspect, no epistemic representation, no matter how unfaithful, can be so unfaithful as to meet such a stringent criterion.

I suspect that the reason why the structural account focuses on the wrong question when it comes to partially faithful epistemic representation is that morphisms are an all-or-nothing matter and, therefore, it is tempting to think that, if faithfulness is to be explained in terms of similarity of structure, faithfulness is also an all-or-nothing matter. However, as I have argued repeatedly, faithfulness is a matter of degree and not an all-or-nothing matter and our account of faithful epistemic representation should reflect this fact. In the next chapter, I develop and defend an account of faithfulness that does so—the structural similarity account of faithful epistemic representation. As its name suggests, the structural similarity account combines elements of both the similarity account and the structural account. The account is based on a notion, that of *structural similarity*, which uses some of the resources that I have developed in this chapter but reflects the fact that faithfulness is a matter of degree.

8 The Structural Similarity Account

8.1 Introduction

In this chapter, I develop what I call 'the structural similarity account (of faithful epistemic representation)'. As its name suggests, the structural similarity account combines aspects of both the structural account and the similarity account. More specifically, it employs the technical resources of the structural account to articulate in a clear and precise manner the intuitions that underlie the similarity account. As I mentioned at the end of Chapter 6, the structural account can be construed as a version of the similarity account, for a morphism is just a way to specify a very abstract and "global" sort of similarity, which, for lack of better label, might be called 'structural similarity'. The structural similarity account, however, takes this idea one step further. As we have seen at the end of Chapter 7, one of the problems with the structural account is that morphisms are an all-or-nothing matter (either a certain *x*-morphism obtains between two structures or it does not) while faithfulness, on the other hand, seems to be a matter of degree. As I have argued in Chapter 7, any epistemic representation of any target is to some extent a faithful epistemic representation of that target; the crucial question is therefore not 'What makes an epistemic representation of a certain target a faithful one (as opposed to a completely unfaithful one)?' (for, as I have argued, there might be no *completely* unfaithful epistemic representations) but rather 'What makes one of two epistemic representation of a certain target a more faithful epistemic representation than the other?'. If my arguments are correct, the structural account is therefore ill-suited to explicate the notion of faithfulness.

The structural similarity account, on the other hand, tries to explain (overall) faithfulness in terms of what I call '*(overall) structural similarity*'. More specifically, the structural similarity account of faithful epistemic representation maintains that:

(L) *v** is *(overall) a (strictly) more faithful epistemic representation* of *t* than *v* (for *u*) if and only if:

(L.1) :

(L.1.1) *u* takes both *v* and *v** to denote *t*, *and*

(L.1.2) *u* adopts an interpretation of *v* in terms of *t*, $i°(v \rightarrow t)$, and an interpretation of *v** in terms of *t*, $i°(v^* \rightarrow t)$,

(L.2) *v* and *v** have the same scope, *and*

(L.3) the relevant structure of *v** relative to $i°(v^* \rightarrow t)$, $\mathtt{V}^*$, is (overall) (strictly) more structurally similar to the relevant structure of *t* relative to $i°(v^* \rightarrow t)$, , $\mathtt{T}$, than the relevant structure of *v* relative to $i°(v \rightarrow t)$, $\mathtt{V}^*$,

and that:

(M) *v** and *v* are *(overall) equally faithful epistemic representations* of *t* (for *u*) if and only if:

(M.1) :

(M.1.1) u takes both v^* and v to denote t, *and*

(M.1.2) u adopts an interpretation of v in terms of t, $i^\circ(v \rightarrow t)$, and an interpretation of v^* in terms of t, $i^\circ(v^* \rightarrow t)$,

(M.2) v and v^* have the same scope, *and*

(M.3) the relevant structure of v^* relative to $i^\circ(v^* \rightarrow t)$, $\mathbf{V}^*$, is (overall) as structurally similar to the relevant structure of t relative to $i^\circ(v^* \rightarrow t)$, $\mathbf{T}$, as the relevant structure of v relative to $i^\circ(v \rightarrow t)$, $\mathbf{V}$.

In conjunction with (H) (i.e. what I called the isomorphism account of completely faithful epistemic representation), (M) and (L) form what I call '*the structural similarity account of faithful epistemic representation*'. In the next section, I first define the notion of (overall) structural similarity and then that of the strength of a morphism.

8.2 Morphism Strength and Structural Similarity

The two fundamental ideas that underlie the notion of structural similarity are that (i) morphisms between two structures can be (partially) ordered with respect to their strength (with isomorphism being the strongest possible morphism between them and what I called 'a minimal partial isomorphism' being possibly the weakest) and that (ii) the stronger the strongest morphism between two structures is, the more structurally similar they are. More formally,

(30) $\mathbf{A}^*$ is *(overall) (strictly) more structurally similar to* $\mathbf{B}$ *than* $\mathbf{A}$ if and only if:

[30.1] All the strongest intended morphism between $\mathbf{A}^*$ and $\mathbf{B}$ are stronger than any of the strongest intended morphism between $\mathbf{A}$ and $\mathbf{B}$.

All we need now is a general notion of the comparative strength of morphisms. I start by introducing the notion of a morphism being at least as strong as another and then I define the notion of two morphisms being equally strong, that of a morphism being (strictly) stronger/weaker than another, and that of the strongest morphism between two structures. I then discuss the role these definitions play in the structural similarity account. Here are the definitions:

(31) A morphism between $\mathbf{A}^*$ and $\mathbf{B}$ (f^*: $\Omega^{\mathbf{A}^*} \rightarrow \Omega^{\mathbf{B}}$) is *at least as strong as* a morphism between $\mathbf{A}$ and $\mathbf{B}$ (f: $\Omega^{\mathbf{A}} \rightarrow \Omega^{\mathbf{B}}$) if and only if:

[31.1] For every $o_i^{\mathbf{A}} \in \Omega^{\mathbf{A}}$, if there is an $o_i^{\mathbf{B}} \in \Omega^{\mathbf{B}}$ such that $f(o_i^{\mathbf{A}}) = o_i^{\mathbf{B}}$, then there is a $o_i^{\mathbf{A}^*} \in \Omega^{\mathbf{A}^*}$ such that $f^*(o_i^{\mathbf{A}^*}) = o_i^{\mathbf{B}}$,

[31.2] For every $o_i^{\mathbf{A}^*} \in \Omega^{\mathbf{A}^*}$, if there is no $o_i^{\mathbf{B}} \in \Omega^{\mathbf{B}}$ such that $f^*(o_i^{\mathbf{A}^*}) = o_i^{\mathbf{B}}$, then there is a $o_i^{\mathbf{A}} \in \Omega^{\mathbf{A}}$ such that there is no $o_i^{\mathbf{B}} \in \Omega^{\mathbf{B}}$ such that $f(o_i^{\mathbf{A}}) = o_i^{\mathbf{B}}$,

[31.3] The cardinality of the set of $o_i^{\mathbf{A}} \in \Omega^{\mathbf{A}}$ for which there is no $o_i^{\mathbf{B}} \in \Omega^{\mathbf{B}}$ such that $f(o_i^{\mathbf{A}}) = o_i^{\mathbf{B}}$ is greater than or equal to the cardinality of the set of $o_i^{\mathbf{A}^*} \in \Omega^{\mathbf{A}^*}$ for which there is no $o_i^{\mathbf{B}} \in \Omega^{\mathbf{B}}$ such that $f^*(o_i^{\mathbf{A}^*}) = o_i^{\mathbf{B}}$,

[31.4] For every ${}^nR_k^{\mathbf{A}^*} \in \mathbf{P}^{\mathbf{A}^*}$, ${}^nR_k^{\mathbf{A}} \in \mathbf{P}^{\mathbf{A}}$, and ${}^nR_k^{\mathbf{B}} \in \mathbf{P}^{\mathbf{B}}$, and every $<o_1^{\mathbf{A}^*}, \ldots, o_n^{\mathbf{A}^*}> \in (\Omega^{\mathbf{A}^*})^n$, $<o_1^{\mathbf{A}}, \ldots, o_n^{\mathbf{A}}> \in (\Omega^{\mathbf{A}})^n$ and $<o_1^{\mathbf{B}}, \ldots, o_n^{\mathbf{B}}> \in (\Omega^{\mathbf{B}})^n$ such that

$f(o_1^{\mathtt{A}})=o_1^{\mathtt{B}}$, …, and $f(o_n^{\mathtt{A}})=o_n^{\mathtt{B}}$ and $f^*(o_1^{\mathtt{A}*})=o_1^{\mathtt{B}}$, …, and $f^*(o_n^{\mathtt{A}*})=o_n^{\mathtt{B}}$, if $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\in{}^{n}R_k^{\mathtt{A}}$ and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\in{}^{n}R_k^{\mathtt{B}}$ then $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\in{}^{n}R_k^{\mathtt{A}*}$,

[31.5] For every ${}^{n}R_k^{\mathtt{A}}\in\mathbf{P}^{\mathtt{A}}$, ${}^{n}R_k^{\mathtt{A}*}\in\mathbf{P}^{\mathtt{A}*}$, and ${}^{n}R_k^{\mathtt{B}}\in\mathbf{P}^{\mathtt{B}}$, and every $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\in(\mathbf{\Omega}^{\mathtt{A}})^n$, $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\in(\mathbf{\Omega}^{\mathtt{A}*})^n$ and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\in(\mathbf{\Omega}^{\mathtt{B}})^n$ such that $f(o_1^{\mathtt{A}})=o_1^{\mathtt{B}}$, …, and $f(o_n^{\mathtt{A}})=o_n^{\mathtt{B}}$ and $f^*(o_1^{\mathtt{A}*})=o_1^{\mathtt{B}}$, …, and $f^*(o_n^{\mathtt{A}*})=o_n^{\mathtt{B}}$, if $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\notin{}^{n}R_k^{\mathtt{A}}$ and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\notin{}^{n}R_k^{\mathtt{B}}$ then $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\notin{}^{n}R_k^{\mathtt{A}*}$,

[31.6] For every ${}^{n}R_k^{\mathtt{A}}\in\mathbf{P}^{\mathtt{A}}$, ${}^{n}R_k^{\mathtt{A}*}\in\mathbf{P}^{\mathtt{A}*}$, and ${}^{n}R_k^{\mathtt{B}}\in\mathbf{P}^{\mathtt{B}}$, and every $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\in(\mathbf{\Omega}^{\mathtt{A}})^n$, $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\in(\mathbf{\Omega}^{\mathtt{A}*})^n$ and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\in(\mathbf{\Omega}^{\mathtt{B}})^n$ such that $f(o_1^{\mathtt{A}})=o_1^{\mathtt{B}}$, …, and $f(o_n^{\mathtt{A}})=o_n^{\mathtt{B}}$ and $f^*(o_1^{\mathtt{A}*})=o_1^{\mathtt{B}}$, …, and $f^*(o_n^{\mathtt{A}*})=o_n^{\mathtt{B}}$, if $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\in{}^{n}R_k^{\mathtt{A}*}$ and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\notin{}^{n}R_k^{\mathtt{B}}$ then $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\in{}^{n}R_k^{\mathtt{A}}$,

[31.7] For every ${}^{n}R_k^{\mathtt{A}}\in\mathbf{P}^{\mathtt{A}}$, ${}^{n}R_k^{\mathtt{A}*}\in\mathbf{P}^{\mathtt{A}*}$, and ${}^{n}R_k^{\mathtt{B}}\in\mathbf{P}^{\mathtt{B}}$, and every $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\in(\mathbf{\Omega}^{\mathtt{A}})^n$, $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\in(\mathbf{\Omega}^{\mathtt{A}*})^n$, and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\in(\mathbf{\Omega}^{\mathtt{B}})^n$ such that $f(o_1^{\mathtt{A}})=o_1^{\mathtt{B}}$, …, and $f(o_n^{\mathtt{A}})=o_n^{\mathtt{B}}$ and $f^*(o_1^{\mathtt{A}*})=o_1^{\mathtt{B}}$, …, and $f^*(o_n^{\mathtt{A}*})=o_n^{\mathtt{B}}$, if $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\notin{}^{n}R_k^{\mathtt{A}}$ and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\in{}^{n}R_k^{\mathtt{B}}$ then $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\notin{}^{n}R_k^{\mathtt{A}}$,

[31.8] For every ${}^{n}F_k^{\mathtt{A}}\in\mathbf{\Phi}^{\mathtt{A}}$, ${}^{n}F_k^{\mathtt{A}*}\in\mathbf{\Phi}^{\mathtt{A}*}$, and ${}^{n}F_k^{\mathtt{B}}\in\mathbf{\Phi}^{\mathtt{B}}$, every $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\in(\mathbf{\Omega}^{\mathtt{A}})^n$, $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\in(\mathbf{\Omega}^{\mathtt{A}*})^n$, and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\in(\mathbf{\Omega}^{\mathtt{B}})^n$, and every $o_i^{\mathtt{A}}\in\mathbf{\Omega}^{\mathtt{A}}$, $o_i^{\mathtt{A}*}\in\mathbf{\Omega}^{\mathtt{A}*}$ and $o_i^{\mathtt{B}}\in\mathbf{\Omega}^{\mathtt{B}}$ such that $f(o_1^{\mathtt{A}})=o_1^{\mathtt{B}}$, …, $f(o_n^{\mathtt{A}})=o_n^{\mathtt{B}}$, and $f(o_i^{\mathtt{A}})=o_i^{\mathtt{B}}$ and $f^*(o_1^{\mathtt{A}*})=o_1^{\mathtt{B}}$, …, $f^*(o_n^{\mathtt{A}*})=o_n^{\mathtt{B}}$, and $f^*(o_i^{\mathtt{A}*})=o_i^{\mathtt{B}}$, if ${}^{n}F_k^{\mathtt{A}}(o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}})=o_i^{\mathtt{A}}$ and ${}^{n}F_k^{\mathtt{B}}(o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}})= o_i^{\mathtt{B}}$, then ${}^{n}F_k^{\mathtt{A}*}(o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*})=o_i^{\mathtt{A}*}$,

[31.9] For every ${}^{n}F_k^{\mathtt{A}}\in\mathbf{\Phi}^{\mathtt{A}}$, ${}^{n}F_k^{\mathtt{A}*}\in\mathbf{\Phi}^{\mathtt{A}*}$, and ${}^{n}F_k^{\mathtt{B}}\in\mathbf{\Phi}^{\mathtt{B}}$, every $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\in(\mathbf{\Omega}^{\mathtt{A}})^n$, $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\in(\mathbf{\Omega}^{\mathtt{A}*})^n$, and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\in(\mathbf{\Omega}^{\mathtt{B}})^n$, and every $o_i^{\mathtt{A}}\in\mathbf{\Omega}^{\mathtt{A}}$, $o_i^{\mathtt{A}*}\in\mathbf{\Omega}^{\mathtt{A}*}$ and $o_i^{\mathtt{B}}\in\mathbf{\Omega}^{\mathtt{B}}$ such that $f(o_1^{\mathtt{A}})=o_1^{\mathtt{B}}$, …, $f(o_n^{\mathtt{A}})=o_n^{\mathtt{B}}$, and $f(o_i^{\mathtt{A}})=o_i^{\mathtt{B}}$ and $f^*(o_1^{\mathtt{A}*})=o_1^{\mathtt{B}}$, …, $f^*(o_n^{\mathtt{A}*})=o_n^{\mathtt{B}}$, and $f^*(o_i^{\mathtt{A}*})=o_i^{\mathtt{B}}$, if ${}^{n}F_k^{\mathtt{A}*}(o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*})\neq o_i^{\mathtt{A}*}$ and ${}^{n}F_k^{\mathtt{B}}(o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}})=o_i^{\mathtt{B}}$, then ${}^{n}F_k^{\mathtt{A}}(o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}})\neq o_i^{\mathtt{A}}$,

[31.10] For every ${}^{n}F_k^{\mathtt{A}}\in\mathbf{\Phi}^{\mathtt{A}}$, ${}^{n}F_k^{\mathtt{A}*}\in\mathbf{\Phi}^{\mathtt{A}*}$, and ${}^{n}F_k^{\mathtt{B}}\in\mathbf{\Phi}^{\mathtt{B}}$, every $<o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}}>\in(\mathbf{\Omega}^{\mathtt{A}})^n$, $<o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*}>\in(\mathbf{\Omega}^{\mathtt{A}*})^n$, and $<o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}}>\in(\mathbf{\Omega}^{\mathtt{B}})^n$, and every $o_i^{\mathtt{A}}\in\mathbf{\Omega}^{\mathtt{A}}$, $o_i^{\mathtt{A}*}\in\mathbf{\Omega}^{\mathtt{A}*}$ and $o_i^{\mathtt{B}}\in\mathbf{\Omega}^{\mathtt{B}}$ such that $f(o_1^{\mathtt{A}})=o_1^{\mathtt{B}}$, …, $f(o_n^{\mathtt{A}})=o_n^{\mathtt{B}}$, and $f^*(o_1^{\mathtt{A}*})=o_1^{\mathtt{B}}$, …, $f^*(o_n^{\mathtt{A}*})=o_n^{\mathtt{B}}$, if ${}^{n}F_k^{\mathtt{A}}(o_1^{\mathtt{A}}, \ldots, o_n^{\mathtt{A}})=o_i^{\mathtt{A}}$, ${}^{n}F_k^{\mathtt{A}*}(o_1^{\mathtt{A}*}, \ldots, o_n^{\mathtt{A}*})=o_i^{\mathtt{A}*}$, and ${}^{n}F_k^{\mathtt{B}}(o_1^{\mathtt{B}}, \ldots, o_n^{\mathtt{B}})=o_i^{\mathtt{B}}$, then:

[31.10.1] if $f(o_i^{\mathtt{A}})$, $f^*(o_i^{\mathtt{A}*})$, and $o_i^{\mathtt{B}}$ are scalars, then $(f^*(o_i^{\mathtt{A}*})-o_i^{\mathtt{B}})^2\leq(f(o_i^{\mathtt{A}})-o_i^{\mathtt{B}})^2$,

[31.10.2] if $f(o_i^{\mathtt{A}})$, $f^*(o_i^{\mathtt{A}*})$, and $o_i^{\mathtt{B}}$ are vectors (and '$||v||$' denotes the norm of v), then $(||f^*(o_i^{\mathtt{A}*})-o_i^{\mathtt{B}}||)^2\leq(||f(o_i^{\mathtt{A}})-o_i^{\mathtt{B}}||)^2$.

(32) A morphism between $\mathtt{A}^*$ and $\mathtt{B}$ (f^*: $\Omega^{\mathtt{A}*}\rightarrow\Omega^{\mathtt{B}}$) and a morphism between $\mathtt{A}$ and $\mathtt{B}$ (f: $\Omega^{\mathtt{A}}\rightarrow\Omega^{\mathtt{B}}$) *are equally strong* if and only if:

[32.1] f^* is at least as strong as f, *and*

[32.2] f is at least as strong as f^*.

(33) A morphism between $\mathtt{A}^*$ and $\mathtt{B}$ (f^*: $\Omega^{\mathtt{A}^*}\rightarrow\Omega^{\mathtt{B}}$) is *(strictly) stronger than* morphism between $\mathtt{A}$ and $\mathtt{B}$ (f: $\Omega^{\mathtt{A}}\rightarrow\Omega^{\mathtt{B}}$) if and only if:

[33.1] f^* is at least as strong as f *and*

[33.2] f and f^* are not equally strong.

(34) A morphism between $\mathbb{A}$ and $\mathbb{B}$ (f: $\Omega^{\mathbb{A}} \rightarrow \Omega^{\mathbb{B}}$) is *(strictly) weaker than* morphism between $\mathbb{A}^*$ and $\mathbb{B}$ (f^*: $\Omega^{\mathbb{A}^*} \rightarrow \Omega^{\mathbb{B}}$) if and only if

[34.1] f^* is (strictly) stronger than f.

Finally,

(35) A morphism between $\mathbb{A}$ and $\mathbb{B}$ (f^*: $\Omega^{\mathbb{A}} \rightarrow \Omega^{\mathbb{B}}$) is one of the strongest morphisms between $\mathbb{A}$ and $\mathbb{B}$ if and only if:

[35.1] For any morphism between $\mathbb{A}$ and $\mathbb{B}$ (f: $\Omega^{\mathbb{A}} \rightarrow \Omega^{\mathbb{B}}$), f^* is at least as strong as f.

Since definition (31) plays a pivotal role here, let me give a brief informal explanation for each clause. Condition [31.1] and [31.2] require, respectively, that every object in $\mathbb{B}$ that has a counterpart in $\mathbb{A}$ also has a counterpart in $\mathbb{A}^*$ and that every object in $\mathbb{B}$ that has no counterpart in $\mathbb{A}^*$ also has no counterpart in $\mathbb{A}$. These conditions are there to ensure that insofar as the universe of $\mathbb{A}$ "mirrors" the universe of $\mathbb{B}$, so does the universe of $\mathbb{A}^*$, and that insofar as the universe of $\mathbb{A}^*$ fails to "mirror" the universe of $\mathbb{B}$, so does the universe of $\mathbb{A}$. Condition [31.3] requires that the cardinality of the set of objects in $\mathbb{A}$ that have no counterpart in $\mathbb{B}$ is greater than or equal to the cardinality of the set of objects in $\mathbb{A}^*$ that have no counterpart in $\mathbb{B}$. This is to ensure that there is no more "excess structure" in $\mathbb{A}^*$ than there is in $\mathbb{A}$. Conditions [31.4], [31.5], [31.6], and [31.7] require that, insofar as the extension of a property (or relation) in $\mathbb{B}$ is "mirrored" by the extension of the corresponding property (or relation) in $\mathbb{A}$, it is also "mirrored" by the extension of the corresponding property (or relation) in $\mathbb{A}^*$ and that, insofar as the extension of a property (or relation) in $\mathbb{B}$ is not "mirrored" by the extension of the corresponding property (or relation) in $\mathbb{A}^*$, it is not "mirrored" by the extension of the corresponding property (or relation) in $\mathbb{A}$ either. Conditions [31.8] and [31.9] require that, insofar a function in $\mathbb{A}$ "mirrors" the corresponding function in $\mathbb{B}$, so does the corresponding function in $\mathbb{A}^*$ and, insofar as a function in $\mathbb{A}^*$ fails to "mirror" the corresponding function in $\mathbb{B}$, so does the corresponding function in $\mathbb{A}$. Condition [31.10] requires that every function in $\mathbb{B}$ that takes a scalar or a vector as its value is approximated by the corresponding function in $\mathbb{A}^*$ at least as well as the corresponding function in $\mathbb{A}$.

Before proceeding, let me note that condition [31.10], as formulated here, is likely to be neither sufficiently general nor sufficiently fine-grained for a fully satisfactory account of structural similarity. First of all, the condition only applies to scalar and vector quantities; however, many mathematical objects other than scalars and vectors are used to represent the behaviour of objects in mathematical models, so condition [31.10] is likely not sufficiently general to adequately account for *all* cases of approximation. Second, in many cases, the requirement is likely to be too coarse-grained for our purposes. As I have argued in Chapter 2, closeness to truth is not simply a matter of approximation. A floor plan that represents a 5.9m wide room as being 6m wide might be more accurate than one that represents it as being 5.5m wide but, if all we need to do is to figure out whether a 6m wide bookshelf fits in the room, then the former representation is no closer to the truth than the latter. So it might be advisable to formulate subcondition [31.10.1] as 'if $f(o_i^{\mathbb{A}})$, $f^*(o_i^{\mathbb{A}^*})$, and $o_i^{\mathbb{B}}$ are scalars, then $f(o_i^{\mathbb{A}}) \geq f^*(o_i^{\mathbb{A}^*}) \geq (o_i^{\mathbb{B}})$ or $f(o_i^{\mathbb{A}}) \leq f^*(o_i^{\mathbb{A}^*}) \leq (o_i^{\mathbb{B}})$'. Since making analogous adjustments to subcondition [31.10.2] would be much more complicated and would require an in-depth discussion of closeness to truth in relation to vector quantities and since condition [31.10], as formulated here, performs equally well in a large number of cases, however, I rely on [31.10] here.

In this section, I have introduced all the notions that are needed to formulate the structural similarity account. In the next few sections, I make a few preliminary general comments on the structural similarity account, I illustrate how the account handles the two main examples of epistemic representation I have discussed so far, and explain on some technical details of the above definitions.

8.3 Faithfulness and Structural Similarity

Like the similarity account and unlike the structural account, the structural similarity account can account for the fact that faithfulness is a matter of degree and not an all-or-nothing matter. Like the similarity account, the structural similarity account does so by appealing to the notion of similarity, but, unlike the similarity account, the structural similarity account provides us with a clear definition of the notion of similarity involved. In doing so, the account vindicates the intuition that similarity plays a role in representation, while avoiding the pitfalls associated with the less sophisticated ways of expressing that intuition. In particular, the structural similarity account maintains that the relevant similarity is an abstract and "global" sort of similarity. What matters to faithful epistemic representation is not the nature of the things that are doing the representing or the nature of their properties and relations, but only their arrangement.

The structural similarity account also accommodates the fact that the relation '*x* is overall a (strictly) more faithful epistemic representation of *t* than *y*' is a partial order, for '*x* is overall (strictly) more structurally similar to *t* than *y*' is also a partial order. This is due both to the fact that only epistemic representations of a target that have the same scope can be compared for both faithfulness and structural similarity and to the fact that it is possible for two epistemic representations of the same target with the same scope to be such that neither of them is *strictly* more faithful (or more structurally similar) to the target than the other or as faithful (or structurally similar) to the target as the other. When this is the case, it might still be the case that one epistemic representation of the target is still (overall) more faithful (or more structurally similar) to the target the other, even if it is not *strictly* so. The example of the two maps of Rome I discussed in Chapter 2 (in which one map is only a faithful epistemic representation of a small area of Rome but not a faithful epistemic representation of the rest of Rome and the other map is a faithful epistemic of the rest of Rome but not of the area in question), for example, is a case in point. Intuitively, the second map seems to be (overall) more faithful than the first map, but, nevertheless, neither map is strictly more faithful than the other. Analogously, the second map seems to be more structurally similar to the city of Rome than the other, and yet neither is *strictly* more structurally similar to Rome than the other.

8.4 Maps, Models, and Structural Similarity

Let me now turn to how the structural similarity account handles the two main examples of epistemic representation I have discussed throughout this book. Consider first the case of the new and the old London Underground maps. Any satisfactory account of faithful epistemic representation should explain why the new map is a more faithful epistemic representation of today's network than the old one. The structural similarity account does so by appealing to the fact that the strongest intended morphism that holds between the relevant structure of the old map and that of today's network (which is a partial isomorphism) is weaker than the strongest intended morphism that obtains between the relevant structure of the new map and that of the network (which is an isomorphism). The intended isomorphism between the new map and the network must be at least as strong as the strongest intended morphism between the old map and the network, as isomorphism is the strongest morphism that can obtain between two structures. The morphism between the old map and the network, on the other hand, meets some of the conditions for being weaker than the one between the new map and the network. First, while all stations on the network have a counterpart on the new map, some of them (e.g. Borough

Station) have no counterpart on the old map (which means that the strongest intended morphism between the old map and the network does not satisfy [34.1] when compared to the strongest intended isomorphism between the new map and the network). Second, while all circles on the new map have a counterpart in the corresponding station on the network, some circles on the old map (e.g. British Museum Station) do not (which means that the strongest intended morphism between the old map and the network does not satisfy [34.2] when compared to the strongest intended isomorphism between the new map and the network). Third, while all relevant properties of and relations among stations have a counterpart in the new map, some do not have a counterpart in the old map. On the old map, for example, there is no connection between the circle marked 'Stockwell' and the one marked 'Victoria'; however, Stockwell Station and Victoria Station are today connected by Victoria Line trains (which means that the strongest intended morphism between the old map and the network does not satisfy [34.3] when compared to the strongest intended isomorphism between the new map and the network). Fourth, while all relevant properties of and relations among circles on the new map have a counterpart in properties of and relations among stations on the network, some properties of and relations among circles on the old map do not. On the old map, for example, the circle on the red line between the circles labeled 'Bank' and 'Chancery Lane' is labeled 'Post Office', but the name of the station on the Central Line between Bank and Chancery Lane Stations is 'St. Paul's' (which means that the strongest intended morphism between the old map and the network does not satisfy [34.4] when compared to the strongest intended isomorphism between the new map and the network).[40]

Consider now the toboggan case. The inclined plane model is a partially faithful epistemic representation of the tobogganing hill system. In the model, there are only two forces acting on the box. The first is the gravitational force and the second is the normal force that the plane exerts on the box. As we have already noted, however, in the system, there are many more forces acting on the toboggan than just the gravitational pull of the Earth and the normal force that the hill exerts on the toboggan. This is one of the main reasons why the inclined plane model, overall, is not a particularly faithful epistemic representation of the tobogganing hill system. However, instead of the frictionless inclined plane model, one might use a model, which, overall, is slightly more faithful—i.e. the inclined plane model with drag. In the inclined plane model with drag, a third external force acts on the box, a force that stands for the air resistance that my daughters and their toboggan would experience on their downhill journey. In the model, we can set the magnitude of the force air resistance exerts on the box equal to $-\frac{1}{2}dA(||\mathbf{v}^{\mathbb{M}^*}(b, t_i)||)^2(\mathbf{v}^{\mathbb{M}^*}(b, t_i)/|| \mathbf{v}^{\mathbb{M}^*}(b, t_i)||)$, where A is the cross-sectional area of the box (perpendicular to the direction of motion), $\mathbf{v}^{\mathbb{M}^*}(b, t_i)$ is the vector that represents the velocity of the box at t_i, and $||\mathbf{v}^{\mathbb{M}^*}(b, t_i)||$ is its norm (so that '$(\mathbf{v}^{\mathbb{M}^*}(b, t_i)/|| \mathbf{v}^{\mathbb{M}^*}(b, t_i)||$' denotes the unit vector in the direction of $\mathbf{v}^{\mathbb{M}^*}(b, t_i)$), d is a drag coefficient proportional to the air density. In the model, the air drag on the box is therefore an external force, whose direction is the opposite of the direction of the velocity of the box and whose magnitude is proportional to both the cross sectional area of the box and to the square of its velocity. So, in the inclined plane model with air friction, $\mathbf{g}^{\mathbb{M}^*}(b, t_i, 3)=-\frac{1}{2}dA(||\mathbf{v}^{\mathbb{M}^*}(b, t_i)||)^2(\mathbf{v}^{\mathbb{M}^*}(b, t_i)/||\mathbf{v}^{\mathbb{M}^*}(b, t_i)||)$ (as opposed to $\mathbf{g}^{\mathbb{M}}(b, t_i, 3)= \mathbf{0}$ in the standard inclined plane model). This means that, in the model with drag, the box gains speed more slowly than it does in the standard inclined plane model and ultimately reaches a lower terminal velocity than the one reached by the box in the standard inclined plane model.

Overall, the inclined plane model with drag would seem to be a more faithful epistemic representation of my daughters tobogganing down the hill than the inclined plane model without

---

[40] Here I assume that *being labeled 'Post Office'* and *being named 'St. Paul's'* are properties of, respectively, the circle on the old map and the station in the network. When the old map was originally printed, the Underground station that today is known as 'St. Paul's' used to be named 'Post Office'. It seems to be natural to assume that, on the standard interpretation of the old map, the circle labeled 'Post Office' on the old map still denotes St. Paul's Station while incorrectly attributing that station the property of *being named 'St. Paul's'*

drag (or so I assume here). If my daughters were to go down the hill on their toboggan, the drag they would experience due to air resistance would be one of the main factors in determining how fast they would go and, although the inclined plane with drag represents the effects of that drag somewhat crudely, it is still an improvement on the standard inclined plane model, which does not take air friction into account at all.

The structural similarity account of faithful epistemic representation vindicates the intuition that the inclined plane model with drag is a more faithful epistemic representation of my daughters tobogganing down the hill than the standard inclined plane model. According to the structural similarity account, this is due to the fact that the former is more structurally similar to my daughters tobogganing down the hill than the latter. The inclined plane model with drag trivially meets conditions [34.1]–[34.7] for being at least as structurally similar to my daughters going down the hill as the standard inclined plane model, for the relevant structures of the two models contain the same objects and do not contain any (*v*-relevant) properties or relations. Furthermore, the functions $\mathbf{g}^{\mathtt{M}}(b, t_i, 1)$ and $\mathbf{g}^{\mathtt{M}^*}(b, t_i, 2)$, and $\mathbf{g}^{\mathtt{M}}(b, t_i, 2)$ and $\mathbf{g}^{\mathtt{M}^*}(b, t_i, 2)$ trivially meet conditions [35.8]–[35.10] because the gravitational and normal forces acting on the box in the two models are the same. The inclined plane model with drag, therefore, meets the first condition ([35.1]) for being (strictly) more structurally similar to my daughters going down the hill than the inclined plane model without drag. It also meets the second condition (i.e. [35.2]) for being (strictly) more structurally similar to my daughters going down the hill than the standard inclined plane model without drag, as the latter fails to meet some of the conditions for being at least as structurally similar to the tobogganing hill system as the former. In particular, the standard inclined plane model fails to meet the conditions that require all its functions to approximate the functions in the system at least as closely as the inclined plane model with friction does. Clearly, this is not the case. Consider, for example, the vectors that represent the drag on the boxes in the two models ($\mathbf{g}^{\mathtt{M}^*}(b, t_i, 3)$ and $\mathbf{g}^{\mathtt{M}}(b, t_i, 3)$) and the one that represents the drag my daughters and their toboggan would experience at t_i ($\mathbf{g}^{\mathtt{S}}(f(b), f(t_i), f(3))$). Since $\mathbf{g}^{\mathtt{M}}(b, t_i, 3)$ is equal to $\mathbf{0}$ in the standard inclined plane model and since my daughters would experience a significant drag in their downhill journey, for all t_i, $\mathbf{g}^{\mathtt{M}^*}(b, t_i, 3)$ would approximate $\mathbf{g}^{\mathtt{S}}(f(b), f(t_i), f(3))$ more closely than $\mathbf{g}^{\mathtt{M}}(b, t_i, 3)=\mathbf{0}$ (or, more precisely, $||\mathbf{g}^{\mathtt{M}}(b, t_i, 3)-\mathbf{g}^{\mathtt{S}}(f(b), f(t_i), f(3))||^2>||\mathbf{g}^{\mathtt{M}^*}(b, t_i, 3)-\mathbf{g}^{\mathtt{S}}(f(b), f(t_i), f(3))||^2$). As a result, the functions that represent the position, velocity, and acceleration of the box at t_i in the inclined plane model with drag also approximate the position, velocity, and acceleration my daughters and their toboggan would have at t_i more closely than the corresponding functions in the standard inclined plane model. In light of the above, we can conclude that inclined plane model with drag is more structurally similar to my daughters tobogganing down the hill than the standard inclined plane model, as we would expect to be the case pre-theoretically.

I should note, once more, that the account I developed here is meant to be an account of *overall* faithfulness and not one of *specific* faithfulness. For my specific purposes (i.e. determining whether my daughters would reach a velocity that exceeds the velocity that I deem to be safe), even if one model is *overall* a (strictly) more faithful epistemic representation than the other, both might be equally specifically faithful epistemic representations of the situation, insofar as they both give a "good enough" answer to the question I happen to be interested in.

8.5 Conclusions

In this chapter, I developed and defended a structural similarity account of faithful epistemic representation. The structural similarity account, I believe, inherits the respective benefits of the similarity account and the structural account while avoiding their respective defects. In particular, like the similarity account and unlike the structural account, the structural similarity account can account for the fact that faithfulness is a matter of degree and not an all-or-nothing matter. Like the structural account and unlike the similarity account, the structural similarity account can clearly define what it is for a vehicle and a target to be similar in the relevantly abstract and "global" sense that matters when it comes to epistemic representation. So, while faithful epistemic representation is a matter of similarity and, the more similar a vehicle is to a target, the more faithfully it represents it, the similarity in question is the abstract and "global" sort of similarity that I have called structural similarity, and not a concrete, "local" similarity, as we might initially have thought. In other words, our focus on certain sorts of epistemic representations might mislead us into believing that, say, a portrait is a faithful epistemic representation of its subject in virtue of certain concrete, "local" similarities. The shape of the area that represents the nose is similar to the shape of the nose and the colour of the area that represents the eyes is similar to the colour of the eyes. However, as I have argued, these "local" similarities do not obtain in all cases of epistemic representation, and what we should focus on instead are the abstract, "global" similarities between vehicles and targets (a sort of similarity that only emerges once we interpret the vehicle in terms of the target).

Conclusions

In this book, I have argued that the so-called "problem of scientific representation" is neither a single problem nor a problem about scientific representation in particular. Instead, what is often misleadingly labelled as the problem of scientific representation are, in fact, two general and oft-conflated problems—i.e. the problem of what makes a certain vehicle an epistemic representation of a certain target and the problem of what makes a certain epistemic representation of a certain target a more or less faithful epistemic representation of that target.

Most of this book was devoted to trying to solve these two general problems. According to the account of epistemic representation that I have defended (the interpretational account of epistemic representation), a vehicle is an epistemic representation of a certain target for a certain user if and only if the user takes the vehicle to denote the target and she adopts an interpretation of the vehicle (in terms of the target). In this book, I have focussed exclusively on one specific kind of interpretation, which I have called analytic interpretation. Whether the account I have developed can be developed further into a general account of epistemic representation crucially depends on whether every possible interpretation can be reconstructed as an analytic interpretation. This is one of the crucial issues that are left open by this book and on which further work needs to be done.

One of the main advantages of the interpretational account, I have argued, is that it sheds light on the relation between epistemic representation and valid surrogative reasoning—the fact that a user adopts an interpretation of the vehicle in terms of the target (and takes the vehicle to stand for the target) is both that in virtue of which the vehicle is an epistemic representation of the target for her and that in virtue of which she can perform valid inferences from the vehicle to the target. Without the notion of an interpretation (or some analogous notion), the intimate relation between epistemic representation and valid surrogative reasoning is unnecessarily mysterious.

The notion of an analytic interpretation also plays a central role in the account of faithful epistemic representation. It directly contributes to the solutions of two crucial problems that have haunted the structural account of (faithful) epistemic representation and, indirectly, contributes to the solution of a third problem. The first problem is that of applying the notion of a morphism to objects that are not set-theoretic structures. As I have argued, the notion of an analytic interpretation provides us with a principled way to reconstruct the vehicle and the target as set-theoretic structures.

The second problem is that of determining which morphisms need to obtain between the vehicle and the target in order for the vehicle to be a faithful epistemic representation of the target (to a certain degree), since a morphism may obtain between the vehicle and the target without the first being a faithful epistemic representation of the target (on a certain interpretation of it). The notion of analytic interpretation provides us with a principled way to single out some of the morphisms that may obtain between the vehicle and the target as the intended morphisms—i.e. the only ones that are relevant to the faithfulness of the epistemic representation in question.

The third problem is that the notion of the faithfulness of an epistemic representation comes in degrees, whereas two structures are either x-morphic or they are not. However, the account of faithful epistemic representation that I have developed uses the notions of relevant structure and intended morphism to develop a third crucial notion, namely that of the structural similarity between the vehicle and the target (under a certain interpretation of the former in terms of the latter). Intuitively, the stronger the strongest intended morphism between the vehicle and the target is, the more structurally similar the vehicle and the target are. The central idea that underlies the structural similarity account is that the more structurally similar the vehicle and the target are, under a certain interpretation of the former in terms of the latter, the more faithful an epistemic representation of the latter the former is under that interpretation. The account of faithful epistemic representation that I have developed, I think, vindicates the intuitions that underlie the main accounts of faithful representation but avoids the pitfalls that characterize the other versions of these views that I have considered.

The account of epistemic representation and that of faithful epistemic representation that I have developed and defended in this book are more than just complementary—they are deeply interconnected. It is only when one attempts to develop some of the intuitions and ideas that can be found in the literature into a coherent whole that one can see how everything falls into place in the overall picture, and what might initially seem like rival attempts to solve a single problem are actually best interpreted as accounts of different notions of representation. In this book, I hope to have provided a good initial sketch of that picture.

Bibliography

Andrade, Edward Neville da Costa (1965). *Rutherford and the Nature of the Atom*, London: Heinman.

Bailer-Jones, Daniela M. (1999). 'Tracing the Development of Models in the Philosophy of Science', in L. Magnani, N. J. Nersessian und P. Thagard (Eds.), *Model-Based Reasoning in Scientific Discovery*, New York: Kluwer Academic/Plenum Publishers, 23–40.

Bailer-Jones, Daniela M. (2000). 'Modelling Extended Extragalactic Radio Sources', *Studies in History and Philosophy of Modern Physics*, 31B, S. 49-74.

Bailer-Jones, Daniela M. (2002). 'Scientists' Thoughts on Scientific Models', *Perspectives on Science* 10, 275–301.

Bailer-Jones, Daniela M. (2003). 'When Scientific Models Represent', *International Studies in Philosophy of Science*, 17, 59–74.

Balzer, Wolfgang, Moulines, Ulises C., and Sneed, Joseph D. (1987). *An Architectonic for Science: The Structuralist Program*, Dordrecht: Reidel.

Bohr, Niels (1913). 'On the Constitution of Atoms and Molecules', *Philosophical Magazine*, Series 6, 26, 1–25.

Brown, James R. (1991). *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, London: Routledge.

Callender, Craig and Jonathan Cohen (2006). 'There is No Problem of Scientific Representation', *Theoria* 55, 67–85.

Cartwright, Nancy (1983). *How The Laws of Physics Lie*, Oxford: Clarendon Press.

Cartwright, Nancy (1999a). *The Dappled World: A Study of the Limits of Science*, Cambridge: Cambridge University Press.

Cartwright, Nancy (1999b). 'Models and the Limits of Theory: Quantum Hamiltonians and the BCS Model of Superconductivity', in Morgan, M.S. and M. Morrison (eds.) (1999). *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press, 241–281.

Cartwright, Nancy, Twofic Shomar and Mauricio Suárez (1995). 'The Tool-Box of Science: Tools for the Building of Models with a Superconductivity Example', *Poznan Studies in the Philosophy of the Sciences and the Humanities* 44, 137–149.

Chakravartty, Anjan (2001): 'The Semantic or Model-Theoretic View of Theories and Scientific Realism', *Synthese*, 127, 325–345.

Contessa, Gabriele (2007). 'Scientific Representation, Interpretation, and Surrogative Reasoning', *Philosophy of Science* 74, 48–68

Contessa, Gabriele (2009). 'Who is Afraid of Imaginary Objects?' in N. Griffin & D. Jacquette (eds.), *Russell vs. Meinong: The legacy of 'on denoting'* (pp. 248–265). London: Routledge.

Contessa, Gabriele (2010). 'Scientific Models and Fictional Objects', *Synthese* 172, 215–229.

Contessa, Gabriele (2011). 'Scientific Models and Representation', in Steven French & Juha Saatsi (eds.), *The Continuum Companion to the Philosophy of Science*, London: Continuum.

Da Costa, N. C. A., & French, S. (2003). *Science and partial truth: A unitary approach to models andscientific reasoning*. Oxford: Oxford University Press.

da Costa, Newton and French, Steven (2003). *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*, Oxford: Oxford University Press.

da Costa, Newton and Steven French (1990). 'The Model-Theoretic Approach in the Philosophy of Science', Philosophy of Science 57, 248–265.

da Costa, Newton and Steven French (2000). 'Models, Theories and Structures: Thirty Years On', *Philosophy of Science* (Supplement) 67, S116–S127.

Fine, A. (1993). Fictionalism. *Midwest Studies in Philosophy, 18*, 1–18.

Fisher, Irving and Miller, O. M. (1944). *World Maps and Globes*, New York: Essential Books.

Fowler, Michael (1997). 'Rutherford Experiment and the Beginning of Nuclear Physics', Modern Physics 252 Lecture Notes, Charlottesville: University of Virignia.

French, Steve and Juha Saatsi (2004). 'Realism About Structure: The Semantic View and Non-Linguistic Representations' in *Proceedings of the Philosophy of Science Association 19th Biennial Meeting – PSA 2004: Contributed Papers*, Austin, TX.

French, Steven (2000). 'The Reasonable Effectiveness of Mathematics: Partial Structures and the Application of Group Theory to Physics', *Synthese* 125, 103–120.

French, Steven (2003). 'A Model-Theoretic Account of Representation (or I Don't Know Much about Art … But I Know It Involves Isomorphism)', *Philosophy of Science* 70, 1472–1483.

French, Steven and James Ladyman (1999). 'Reinflating the Semantic Approach', *International Studies in the Philosophy of Science,* 13, 103–121.

Friedman, Michael (1974). 'Explanation and Scientific Understanding', *The Journal of Philosophy*, 71, 5–19.

Frigg, Roman (2002). 'Models and Representation: Why Structures are not enough' *Measurement in Physics and Economics Discussion Paper Series*, DP MEAS 25/02, London: Centre for the Philosophy of Natural and Social Sciences.

Frigg, Roman (2003). *Re-presenting Scientific Representation*, Ph.D. Dissertation, London: University of London.

Frigg, Roman (2006). 'Scientific Representation and the Semantic View of Theories', *Theoria*, 55, 49–65.

Geiger, Hans and Ernest Marsden (1909). 'On Diffuse Reflection of the α-Particles', *Proceedings of the Royal Society*, 82, 495–500.

Gendler, Tamar Szabó and Hawthorne, John (2002). *Conceivability and Possibility*. Oxford, Oxford University Press.

Giere, Ronald N. (1985). 'Constructive Realism' in P.M. Churchland and C. Hooker (eds.), *Images of Science. Essays on Realism and Empiricism with a Reply from Bas C. van Fraassen*. Chicago: University of Chicago Press, 75–98.

Giere, Ronald N. (1988). *Explaining Science: A Cognitive Approach*, Chicago, University of Chicago Press.

Giere, Ronald N. (1997). *Understanding Scientific Reasoning*, 4th edition, Fort Worth: Harcourt Brace College Publishers.

Giere, Ronald N. (1999). 'Using Models to Represent Reality' in L. Magnani, N. J. Nersessian, and P. Thagard (eds.) *Model-Based Reasoning in Scientific Discovery*, New York, Kluwer/Plenum, 41–57.

Giere, Ronald N. (2004). 'How Models are Used to Represent Reality', *Philosophy of Science*, 71, 742–752.

Giere, Ronald N. (2010). An agent-based conception of models and scientific representation. *Synthese*, 172, 269–281.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21*, 725–740.

Goodman, Nelson (1968). *The Languages of Art: An Approach to a Theory of Symbols*, Indianapolis: Hackett.

Goodman, Nelson (1970). 'Seven Strictures on Similarity', in L. Foster and J.W. Swanson (eds.), *Experience and Theory*, London: Duckworth 19–29.

Hughes, R.I.G. (1997). 'Models and Representation', PSA 1996: Proceedings of the 1996 Biennial Meeting of the Philosophy of Science Association, 2, S325–S336.

Hughes, R.I.G. (1999). 'The Ising Model, Computer Simulation, and Universal Physics' in M. S. Morgan and M. Morrison (eds.), *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge, Cambridge University Press, 97–145.

Kincaid, Peter (1986). *The Rule of the Road: An International Guide to History and Practice*, New York, Greenwood Press, 1986.

Knuuttila, T. (2004). 'Models, Representation, and Mediation', in *Proceedings of the 19th Biennial Meeting of the Philosophy of Science Association*.

Lewis, David (1969). *Convention: A Philosophical Study*, Cambridge, MA.: Harvard University Press.

Lopes, Dominic (1996). *Understanding Pictures*, Oxford: Oxford University Press.

Maling, D. H. (1992). *Coordinate Systems and Map Projections* (2nd edition), Oxford: Pergamon Press.

McMullin, Ernan (1985). 'Galilean Idealization', *Studies in History and Philosophy of Science*, 16, 247–273.

Monmonier, Mark S. (1977). *Maps, Distortion, and Meaning,* Washington: Association of American Geographers.

Morgan, Mary S. (2001), 'Models, Stories and the Economic World', *Journal of Economic Methodology*, 8, 361-384.

Morgan, Mary S. and Margaret Morrison, eds. (1999). *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge, Cambridge University Press.

Morrison, Margaret (1997). 'Modelling Nature: Between Physics and the Physical World', *Philosophia Naturalis*, 35, 65–85.

Morrison, Margaret (1999). 'Models as Autonomous Agents', in M.S. Morgan and M. Morrison, eds. (1999). *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press, 38–65.

Morrison, Margaret (2008). 'Models as Representational Structures', in L. Bovens and S. Hartmann (eds.) *Nancy Cartwright's Philosophy of Science*.

Petruccioli, Sandro (1993). *Atoms, Metaphors, and Paradoxes: Niels Bohr and the Construction of a New Physics*, Cambridge: Cambridge University Press.

Putnam, Hilary (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

Rutherford, Ernest (1911). 'The Scattering of α and β Particles by Matter and the Structure of the Atom', *Philosophical Magazine*, Series 6, 21, 669–688.

Rutherford, Ernest (1914). 'The Structure of the Atom', *Philosophical Magazine (Series 6)*, 27, 488–498.

Sismondo, Sergio (1999). 'Models, Simulations, and their Objects', *Science in Context*, 12, 247–260.

Sismondo, Sergio and Nicholas Chrisman (2001) 'Deflationary Metaphysics and the Natures of Maps', *Philosophy of Science*, 68 (Proceedings), S38–S49.

Suárez, Mauricio (2002). 'The Pragmatics of Scientific Representation', CPNSS *Discussion Paper Series*, DP 66/02.

Suárez, Mauricio (2003). 'Scientific Representation: Similarity and Isomorphism' *International Studies in the Philosophy of Science*, 17, 225–244.

Suárez, Mauricio (2004). 'An Inferential Account of Scientific Representation' *Philosophy of Science*, 71, 767–779.

Suppe, Frederick (1974b). 'The Search for Philosophic Understanding of Scientific Theories' in (Suppes 1974a), 3–233.

Suppe, Frederick (1989). *The Semantic Account of Theories and Scientific Realism*, Urbana, IL.: University of Illinois Press.

Suppe, Frederick, ed. (1974a) *The Structure of Scientific Theories*, Urbana, IL.: University of Illinois Press.

Suppes, Patrick (1960). 'A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences', *Synthese*, 12, 287–301.

Suppes, Patrick (2002). *Representation and Invariance of Scientific Structures*, Stanford, CA.: CSLI Publications.

Swoyer, Chris (1991). 'Structural Representation and Surrogative Reasoning', *Synthese*, 87, 449–508.

Taylor, John R. (2005). *Classical Mechanics*. Dulles, VA: University Science Books.

Teller, Paul (2001). 'The Twilight of the Perfect Model Model' *Erkenntnis* 55, 393–415.

Thomson-Jones, M. (2009). 'Missing Systems and the Face Value Practice', *Synthese*, 127, 283–299.

van Dalen, Dirk (1994). *Logic and Structure*, 3rd edition, Berlin: Springer Verlag.

van Fraassen, Bas C. (1980). *The Scientific Image*, Oxford: Oxford University Press.

van Fraassen, Bas C. (1985) "Empiricism in the Philosophy of Science" in (Churchland and Hooker 1985), 245–308.

van Fraassen, Bas C. (1989). *Laws and Symmetry*, Oxford: Oxford University Press.

van Fraassen, Bas C. (1997) "Structure and Perspective: Philosophical Perplexity and Paradox", in M. L. Dalla Chiara et al. (eds.) *Logic and Scientific Methods*, Dordrecht: Kluwer, pp. 511–530.

van Fraassen, Bas C. (2010) *Scientific Representation: Paradoxes of Perspective*, Oxford: Oxford University Press.

Willoughby, Myrtle (2012). "Stoppism", *Journal of Philosophy*, 109: 281–294.