RomanFrigg

Fiction in Science

Published in John Woods (ed.): Fictions and Models: New Essays, Munich: Philosophia Verlag 2010, 247-287.

1. Two Kinds of Fiction

At first blush, the idea that fictions play a role in science seems to be off the mark. Realists and antirealists alike believe that science instructs us about how the world is (they part ways only over the question of what exactly science tells us about the world). Fiction not only seems to play no role in such an endeavour; it seems to detract from it. The aims of science and fiction seem to be diametrically opposed and a view amalgamating the two should be a cause of discomfort and concern.

This impression is mistaken. In fact, fictions play an essential role in many aspects of science. But what role could that be? What contribution could fictions possibly make to understanding how the world actually is? This essay aims to map out what these roles are and present a detailed analysis of one of them, the construction and use of scientific models.[1]

---

[1] My claim that fictions play an essential role in science should not be conflated with the more radical claim (often associated with postmodern-

'Fiction' means different things in different contexts, and clarifying the various uses of the term is a natural starting point for our discussion. Setting aside subtleties irrelevant to the current discussion, the different uses of 'fiction' fall into two groups which I call 'fiction as non-existence' and 'fiction as imagination'.[2] After a brief general sketch of each, I discuss whether there are fictions in each of those senses in science. My answer is affirmative: there are fictions of both kinds in science. But space constraints prevent me from discussing both in detail. This section provides a synoptic discussion of fictions of the first kind; the remainder of the chapter focuses on fictions of the second kind, which, I claim, is key to understanding how scientific models work.

The first use of 'fiction' characterises something as deviating from reality. Both sentences (propositions, statements) and entities (objects, states of affairs) can be categorised as fictions.[3] A sentence is a fiction if it is false when put forward as a claim about the world; an entity is a fiction if it does not exist. Although seemingly different, these are often the two sides of the same coin because the falsity of the proposition is due to the fact that they presuppose the

---

ism) that science *is* fiction, or, more specifically, that science is nothing but a particular kind of fiction alongside other kinds of fiction. In what follows I presuppose a broadly realist picture according to which there is a mind-independent world which has a certain structure independently of how we choose to describe it, and science aims to discover features of this world. Different positions in the debate over scientific realism diverge on how much structure there is in the world and on how much of this structure we can (possibly) come to know.

[2] Throughout this essay I use the 'to exist' in a timeless sense: Aristotle exists, the Byzantine Empire exists, and World War II exists.

[3] Classifying states of affairs as fictions – thereby expressing that they do not obtain – stretches the ordinary use of the term, but not beyond breaking point. On this understanding the state of affairs of Napoleon being a ballet dancer is a fiction not because Napoleon does not exist, but because he had no involvement in dance. In what follows I only discuss non-existent entities. This is only for the easy of discussion; what I say about entities carries over to states of affairs.

existence of entities that do not exist. For instance, the claim that Emma Bovary is 5 foot tall is a fiction because Emma Bovary does not exist, or, in other words, because Emma Bovary is a fiction (or 'fictional entity').

'Existence' here refers to physical existence - existence in space and time. Hamlet and Emma Bovary have no physical existence. Yet, there is a pervasive intuition that they somehow *are*: we think about them, make claims about them, discuss their properties, and so on, which would be not be possible if they were simply nothing. But how should we characterise the 'mode of being' of Hamlet and Emma Bovary, and how is discourse about them to be understood? This is a vexing question on which much ink has been spilled; for a survey of the different positions in this debate see Friend (2007). But since metaphysical concerns about fictional entities and issues surrounding the semantics of discourse (putatively) about them are tangential to the questions raised by fictions in science, I will bypass them here.

If we brand something as a fiction in this sense, we can do so with different intent. Two cases need to be distinguished. In the first case a fiction is a counterfeit, forgery, or fake, produced with the intention of deceiving and misleading; it is an invention deliberately opposed to fact. We say that Peter's account of the course of events is a fiction if Peter does not report truthfully how things have happened; his exasperated colleagues may at some point proclaim that the time has come to 'distinguish between fact and fiction', and if Peter then repeats his account of events they may dismiss it 'nothing but fiction'. In the second case 'fiction' refers to a supposition known (by everybody involved) to be at variance with fact, but which we nevertheless accept because it serves certain purposes. We know that Santa Claus does not exist, yet we act as if Santa came to town and organise celebrations because accepting the Santa Claus fiction serves all kinds of social functions (it is an opportunity to make gifts, gather the family, etc.). Fictions of this kind, far from being execrable, are something we cherish, and we do so exactly because they are not real.

In the second use, ‘fiction’ refers to a kind of literature, *literary fiction*, which is concerned with the narration of events and the portraiture of characters. Novels, stories, and plays are fictions in this sense.[4] This is a ‘global’ notion of fiction in that it applies to entire works, whereas the first notion of fiction can be applied piecemeal to individual sentences or entities. Rife prejudice notwithstanding, the defining feature of literary fiction is not falsity. Neither is everything that is said in, say, a novel untrue: historical novels, for instance, contain plenty of correct factual information. Nor does every text containing false reports qualify as fiction: a wrong news report or a faulty documentary do not by that token turn into fiction – they remain what they are, namely wrong factual statements. What makes a text fictional is not its truth or falsity (or a particular ratio of false to true claims), but the attitude that the reader is expected to adopt towards it. There is controversy over how exactly this attitude should be characterised, but in essence it is one of imaginary engagement. When reading a novel we are not meant to take the sentences we read as reports of fact (if we do we are simply missing the point); rather we are supposed to imagine the events described. When reading *Le Rouge et Le Noir* we are invited to imagine a plot involving a young man in emotional turmoil, having a romance with a married woman, etc.; whether there ever was a young man to whom these things happened is immaterial.

Needless to say, these senses of ‘fiction’ are not mutually exclusive, let alone independent of each other. In fact, many of the places and persons that appear in literary fiction are in fact fictions in the first sense of the term (in that they do not exist). Yet, as will become clear later, for the purpose of analysis it is helpful to keep the two separate.

[4] This notion of fiction can easily be extended to stage performances, radio plays, screenplays, movies, and different kinds of visual art. Since my focus in what follows will be on literature I do not discuss these at this point.

Let us now turn to the question of what, if any, role these different notions of fiction play in science. As I mentioned above, fictions in the second sense will be discussed in detail in Sections 2 – 4; in the remainder of this section I will briefly discuss fictions in the first sense. As we have just seen, we need to distinguish two cases. The more important and interesting one is the latter: the case of suppositions known to be at variance with fact which we nevertheless accept because they serve a certain purpose. Science is rife with fictions of this kind; in many parts of science we consider objects we know not to exist yet we keep working with them because they are useful in achieving certain goals. What goals exactly we have in mind depends on the specific scientific context, and there may be a variety of ways in which these kinds of fictions can be useful in science. Traditionally fictions have been used as calculational devices for generating predictions. In recent discussions further items have been added to this list: Bokulich (2009) emphasises the explanatory function of fictions, Suárez (2009) claims that the expediency in inference is the main defining feature of a scientific fiction, and Winsberg (2009) points out that especially in computationally intensive sciences fictions serve the purpose of extending the scope of theories beyond their traditional domain of application. As long as fictions serve an accepted goal of science, their use in science is legitimate.

A clear-cut example of the fruitfulness of fictions is D'Alembert's Principle in classical mechanics.[5] The problem we are facing is predicting the motion of particle whose path is constrained by presence of external obstacles which can change over time, for instance the motion of a marble in salad bowl that is itself being shaken. Although one can, in principle solve this problem using Newtonian mechanics, it is not advisable to do so because the mathematics gets virtually intractable even for simple constraints. To get around this problem D'Alembert introduced the concept of a *virtual displacement*, an infinitesimal but infinitely fast displacement of the particle compatible with the constraints, and he postulated that the nature of the constraints be such that the virtual

[5] See, for instance, Kuypers (1992, 13-22).

displacements do no work on the system. From this it follows that the differences between the forces acting on a system and the time derivatives of the momenta of the system itself along a virtual displacement consistent with the constraints is zero. This posit, now known as D'Alembert's Principle, is a powerful tool to calculate the path of objects moving under external constraints. But, needless to say, there are no virtual displacements (as their name indicates!); they are a tool of thought and nothing in nature corresponds to them.

We encounter a similar situation in classical thermodynamics, but with the fictionalization being the opposite: instead of infinitely fast we have infinitely slow state transitions.[6] In equilibrium thermodynamics transitions have to go through equilibrium states and to assure that the system is never pushed out of equilibrium the change of state has to be brought about by a so-called quasi-static transformation: a transformation that is infinitely slow. Again, there are no such transformations. And this not only because transformations in the world take place in finite time; in fact the very notion of an infinitely slow transformation is contradictory: if change is infinitely slow, there is no change at all. Nevertheless, quasi-static transformations lie at the heart of thermodynamics and are used in countless calculations that lead to empirically correct predictions.

No one ever believed that virtual displacements or quasi-static transformations were real. But things may not always be so clear-cut. Sometimes entities are postulated or assumptions made and it is either unclear whether the entities in question are real, or it is assumed that they are and yet later on that turns out to be wrong. Once it is acknowledged that fictions can play a role in science, this is no cause for concern. Something can be tentatively accepted on grounds of expediency, or even when discarded kept as a useful tool. A case in point is Bohr's theory of the atom, which postulates that an atom consists of a dense nucleus and a 'shell' of electrons orbiting around it on classical orbits which satisfy what is now

---

[6] See, for instance, Fermi (1936).

referred to as the Bohr-Sommerfeld quantisation rule. However, about a decade after its inception, Bohr's semi-classical theory was overthrown by Schrödinger's quantum mechanics, according to which electrons do not move on definite trajectories (irrespective of whether they satisfy the Bohr-Sommerfeld quantisation rule). Classical electron orbits have turned out to be fictions. This, however, does not render them useless. In fact, Bokulich (2009) argues that these orbits perform an important explanatory function, and hence are, their fictional character notwithstanding, by no means obsolete.

So far the status of a fiction has been conferred upon particular elements of science. Depending on where one stands in the realism versus antirealism debate, the class of fictions consists not only of convenient inventions like virtual displacements or fallen posits like electron orbits, but in fact the entire theoretical machinery of science.[7] Scientific realists hold that mature scientific theories provide, at least, an approximately true account of the parts of the world that fall within its scope. Anti-realists disagree and submit that we should only take claims about observables at face value and, depending on the kind of anti-realism one advocates, either remain agnostic about, or downright renounce commitment to, the theoretical claims of scientific theory. In our current idiom, the anti-realist regards the theoretical posit as fictions. Arthur Fine advocates this position and calls it 'fictionalism':[8]

'Fictionalism' generally refers to a pragmatic, antirealist position in the debate over scientific realism. The use of a theory or concept can be reliable without the theory being true and without the entities mentioned actually existing. When truth (or existence) is lacking we are dealing with a fiction. Thus fictionalism is a corollary of instrumentalism,

---

[7] Psillos (1999) provides a survey of different positions in this debate.

[8] The term 'fictionalism' is now also used in wider sense: you are a fictionalist about *X* if you think that *X* is somehow like fiction, where X can be moral rules, numbers, properties, etc. For a discussion of fictionalism in this broader sense see the contributions to (Kalderon 2005).

the view that what matters about a theory is its reliability in practice, adding to it the claim that science often employs useful fictions. [...] Fictionalism is allied to instrumentalism, the brand of pragmatism associated with Dewey's 'Chicago School of Thought'. (Fine 1998; cf. 1993)

Let us now turn to the other case of fiction as falsity: fiction as counterfeit, forgery, or fake, produced with the intention of deceiving and misleading. Fictions of this kind do not play an intrinsic role in science and certainly are not conducive to its goals. In fact one would wish that they played not role in science at all, but unfortunately science is no stranger to fictions of this kind. There have been cases in the past in which scientist misrepresented their achievements, stylised the findings beyond breaking point, or simply invented results that have never been obtained, with the aim of making others believe that the results were robust and thereby foster their reputations and careers. A recent high profile case is the one of the disgraced Korean scientist Hwang Woo-Suk who fraudulently reported to have created human embryonic stem cells by cloning. His alleged breakthrough in cloning stem cells had raised hopes for developing cures to diseases such as Alzheimer's, but they were deemed bogus in late 2005. He was subsequently put on trial and found guilty of accepting funds under false pretence, fabricating a series of experiments, and misleading both the scientific community and the general public. His alleged findings were fictions in the current sense, which is why they caused outrage

2. The Fiction view of Models

Models are of central importance in many scientific contexts. We often study a model to discover features of the thing it stands for. For instance, we study the nature of the hydrogen atom, the dynamics of populations, or the behaviour of polymers by studying their respective models. How is this possible and what is involved in constructing and using a model? In this section I offer a comprehensive answer to this question, in which, as it turns out, fiction plays an essential role.

Let us pump our intuitions with an example, the Newtonian model of the sun-earth system. The aim is to determine the orbit of the earth's motion around the sun.[9] The first step in the construction of the model is making various idealising assumptions about the *target-system*, the sun-earth system. We first posit that the only force relevant to the earth's motion is its gravitational interaction with the sun; we neglect all other forces, most notably the gravitational interaction with other planets in the solar system. This force is given by Newton's law of gravity, $F_g = Gm_e m_s / d^2$, where m_e and m_s are the masses of the earth and the sun respectively, d the distance between the two, and G the constant of gravitation. We then make the idealising assumption that both the sun and the earth are perfect spheres with a homogeneous mass distribution (i.e. that the mass is evenly distributed over the sphere), which allows us to calculate the strength of their gravitational interaction as if the mass of both spheres was concentrated in their centres. The sun's mass is vastly larger than the earth's and so we assume that the sun is at rest and the earth orbits around it.

With this in place we turn to classical mechanics and use Newton's equation of motion, $\vec{F} = m\vec{a}$, where $\vec{a}$ is the acceleration of a particle, m its mass and $\vec{F}$ the force acting on it, to determine the trajectory of the earth. We place the sun at the origin of the coordinate system and let $r = (x, y, z)$ be the position of the earth in that coordinate system. Plugging the above force law into Newton's equation and using $\vec{a} = \ddot{\vec{r}}$ (i.e. that the acceleration is equal to the second derivative of the position) yields $\ddot{\vec{r}} = -Gm_s \vec{r} / |\vec{r}|^3$, which

[9] See, for instance, Feynman, Leighton, and Sands (1963, Secs. 9.7 and 13.4) and Young and Freedman (2000, Ch. 12).

is the differential equation describing the earth's trajectory. Now we use various mathematical techniques to solve this equation. From an abstract point of view, solving an equation means finding those geometrical structures of which the equation is a true description. This structure consists of the system's phase space – essentially a mathematical space consisting of the position $\vec{r} = (x, y, z)$ of the earth and the corresponding velocity $\vec{v} = (v_x, v_y, v_z)$ – and the trajectory on which the earth moves.[10] It turns out that this trajectory is an ellipse.

These calculations refer to the idealised situation described above. So the last step is to carry over the results to the real target-system. To this end we argue that both the real earth and the real sun are homogenous spheres to a good degree of approximation and that all other forces acting on them are negligibly small compared to the gravitational pull between them, and that therefore the calculations made on the basis of these assumptions yield results that are true of the real sun and earth to a good degree of approximation. In order to test this claim astronomers gather data from observations. These data are then processed: obviously faulty data points are eliminated, and then statistical methods are used to fit a smooth curve through the remaining points. The result of this data processing is then compared to the model calculations and we find that the earth indeed moves around the sun on an orbit that is an ellipse to a good degree of approximation.

This example makes it clear that modelling a phenomenon involves different elements.[11] Our task is to identify these elements, analyse them, and account for how they work together.

---

[10] The details of this are rather involved. For a thorough discussion of the structure of the sun-earth system see Balzer, Moulines, and Sneed (1987, 29-34, 103-108, 180-191).

[11] Some scientific models are material objects (for instance the wood models of a car that we put into a wind tunnel), but most models are not of this kind. I here focus on models that are, in Hacking's (1983, 216) words, 'something you hold in your head rather than your hands'.

The centre piece of the Newtonian model occurs right at the beginning: we are asked to consider a situation in which the sun and the earth are perfect spheres with a homogeneous mass distribution that interact gravitationally with each other, have no interaction with anything else, etc. This is not a true description of the sun-earth system, and it is not offered as one. Rather, when modelling the solar system in this way physicists describe (and *take themselves* to be describing) an imaginary physical system. This fictional system is like the places and characters in works of fiction like Madame Bovary and Sherlock Holmes: they are the subject of thought and debate, we make claims about them that we judge right and wrong, but they live in our imagination rather then the real world. I refer to the view that scientific models essentially involve fictions of the same kind as places and characters in novels as the *fiction view of models*; it is the view that I want to develop and defend in this chapter.[12]

At this point it is helpful to return briefly to the above distinction between fiction as non-existence and fiction as imagination. Why is the sun-earth model-system like Sherlock Holmes rather than like virtual displacements or quasi-static transformations? The point to emphasise is that although it is *de facto* the case that many components of model-systems have no physical existence, this is not a *defining* feature of them; it is not the case that something must not have existing parts in it to be a model-system. In fact, models-systems are a mixture of things that do and things that do not exist: there are no spherical planets, yet there is gravitational interaction between the sun and the earth of the kind assumed in the model-system. What matters is that the model-system is considered *as a whole*, that it is studied as an ensemble, and that we consider what is the case in the given scenario. In that model-systems are like literary plots: they too are mixtures of existent and

[12] The view has recently been stated explicitly and advocated by Godfrey-Smith (2006) and myself (Frigg 2003, 2009, 2010). Ideas along the same have been developed earlier by Vaihinger's (1911), Cartwright (1983), and Sugden (2000), among others.

non-existent elements and what makes them fictional plots is not their non-existence, but rather the fact that they lead the reader to engage with them in a certain way.

Like in literature, we introduce a model-system by giving a description: sentences specifying its features.[13] Yet it is important to notice that the model-system is not the same as its description; in fact, we can re-describe the same system in many different ways, possibly also using different languages. I refer to descriptions of this kind as *model-descriptions* and the relation they bear to the model-system as *p-representation*; so we can say that the model-description p-represents the model-system. Introducing this artificial term is necessary to set off the relation between the model-description and the model-system from another form of representation to which I turn now (the choice of the term 'p-representation' will become clear in the next section).

The rationale for introducing a fiction of that kind is twofold. First, it is chosen such that it is easier to study than the target-system and therefore allows us to derive results. Second, it is assumed to represent its target system, and representation is something like a 'licence to draw inferences': representation allows us to 'carry over' results obtained in the model to the target-system and hence it enables us to learn something about that system by studying the model. I refer to the representation relation between model-system and target as *t-representation* ('t' for 'target').

Thus, scientists actually perform two acts when they propose a model: they introduce a hypothetical system as the object of study, and they claim that this system is a representation of a target-system of interest. This is reflected in the promiscuous usage of the term 'model' in the sciences. On the one hand 'model' is often used to denote the hypothetical system we study (e.g. when we say

[13] Not all model-systems are introduced by verbal descriptions; sometimes we use drawings, sketches, or diagrams to specify the model-system. The framework I introduce below can accommodate such models, but for ease of presentation I stick to cases of verbal description.

that the model consists of two spheres). On the other hand it is employed to indicate that a certain system represents, or stands for, another system (e.g. when we observe that the Newtonian model of the solar system misrepresents its target in various ways). In practice, however, these two acts are often carried out in tandem and scientists therefore rarely, if ever, clearly distinguish the two.

While this may well be a legitimate way of proceeding efficiently in the heat of battle, it is detrimental to philosophical analysis where it is germane that these two acts be kept separate. In this chapter I endeavour to clearly separate the two and to present an analysis of each. To this end, I employ the following terminology. I use the term 'model-system' to denote the hypothetical system proffered as an object of study. I call those descriptions that are used to introduce the model-system as 'model-descriptions'; t-representation then is the relation between a model-system and its target-system. The term 'model' could refer to either the model-system or representation, or the combination of the two, or yet other things; I will therefore avoid it in what follows. I use the term 'modelling' to refer to the practice of devising, describing and using a model-system. In this more regimented language, the two acts performed in utterances of the kind mentioned above are, first, presenting a model-system, and, second, endowing this model-system with representational power.

Hence, understanding scientific modelling can be divided into two sub-projects: analysing what model-systems are, and understanding how they are used to represent something beyond themselves.

Before turning to these issues, some attention needs to be paid to the other elements used above. Most notably there is the use of mathematics. The model-system itself is not mathematised, and so mathematics 'comes from the outside'. How mathematics applies to something non-mathematical is a time-honoured philosophical puzzle, and much has been written about it. However, since this is somewhat peripheral to the concerns of this chapter, I will not discuss it further; a survey of different positions regarding the applicability of mathematics can be found in Shapiro (2000). What matters for the purpose of the current discussion is that the mathematisation of the model is driven by the background theory, here classical mechanics. The theory provides a general formal framework. This framework has many blanks: the number of particles, the nature of the forces, and boundary conditions. The model-system fills these blanks: it specifies how many particles there are (namely two), what forces act upon them (namely gravity between the two), and what boundary conditions there are (namely that only periodic functions are acceptable as solutions). None of this is part of the theory, and without the model-system, the model-equation could not have been formulated, and the model-structure could not have been obtained. Given that the model-equation is derived using only properties of the model and the model-structure is the structure of which that equation is true, we can say that the model-system possesses (or instantiates) the model-structure.

Finally there are data. When observing the motion of the earth, astronomers choose a coordinate system and observe the position of the earth in this coordinate system at consecutive instants of time. They then write down these observations. This can be done in different ways. We can simply write a list with the coordinates of the moon at certain instants of time; we can draw a graph consisting of various points standing for the position of the moon at different times; or we can choose yet another form of taking down the data. The data thus gathered are called the *raw* data. The raw data then undergo a process of cleansing, rectification and regimentation: we throw away data points that are obviously faulty, take measurement errors into consideration, calculate averages, etc. Often (but not always) the aim of this process is to fit a smooth curve through the various data points so that the curve satisfies certain theoretical desiderata (such as having minimal least-square-distance from the actual data points). The end result of this process is a so-called *data-model*. This data-model is then compared to the model structure; if the two match, the model is (said to be) good. Much can be said about the construction of data models and about what it means for data to match a model structure. For a lack of space I cannot get into this issue here.

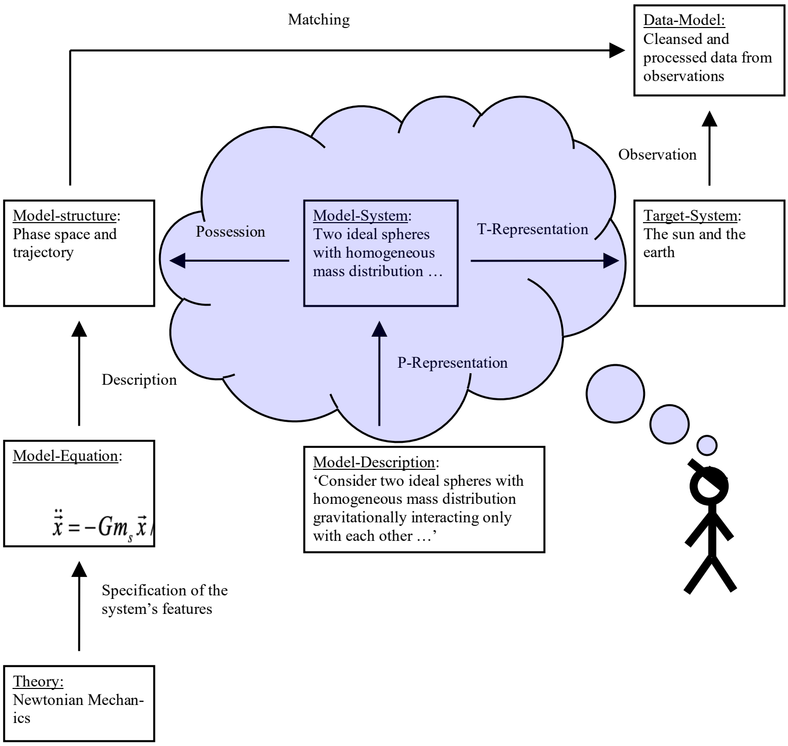

Figure 1 The elements of scientific modelling.

The discussion of the Newtonian model is summarised in Figure 1. And this schema is not only a convenient summary of that particular case; in fact, it provides a template of the basic structure of scientific modelling. In particular, the use of fictional model-systems is common not only in physics, but also in biology, economics, and other disciplines. Population biologists study the evolution of a species procreating at a constant rate in an isolated ecosystem with no deaths. And when studying the exchange of goods, economists consider a situation in which there are only two goods, two perfectly rational agents, no restrictions on available information, no transaction costs, no money, and dealings are done immediately. Examples can be multiplied ceaselessly. Their surface structure notwithstanding, no competent scientist would mistake descriptions of such systems as descriptions of an *actual* system: we know very well that there are no such systems. These descriptions are descriptions of a model-system. Hence, fictional model-systems lie at the heart of scientific modelling in many different scientific contexts.

Other elements of the above diagram are less sacrosanct – yet their absence is as interesting as their presence. Two cases stand out. The first is the absence of structures and equations. Although formalisations play an important role in modelling, not all scientific reasoning is tied to a formal apparatus. In fact, sometimes conclusions are established solely by considering a fictional scenario and without using formal tools at all. If this happens it is common to speak of a *thought experiment*. Although there does not seem to be a clear distinction between modelling and thought-experimenting in scientific practice, there has been little interaction between the respective philosophical debates.[14] This is lamentable because it seems to be important to understand how models and thought experiments relate to each other. In a recent paper Davies (2007) argues that there are important parallels between fictional narratives and thought experiments, and that exploring these parallels sheds light on many aspects of thought experiments. This take on thought experiments is congenial to the view of models presented in this paper and suggests that modelling and thought-experimenting are intrinsically related: thought experiments (at least in the sciences) are models without the formal apparatus.

The second case is the absence of t-representation. Not all models have a target system. Model-systems without targets not only play a role in explaining failures; they are also important as means to explore certain technical tools, in which case they are often referred to as ‘probing models’, ‘developmental models’, ‘study

[14] For an overview see Brown’s and Norton’s contributions on this topic to Hitchcock (2004).

models', 'toy models', or 'heuristic models'. The purpose of such model-systems is not to represent anything in nature; instead they are used to test and study theoretical tools that are later used to build representational models. In field theory, for instance, the so-called φ^4-model was studied extensively, but not because it represents anything in the world (it was well known right from the beginning that it did not), but because its simplicity allowed physicist to study complicated techniques such as renormalization in a simple setting and get acquainted with mechanisms – in this case symmetry breaking – that are important in other contexts (Hartmann 1995).

3. Modelling and Pretence

So far, I have argued that model-systems are best understood as akin to characters and objects of literary fiction.[15] However, to many this may seems to be a Pyrrhic victory because fictions are regarded as even more problematic than models. Hence the burden of proof is on the side of the proponent of the fiction view, who has to show that there is a workable conception of fiction that serves the needs of a theory of scientific modelling. Developing such a view is the aim of this section.

Before delving into the discussion, it is important to get clear on what we expect from an account of fiction in the context of scientific modelling. I think it has to provide responses to five questions:

(Q1) *Identity conditions*. When are two model-systems identical? Model-systems in science are often presented by different authors in different ways. Nevertheless, many different descriptions are actually meant to describe the same model-system. When are the model-systems specified by different descriptions identical?

(Q2) *Attribution of properties*. We frequently attribute properties to parts of model-systems, for instance when we say that rabbits in

---

[15] This section is based on my (2009).

the model reproduce at constant rate. How should we understand such statements, given that there are no such rabbits?

(Q3) *Comparative statements*. Comparing a model and its target-system is essential to many aspects of modelling, and it plays a crucial role in the account of representation developed below. We customarily say things like 'real agents do not behave like the agents in the model' and 'the surface of the real sun is unlike the surface of the model sun'. How can we compare something that does not exist with something that does?

(Q4) *Truth in model-systems*. There is right and wrong in a discourse about model-systems. It is true that the model-earth moves in an ellipse; it is wrong that it moves in a parabola. But on what basis are claims about a model-system qualified as true or false, or, more poignantly, what does it even mean for a claim about a model to be true or false? This issue becomes particularly pressing when we also take into account that we frequently judge statements as true or false about which the model-description itself remains silent. Indeed, that there is truth and falsity in a model-system beyond what is explicitly said in the original description is what makes them useful to science.

(Q5) *Epistemology*. We investigate model-systems and find out about them; truths about a model-system are not forever concealed from us. In fact, we engage with model-systems because we want to explore their properties. How do we do this? How do we find out about truths about them and how do we justify our claims?

It is the contention of this chapter that Kendall Walton's (1990) pretence theory of fiction fits the bill.[16, 17] The point of departure of

---

[16] For want of space I cannot discuss competing approaches. In a nutshell, their problems seem to be the following. The paraphrase account (Russell 1905) does not offer a workable theory of truth in fiction (Crittenden 1991, Ch. 1). The neo-Meinongean view (Parsons 1980) runs into difficulties with incompleteness (Howell 1979, Sec. 1) and as a consequence does not offer a satisfactory answer to (Q5). Finally, Lewis' (1978) ac-

this view is the capacity of humans to imagine things.[18] Sometimes we imagine something without a particular reason. But there are cases in which our imagining something is prompted by the presence of a particular object, in which case this object is referred to as a 'prop'. 'Object' has to be understood in the widest sense possible; anything capable of affecting our senses can serve as a prop. An object becomes a prop due to the imposition of a rule or 'principle of generation' (p. 38), prescribing what is to be imagined as a function of the presence of the object. If someone imagines something because he is encouraged to do so by the presence of a prop he is engaged in a game of make-believe. Someone who is involved in a game of make-believe is pretending; so 'pretence' is just a shorthand way of describing participation in such a game (p. 391) and has (in this context) nothing to do with deception (p. 392). The simplest examples of games of make-believe are cases of child's play (p. 11). In one such case, stumps may be regarded as bears and a rope put around the stump may mean that the bear has been lassoed; or pointing the index finger at someone and saying 'bang' may mean that the person has been shot.

A prop becomes a prompter if someone notices the prop and as a result starts engaging in a rule-guided imaginative activity. The set of prompters and the set of props overlap, but neither is a subset of the other. For one, a prop that is never perceived by anybody and hence never causes anybody to imagine something is not a prompter (but still a prop). For another, an object can prompt imaginations without being part of a game of make-believe (i.e. in the absence of

---

count is too permissive about what counts as true in a fictional context (Currie 1990, Sec. 2.3; Lamarque and Olsen 1994, Ch. 4).

[17] Strictly speaking, Walton (1990) restricts the use of 'pretence' to verbal (or more generally behavioural) participation, which does not include the activity of someone reading on his own. However, it has become customary to use 'pretence' as synonymous with 'make-believe' and I stick to this wider use in what follows.

[18] I here discuss pretence theory as it is presented by Walton (1990); Currie (1990) and Evans (1982, Ch. 10) develop different versions. Parenthetical references in the text of this and the following section are to Walton's book.

rules of generation), for instance when we see faces in the clouds and imagine how these faces talk to each other. Even within a game we can make errors (e.g. mistakenly take a mole heap for a stump and then say that it is a bear), in which case the mole heap is a prompter (because it prompts imaginings) but it is not a prop (because there is not a rule).

Pretence theory considers a vast variety of different props ranging from novels to movies, from paintings to plays, and from music to children's games. In the present context I only discuss the case of literature. Works of literary fiction are, on the current account, regarded as props because they prompt the reader to imagine certain things. By doing so a fiction generates its own game of make-believe. This game can be played by a single player when reading the work, or by a group when someone tells the story to the others.

Some rules of generation are *ad hoc*, for instance when a group of children spontaneously imposes the rule that stumps are bears and play the game 'catch the bear'. Other rules are publicly agreed on and hence (at least relatively) stable. Games based on public rules are 'authorized'; games involving *ad hoc* rules are 'unauthorized'.

By definition, a prop is a representation if it is a prop in an *authorised* game. On this view, then, stumps are not representations of bears because the rule to regard stumps as bears is an *ad hoc* rule that is neither shared by others in the society nor stable over time (stumps may not be props to other people and even the children playing the game now may regarded them as elephants on the next walk). However, *Hamlet* is a representation because everybody who understands English is invited to imagine its content, and this has been so since the work came into existence. Within pretence theory 'representation' is used as a technical term. Representations are not, as is customary, explained in terms of their relation to something beyond themselves; representations are things that possess the social function of serving as props in authorised games of make-believe (I will come back to this point below). This notion of representation is what is at work in what I have called p-representation ('p' for 'prop') above.

Props generate fictional truths by virtue of their features and principles of generation. Fictional truths can be generated directly or indirectly; directly generated truths are ‘primary’ and indirectly generated truths are ‘implied’ (p. 140). Derivatively, one can call the principles of generation responsible for the generation of primary truths ‘principles of direct generation’ and those responsible for implied truths ‘principles of indirect generation’. The leading idea is that primary truths follow immediately from the prop, while implied ones result from the application of some rules of inference. When little Jimmy sees a stump and shouts ‘here is a bear’ this is a direct truth because it follows from fact that there is a stump and the direct rule ‘stumps are bears’, which is constitutive of the game. The boys may then stay away from the bear because they think the bear is dangerous and might hurt them. This fictional truth is inferred because it does not follow from the basic laws of the game that stumps are bears, but from the additional principle that bears in the game have the same properties as real bears.

The distinction between primary and inferred truths is also operative in literary fiction. The reader of *Changing Places* reads that Zapp ‘embarked […] on an ambitious critical project: a series of commentaries on Jane Austen which would work through the whole canon, one novel at a time, saying absolutely everything that could possibly be said about them.’ The reader is thereby invited to imagine the direct truth that Morris Zapp is working on such a project. She is also invited to imagine that Zapp is overconfident, arrogant in an amusing way, and pursues a project that is impossible to complete. None of this is explicitly stated in the novel. These are inferred truths, which the reader deduces from common knowledge about academic projects and the psyche of people pursuing them. What rules can legitimately be used to reach conclusions of this sort is a difficult issue fraught with controversy. I will return briefly to it below; for the time being all that matters is that there are such rules, no matter what they are.

This framework has the resources to explain the nature of model-systems. Typically, model-systems are presented to us by way of descriptions, and these descriptions should be understood as props in games of make-believe. These descriptions usually begin with expressions like ‘consider’ or ‘assume’ and thereby make it clear that they are not descriptions of fact, but an invitation to ponder - in the present idiom, imagine - a particular situation. Although it is often understood that this situation is such that it does not occur anywhere in reality, this is not a prerequisite; models, like literary fictions, are not *defined* in contrast to truth. In elementary particle physics, for instance, a scenario is often proposed simply as a suggestion worth considering. Only later, when all the details are worked out, the question is asked whether this scenario bears an interesting relation to what happens in nature, and if so what the relation is.

The ‘working out’ of the details usually consists in deriving conclusions from the primary assumptions of the model and some general principles or laws that are taken for granted. For instance, we derive that the earth moves in an elliptical orbit from the basic assumptions of the Newtonian model and the laws of classical mechanics. This is explained naturally in the idiom of pretence theory. What is explicitly stated in a model description (that the model-earth is spherical, etc.) are the primary truths of the model, and what follows from them via laws or general principles are the implied truths; the principles of direct generation are the linguistic conventions that allow us to understand the relevant description, and the principles of indirect generation are the laws that are used to derive further results from the primary truths.

We can now address the above questions. The attribution of certain concrete properties to models (Q2) is explained as it being fictional that the model-system possesses these properties. To say that the model-population is isolated from its environment is just like saying that Zapp drives a convertible. Both claims follow from a prop together with rules of generation. In other words, saying that a hypothetical entity possesses certain properties involves nothing over and above saying that within a certain game of make-believe we are entitled to imagine the entity as having these properties. For this reason there is nothing mysterious about ascribing concrete properties to nonexistent things, nor is it a category mistake to do so.

Let us now discuss the issue of truth in model-systems (Q4), which will also provide us with solutions to the other open questions. The question is: what exactly do we assert when we qualify 'Zapp drives a convertible' as true in the fiction while 'Zapp drives a Mini Cooper' as false?[19] To begin with, it is crucial to realise that there are three different kinds of statement in connection with fiction, and that these require a different treatment when it comes to the questions of truth; I refer to these as intrafictional, metafictional, and transfictional statements.[20] For someone sitting in an armchair reading *Changing Places* 'Morris jumped into the paternoster on the downside' is an intrafictional statement because the reader is involved in playing the game defined by the novel and imagines that the sentence's content is the case. Someone who read the novel a while ago and asserts in discussion with a friend that Zapp jumped into a paternoster makes a metafictional statement because he is talking about the fiction. If he then also asserts that Zapp, his quirks notwithstanding, is more likeable than any literature teacher he ever had or that Zapp is smarter than Candide, he makes transfictional statements as he is comparing Zapp to a real person and a character in another fiction.

Intraficational propositions are made within the fiction and we are not meant to believe them, nor are we meant to take them as reports of fact; we are meant to imagine them. Although some statements are true in the fiction as well as true *tout court* ('1968 was the year of student revolts' is true and true in *Changing Places*),

---

[19] There is controversy over this issue even within pretence theory. It is beyond the scope of this paper to discuss the different proposals and compare them to one another. In what follows I develop an account of truth in fiction that is based on elements from different theories and that is tailored towards the needs of a theory of model-systems.

[20] All theories of fiction acknowledge this distinction. My terminology is adapted from Currie (1990, Ch. 4) who speaks about the 'fictive', 'metafictive' and 'transfictive' use of fictional names.

we often qualify false statements as true in the fiction ('Zapp is a literary theorist' is false because there is no Zapp) and true statements as false in the fiction ('white light is composed of light of other colours' is false in Goethe's *Faust*). So truth and truth in fiction are distinct; in fact, truth in fiction is not a species of truth at all (p. 41). For this reason it has become customary when talking about what is the case in a fiction to replace locutions like 'true in the fiction' or 'true in a fictional world' by the term of art 'being fictional'; henceforth '$F_w(p)$' is used as an abbreviation for 'it is fictional in work *w* that *p*', where *p* is a placeholder for an intrafictional proposition like 'Zapp pursues an impossible project'.[21]

The question now becomes: when is *p* fictional in *w*? Let the *w*-game of make-believe be the game of make-believe based on work *w*, and similarly for '*w*-prop' and '*w*-principles of generation'. Then, *p* is fictional in *w* iff *p* is to be imagined in the *w*-game of make-believe (p. 39). In more detail:

p is fictional in work w iff the w-prop together with the w-principles of generation prescribes p to be imagined

This analysis alleviates worries about the (alleged) subjectivity of imaginings. In common parlance, 'imagination' has subjective overtones, which might suggest that an understanding of models as imagined entities makes them subjective because every person imagines something different. This is not so. In pretence theory, imaginations in an authorised game of make-believe are sanctioned by the prop itself and the rules of generation, both of which are public and shared by the relevant community. Therefore, someone's imaginings are governed by intersubjective rules, which guarantee that, as long as the rules are respected, everybody in-

[21] I here follow Currie (1990, Ch. 2) and assume that sentences like 'Zapp drives a convertible' express propositions, something that Walton denies (p. 391). This assumption greatly simplifies the statement of truth conditions for fictional statements, but nothing in the present paper hangs on it. Essentially the same results can be reached only using sentences and pretence (see pp. 400-405).

volved in the game has the same imaginings. So, not only do all participants in the game *de facto* imagine the same things (which could also be the result of happenstance), but they do so because they participate in a rule-governed activity. What is more, participants *know* that they do; they know that they are participants in an authorised game and as long as they trust that the others play by the rules they can trust that other have the same imaginings.

Furthermore, for a proposition to be fictional in work *w* it is not necessary that it is actually imagined by anyone: fictional propositions are ones for which there is a prescription to the effect that they *have to be imagined* (p. 39), and whether a proposition is to be imagined is determined by the prop and the rules of generation. Hence, props, via the rules of generation, make propositions fictional independently of people's actual imaginings (p. 38), and for this reason there can be fictional truths that no one knows of. If there is a stump hidden behind a bush, unknown to those playing the game, it is still fictional that there is a bear behind the bush; the prop itself and the rules of generation are sufficient to generate this fictional truth.

With this in place we can now also render the concept of a 'fictional world' or 'world of a fiction' precise: the world of work *w* is the set of all propositions that are fictional in *w*.[22]

This analysis of truth in fiction carries over to model-systems one-to-one simply by replacing *p* by a claim about the model, *w* by the description of the model-system, and *w*-principles of generation by the laws and principles assumed to be at work in the model. For instance, 'the solar system is stable' is true in the Newtonian model of the solar system systems iff the description of the system together with the laws and principles assumed to hold in the system (the laws of classical mechanics, the law of gravity, and some gen-

---

[22] Fictional worlds thus defined are rather different from possible worlds as used in modal logic, the most significant difference being that the former are incomplete while the latter are not. See Currie (1990, 53-70) for a discussion of possible worlds and fiction.

eral assumptions about physical objects) imply that this is the case. This gives us a straightforward answer to the question about identity conditions (Q1): two models are identical iff the worlds of the two models – the set of all propositions that are fictional in the two models – are identical.[23]

Metafictional propositions make genuine claims that can be true or false in the same way in which claims about chairs and tables can be true or false. But how can such statements be true if the singular terms that occur in them have no referents? A solution emerges when we realise that statements like 'Zapp is a professor' are ellipses for 'in *Changing Places*, Zapp is a professor'. So when we metafictionally assert *p*, what we really assert is 'in work *w*, *p*' (p. 397). Asserting that something is the case in a work of fiction is tantamount to asserting that it is fictional in that work. Hence asserting 'in work *w*, *p*' amounts to asserting '*p* is fictional in work *w*', which in turn in is equivalent to 'it is fictional in work *w* that *p*'. The last sentence is, of course, just $F_w(p)$. Hence metafictionally asserting *p* amounts to asserting $F_w(p)$. The truth condition for this assertion follows from what has been said above:

$F_w(p)$ is true iff p is fictional in w, which in turn is the case iff the w-prop and together with the w-principles of generation prescribe p to be imagined.

Derivatively, *p*, when uttered as a metafictional claim, is true iff *p* is fictional when uttered as an intrafictional claim.[24] In sum, once we understand that a metafictional claim has to be prefixed by 'In

[23] An interesting consequence of this identity condition is that not all models with the same prop are identical, because they can operate with different rules of indirect generation. This is the case, for instance, when the 'same model' is treated first classically and then quantum mechanically; on the current view, the classical and the quantum model are not identical.

[24] In some places Walton ties the truth of such statements to *authorised* games (e.g., p. 397-8). This restriction seems unnecessary as the analysis works just as well for unauthorized games.

fiction *w*', and hence has the structure $F_w(p)$, the truth of the claim is determined by appeal to the *w*-game of make-believe. Again, this analysis translates to scientific statements without further ado.

Transfictional propositions pose a particular problem because they – apparently – involve comparisons with a nonexistent objects, which does not seem to make sense: we cannot compare someone with Zapp if there is no Zapp. Different authors have offered very different solutions to this problem.[25] Fortunately we need not deal with the problem of transfictional statements in its full generality because the transfictional statements that are relevant in connection with model-systems are of a particular kind: they compare features of the model-systems with features of the target-system. For this reason, transfictional statements about model-systems should be read as prefixed with a clause stating what the relevant respects of the comparison are. This allows us to rephrase comparative sentences as comparisons between properties rather than an object, which makes the original puzzle go away.

Crucially, then, truth conditions for transfictional statements in the context of scientific modelling come down to truth conditions for comparative statements between properties, which are unproblematic in the current context (for the problems that attach to them have nothing to do with issues surrounding fictional discourse). For instance, when I say 'my friend James is just like Zapp' I am not comparing my friend to a nonexistent person. What I am asserting is that both James and Zapp possess certain relevant properties (Zapp possesses properties in the sense explained above) and that these properties are similar in relevant ways. Likewise, when I say that the population of rabbits in a certain ecosystem behaves very much like the population in the Fibonacci model, what I assert is that these populations possess certain relevant properties which are similar in relevant respects. What these relevant properties are and

[25] Lamarque and Olsen (1994, Ch. 4), for instance, solve the problem by introducing characters. Walton, by contrast, renounces the commitment to characters and instead analyses transfictional statements in terms of unauthorized games (pp. 405-416).

what counts as being similar in relevant respects may well depend on the context. But this is not a problem. All that matters from a semantic point of view is that the apparent comparison with a nonexistent object eventually comes down to the unproblematic comparison of properties. Further, the statement making this comparison is true iff the statement comparing the properties with each other is true. Obviously, statements comparing two nonexistent objects are analysed in exactly the same way.[26]

These insights provide us with answers to (Q3) and (Q4). And what is more, this take on truth also provides us with an answer to the question about the epistemology of models (Q5): we investigate a model by finding out what follows from the primary truths of the model and the rules of indirect generation. This seems to be both plausible and in line with scientific practice because a good deal of the work that scientists do with models can accurately be described as studying consequences of the basic assumptions of the model.

Finally, let me add a word about rules of generation. Although the general idea is intuitively clear, it turns out to be difficult to give an account of these rules. So what are the rules of generation in scientific fictions? This is a substantial question that needs to be addressed, but we should not expect a single unified answer. On the contrary, it seems plausible to assume that different disciplines have different rules, and understanding what these rules are will shed light on how modelling in these disciplines works. So we should not expect a ready-made answer, but rather regard the study of rules of generation as part of research programme aiming at understanding the practice of modelling in various branches of science.

By way of closing it is worth mentioning that this account is ontologically parsimonious: we have not incurred ontological commitments to fictional entities. Walton's theory is antirealist in that it

---

[26] For a critical discussion of this account see Godfrey-Smith (2009, 113-4).

renounces the postulation of fictional (or abstract) entities, and hence a theory of scientific modelling based on this account is also free of ontological commitments. This, of course, is not a refutation of metaphysically less parsimonious views such as Meinong's, and there may be reasons to eventually prefer such a view over an antirealist one. The point to emphasise here is that whatever these reasons may be, the needs of science are not one among them.

No theory of modelling is complete without an account of t-representation. It is an essential feature of many models that they represent a target system. After having presented an account of what it means for claims about a model-system to be true, how we learn about model-systems, and how we can meaningfully compare them to either things in the world or other model systems, it is now time to discuss how model-systems represent (i.e. t-represent) something beyond themselves. Representation has been discussed controversially and a review of this literature is beyond the scope of this chapter; in what follows I will just present a brief statement of the account of representation that I favour (Frigg 2010, Sec. 6).[27] Understanding t-representation involves establishing and understanding a relation between the fictional scenario and parts (or aspects) of the real world. While we sometimes do this casually (for instance when I compare my friend James with Zapp), there is no canonical way in which this is done, and much seems to be left to reader's personal interpretation. Not so in science. Models not only represent their target; they do so in a clearly specifiable and unambiguous way, and one that allows scientists to 'read off' features of the target from the model. In fact, model-systems are the units on which significant parts of scientific investigation are carried out rather than on the target system itself: we study a model and thereby discover features of the thing it stands for. We do this by first finding out what is true in the model-system itself, and then translating the findings into claims about the target itself. This is possible only if the model-system t-representations of the target.

[27] For recent discussions of scientific representation see Contessa (2007), Frigg (2006), Hughes (1997), Suárez (2004) and Toon (2009).

This realisation provides us with the elements of the general scheme of representation:

X t-*represents Y* iff:
(R1) *X denotes Y.*
(R2) *X comes with a key K specifying how facts about X are to be translated into claims about Y.*

In nutshell, the idea is that the first condition establishes the aboutness of *X*, and the second guarantees the cognitive relevance of *X* for *Y*.[28]

We find this kind of representation not only in models but also in maps, which provide a good example to illustrate the workings of t-representation thus defined. I have in front of me a map of North London. This is the first condition: the map denotes North London. Now I look at the details. I see a black rectangle on a black line and written next to it is 'Camden Road'. The explanations that come with the map say that this rectangle stands for an over-ground railway station, the name next to it is the name of the station, and the black line stands for the rail tracks. A bit further up there is a black dot on a black line. The legend says that the dot stands for a tube station, and the name written next to it is the name of the station, in this case 'Kentish Town'. Between the two there is a thick yellow line, which stand for a main road. Hence, that a black rectangle labelled 'Camden Road' is connected by a thick yellow line to a black dot labelled 'Kentish Town' (a fact in the map) translates into the fact that Camden Road railway station is connected to Kentish Town tube station by a main road (a fact about North London). Furthermore, from the fact that this yellow line is 4.5cm long, I can infer that the actual distance between the two is about 1km since the scale of the map is 4.55cm to 1km. Finally, the 'Kentish Town' dot lies vertically above the 'Camden Road' rectangle, from which I infer that Kentish Town tube station is north of Camden Road railway station.

[28] The first condition is Goodman's (1976, Ch. 1) who argued that denotation lies at the heart of representation.

Our use of a map essentially involves a key, telling us how to translate facts about the map into putative facts about North London. Some elements of the key are stated at the bottom of the map; for instance, we are instructed that rectangles stand for railway stations and dots for tube stations. Other elements are conventions that are so common that they are assumed without further explanation. The top of the map indicates north, for example, and the distances in the map are proportional to distances in the world (where the ‘scale’ of the map gives the proportionality factor).

It is worth emphasising that (R2) defines the role of the key as providing a translation of *facts* about *X* into *claims* about *Y*. This is not a slip. An acceptable definition of t-representation has to make room for misrepresentation. A map can contain errors in the sense that even if we use the right key and use it correctly we may obtain wrong results. For instance, it may have happened that the cartographers failed to connect the black dot and the black rectangle with a yellow line, and so we would have been led to believe that the two stations are not connected by a main road. This would not have turned the map into a non-t-representation; it would still have been a t-representation, but one that misrepresents North London. Saying that we translate facts in the map into *claims* about the target makes room for error because claims can be true or false, while facts cannot. A representation is a *faithful* representation iff if all claims about *Y* are true.

However, (R1) and (R2) only provide the *general form* of an account of t-representation, which needs to be concretised in every particular instance of a t-representation. In fact, ‘denotation’ and ‘key’ are just blanks to be filled. In order to understand how a *particular* representation works, we need to account for how the *X* at hand comes to denote *Y*, and we have to provide a particular key *K*. In the above example, we borrowed denotation from ordinary language by saying ‘this is a map of North London’, and the key was provided to us by cartography. But other cases may work differently since there may be different sources of denotation and there may be any number of keys that can be used to interpret *X*. Moreover, keys are often implicit and determined by context. This is often the case with scientific representations, which unlike maps, rarely, if ever, come with something like a legend. It is one of the challenges facing a philosophical analysis of representation to make hidden assumptions explicit, and present a clear statement of them. The claim that something is a t-representation amounts to an invitation to spell out how exactly *X* comes to denote *Y* and what *K* is.

This generality is an advantage. The class of t-representations is large and its members varied. A view that claims that all t-representations work in exactly the same way would be doomed to failure right from the beginning. Maps, graphs, architectural plans, diagrams, photographs, (certain kinds of) paintings and drawings, and of course scientific models, are all t-representations in that they satisfy (R1) and (R2), but they work in very different ways. The differences between them are that these conditions are realised in very different ways: different keys are used and denotation has different sources. The challenge for a complete account of representation is to come up with a taxonomy of different ways in which the two conditions can be realised, and to explain how they differ from each other.

Hence, the detailed study of different keys is a research programme to be undertaken in the future. However, to get a better idea of what such an investigation involves I now want to discuss two keys often used in science: identity and ideal limits.

The simplest of all keys is *identity*, the rule according to which facts in the model (or at least a suitably defined class of facts) are also facts in the world. Although scientists often talk as if the relation between models and reality was identity, there are actually very few, if any, models that work in this way.

A more interesting key is the *ideal limit* key. Many model-systems are idealisations of the target in one way or another. A common kind of idealisations is to 'push to the extreme' a property that a system possesses. This happens when we model particles as point masses, strings as massless, planets as spherical, and surfaces as frictionless. Two things are needed to render such idealisations benign: experimental refinements and convergence (Laymon 1991). First, there must be the *in principle* possibility of refining actual systems in a way that they are made to approach the postulated limit (that is, we don't actually have to produce these systems; what matters is that we could produce them in principle. With respect to friction, for instance, one has to find a series of experimental refinements that render a tabletop ever smoother and hence allow real systems to come ever closer to the ideal frictionless surface. These experimental refinements together constitute a sequence of systems that come ever closer to the ideal limit. Second, this sequence has to behave 'correctly': the closer the properties of a system come to the ideal limit, the closer its behaviour has to come to the behaviour in the limit. If we take the motion of a spinning top on a frictionless surface to be the ideal limit of the motion of the same spinning top on a non-frictionless surface, then we have to require that the less friction there is, the closer the motion of the real top comes to the one of the idealised model. Or to put it in more instrumental terms, the closer the real situation comes to the ideal limit, the more accurate the predictions of the model. This is the requirement of convergence. If there exists such a sequence of refinements and if the limit is monotonic, then the model is an ideal limit.

If a model is an ideal limit, this implies a key. To see how, let us first briefly recapitulate the mathematical definition of a limit. Consider a function $f(x)$, and then ask the question how $f(x)$ behaves if x approaches a particular value x_0. We say that the number F is the limit of $f(x)$ (in symbols: $\lim_{x \to x_0} f(x) = F$) iff for every positive number ε (no matter how small), there exists another positive number δ such that: if $|x - x_0| < \delta$, then $|f(x) - F| < \varepsilon$. Colloquially, this says that the closer x comes to x_0, the closer $f(x)$ comes to F: if we know that x is less than δ way from x_0, then we also know that $f(x)$ is less than ε away from F. This idea can now be used for ideal limits in the above sense. The sequence of experimental refinements plays the role of x, and the limit itself is x_0; in the example: the ever smoother table tops correspond to different values of x, and the frictionless plane corresponds to x_0. The behaviour of the object corresponds to f. If there is a limit we know that if the difference between the friction of the real plane and the ideal frictionless plane is smaller than δ, then difference between the behaviour of the real spinning top and the ideal spinning top in the model-system is smaller than ε. So if we are given the friction of the table, we know how to translate facts obtaining in the model-system into claims about the world.

4. Replies to Criticisms

I have introduced the fiction view of models and presented a particular version of it based on Walton's pretence theory. However, the fiction view of models is not uncontroversial. In this last section I want to briefly address some criticisms that have been levelled against it (I restrict attention to criticisms against the fiction view in general, and set aside quibbles about my particular version of it).

There are four different lines of attack. The first criticism is driven by worries about fiction. Fictions, so the argument goes, are intrinsically dubious and are beset with so many serious problems that one should steer away from them whenever possible. So a view assigning them a central role in science is downright suicidal.[29]

This, I think, overstates the problems with fictions. Sure enough, there is controversy surrounding fictions; but the problems surrounding fictions aren't more devastating then those surrounding other items on the philosophical curriculum. Furthermore, with the pretence view of models developed in the last section there is a workable suggestion on the table, which, if nothing else, shows that the fiction view can be given a coherent formulation.

The second criticism is that the fiction view – involuntarily – plays into the hands of science sceptics and irrationalists (Giere 2009,

[29] I have been unable to locate this view in print, but it has been put to me many times in conversation.

257). Creationists and other (in particular religiously motivated) science-sceptics will find great comfort, if not powerful rhetorical ammunition, in the fact that respected philosophers of science say that what most scientists do during most of their professional lives - namely working in one way or another with models - is producing fiction. This, so the argument goes, will be seen as a justification of the view that the claims of religion are on par with those of science. Hence the fiction view of models undermines the authority of science and fosters the cause of those who wish to replace science with religious or other unscientific worldviews.

Needless to say, I share Giere's concerns about creationists and would be chagrined if the fiction view of models was used to support their cause. But in order not to misidentify the problem it is important to point out that Giere's claim is not that the view itself – or its proponents – support creationism; his worry is that the view is a dangerous tool when it falls into the wrong hands. What follows from this, however, is not that the fiction view itself should be abandoned; what follows is that some care is needed when dealing with the press office. Improving the impact of your research by having the popular press report that you have discovered that science is fiction may not be a good idea. But as long as the fiction view of models is discussed in informed circles, and, when popularised, is presented carefully and with the necessary qualifications, it is no more dangerous than other ideas, which, when taken out of context, can be put to use that would send shivers down the spines of their progenitors.

The third objection, also due to Giere, has it that the fiction view misidentifies the aims of models. Giere agrees that from an *ontological* point of view scientific models and works of fictions are on par, but emphasises that '[i]t is their differing function in practice that makes it inappropriate to regard scientific models as works of fiction' (*ibid*., 249). Giere identifies three functional differences (*ibid*., 251-2). First, while fictions are the product of a single author's individual endeavours, scientific models are the result of a public effort because scientists discuss their creations with their colleagues and subject them to public scrutiny. Second, there is a clear distinction between fiction and non-fiction books, and even when a book classified as non-fiction is found to contain false claims, it is not reclassified as fiction. Third, unlike works of fiction whose prime purpose is to entertain (although some works can also give insight into certain aspects of human life), scientific models are representations of certain aspects of the world.

These observations, although correct in themselves, have no force against the fiction view of models, at least in the version developed in this chapter. First, whether a fiction is the product of an individual or a collective effort has no impact on its status as a fiction; a collectively produced fiction is just a different kind of fiction. Even if *War and Piece* (to take Giere's example) had been written in a collective effort by all established Russian writers of Tolstoi's time, it would still be a fiction. Vice versa, even if it were true that Newton had never discussed his model of the solar system with anybody before publishing it, it would still be science. The history of production is immaterial to the status of a work as fiction. Second, at least in my version of the fiction view of models, falsity is not a defining feature of a fiction. I have distinguished right from the beginning a use of 'fiction' as falsity from a use of 'fiction' as imagination. I agree with Giere that there is a clear distinction between texts of fiction and non-fiction, but I deny that this distinction is defined by truth or falsity; it is the attitude that we are supposed to adopt towards the text's content that makes the difference. Once this is realised, the problem fades away. Third, proponents of the fiction view are clear that problems of ontology should be kept separate from function and agree that it is one of the prime function of models to represent. This point has been stressed by Godfrey-Smith (2009, 108-111) and it is explicit in my own view, which draws a clear distinction between p-representation and t-representation, where the latter is explained in terms that have very little, if anything, to do with how literary fictions work.

The fourth objection is that fictions are superfluous and hence should not be regarded as forming part of (let alone *being*) scientific models because we can give a systematic account of how scientific models work without invoking fictions. This point has been made in different ways by Pincock (2009) and Weisberg (2009). I cannot do justice to the details of their sophisticated arguments here, and will concern myself only with their main conclusion. They argue that scientific models are mathematical objects and that they relate to the world due to the fact that there is a relationship between the mathematical properties of the model and the properties found in the target system (in Weisberg's version similarity relations to a parametrised version of the target). In other words, models are mathematical structures and they represent due to there being certain mathematical relations between these structures and a mathematical rendering of the target system. (Weisberg includes fictions as convenient 'folk ontology' that may serve as a crutch when thinking about the model, but takes them to be ultimately dispensable when it comes to explaining how models relate to the world).[30]

In the remainder of this section I want to briefly indicate why views (whatever their specifics) of models that make no room for fictional systems miss out on important aspects of scientific modelling. The first point to stress is that most of the time it is a fictional scenario that provides the 'entry ticket' to a mathematical treatment of a scientific problem. Only after thinking of planets as perfect homogeneous spheres can we apply Newtonian mechanics to them; and the equations we write down are true only of such spheres and not of real world planets (which are not spherical). But without such fictions there is no mathematical treatment of the problem; we simply would not know what equations to write down. This point is not new. Cartwright (1983) has pointed out that a 'prepared description' of the target is the first condition for theory entry. In the current idiom this amounts to saying that we present a fictional description.

This description is essential not only for entering the theory, but also for improving the model. If a model fails to make correct pre-

[30] This is reminiscent of the so-called semantic view of theories; yet neither of them endorses this view. For a discussion of the semanitic view *vis a vis* the fiction view of models see my (2006, 2010).

dictions, it is knowledge of the fictional scenario that tells us how to improve the model. If, say, a model of the growth of a population gets the numbers of fish in the pond dramatically wrong, then it helps to realise that the equations of the model are literally true only of a population in which fish never die and which have infinite supplies of food. We can then build in the life span of animals and food shortages, which may improve the performance of the model. We would not know how to do this if we did not know in what fictional scenario the equations hold true.

Finally, there is a pervasive intuition that models have content: the Newtonian Model is about spherical planets. But mathematical structures (or equations) have no content, or at least no content of that kind. It is the fictional scenario that provides the content of the model; this content neither comes from the mathematical skeleton of the model itself, nor from a comparison of this skeleton with data. Data play an evidential role in confirming representational claims, but data are not the content of a representation (I discuss this point in more detail in my 2010).

This is just a sketch and these points need to be developed in grater detail, but I hope that they indicate along what lines the fiction view of models can be defended against the charge that fictions are an idle wheel.

Acknowledgments

I would like to thank José Díez and Foad Dizadji-Bahmani for comments on earlier drafts, and Chris Pincock and Michael Weisberg for helpful discussions.

BIBBLIOGRAPHY

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

Bokulich, Alisa (2009), "Explanatory Fictions", in Mauricio Suárez (ed.), *Fictions in Science. Philosophical Essays on Modelling and Idealization*, London and New York: Routledge, 91-109.

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

Contessa, Gabriele (2007), "Scientific Representation, Interpretation, and Surrogative Reasoning", *Philosophy of Science* 74 (1):48–68.

Crittenden, Charles (1991), *Unreality: The Metaphysics of Fictional Objects*. Ithaca and London: Cornell University Press.

Currie, Gregory (1990), *The Nature of Fiction*. Cambridge: Cambridge University Press.

Davies, David (2007), "Thought Experiments and Fictional Narratives", *Croatian Journal of Philosophy* 7 (19):29-45.

Evans, Gareth (1982), *The Varieties of Reference*. Edited by John McDowell. Oxford: Oxford University Press.

Fermi, Enrico (1936), *Thermodynamics*. Mineola/NY: Dover Publications, 2000.

Feynman, Richard, Robert B. Leighton, and Mathew L. Sands, eds. (1963), *The Feyman Lectures in Physics*. Redwood City and Reading/Ma 1989.

Fine, Arthur (1993), "Fictionalism", *Midwest Studies in Philosophy* 18:1-18.

Fine,Arthur (1998), "Fictionalism", in Edward Craig (ed.), *Routledge Enczclopaedia of Philosophy*, London: Routledge.

Friend, Stacie (2007), "Fictional Characters", *Philosophy Compass* 2 (2):141–156.

Frigg, Roman (2003), *Re-presenting Scientific Representation*, PhD Thesis. London: University of London.

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

Frigg, R oman (2009), "Models and Fiction", *Synthese* (forthcoming).
Frigg, Roman (2010), "Fiction and Scientific Representation", in Roman Frigg and Matthew
Hunter (eds.), *Beyond Mimesis and Convention: Representation in Art and Science*, Berlin and New York: Springer.
Giere, Ronald (2009), "Why Scientific Models Should Not Be Regarded as Works of
Fiction", in Mauricio Suárez (ed.), *Fictions in Science. Philosophical Essays on Modelling and Idealization*, London: Routledge, 248-258.
Godfrey'Smith, Peter (2009), "Models and Fictions in Science", *Philosophical Studies* 143:101-116.
Godfrey-Smith, Peter (2006), "The Strategy of Model-Based Science", *Biology and Philosophy* 21:725-740.
Goodman, Nelson (1976), *Languages of Art*. 2nd ed., Indianapolis and Cambridge: Hacket.
Hacking, Ian (1983), *Representing and Intervening*. Cambridge: Cambridge University Press.
Hartmann, Stephan (1995), "Models as a Tool for Theory Construction: Some Strategies of Preliminary Physics", in William E. Herfel, Wladiyslaw Krajewski, Ilkka Niiniluoto and Ryszard Wojcicki (eds.), *Theories and Models in Scientific Processes (Poznan Studies in the Philosophy of Science and the Humanities 44)*, 49-67.
Hitchcock, Christopher, ed. (2004), *Contemporary Debates in Philosophy of Science*. Oxford: Blackwell.
Howell, Robert (1979), "Fictional Objects: How They Are and How They Aren't", *Poetics* 8:129-177.
Hughes, R. I. G. (1997), "Models and Representation", *Philosophy of Science* Vol. 64, Supplement:S325-S336.
Kalderon, Mark Eli, ed. (2005), *Fictionalism in Metaphysics*. Oxford: Clarendon Press.
Kuypers, Friedhelm (1992), *Klassische Mechanik.* 3rd ed. Weinheim: VHC.
Lamarque, Peter, and Stein Haugom Olsen (1994), *Truth, Fiction, and Literature*. Oxford: Clarendon Press.

Laymon, Ronald (1991), "Thought Experiments by Stevin, Mach and Gouy: Thought Experiments as Ideal Limits and as Semantic Domains", in Tamara Horowitz and Gerald J. Massey (eds.), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield, 167-191.

Lewis, David (1978), "Truth in Fiction", in David Lewis (ed.), *Philosophical Papers, Volume I*, Oxford: Oxford University Press 1983, 261-280.

Parsons, Terrence (1980), *Nonexistent Objects*. New Haven: Yale University Press.

Pincock, Christopher (2009), "Discerning the Truth in Fiction", *Unpublished Manuscript*.

Psillos, Stathis (1999), *Scientific Realism: How Science Tracks Truth*. London.

Russell, Bertrand (1905), "On Denoting", in *Logic and knowledge*, London Routledge 1956, 39-56.

Shapiro, Stewart (2000), *Thinking About Mathematics*. Oxford.

Suárez, Mauricio (2004), "An Inferential Conception of Scientific Representation", *Philosophy of Science (Supplement)* 71:767-779.

Suárez, Mauricio (2009), "Scientific Fiction as Rules of Inference", in Mauricio Suárez (ed.), *Fictions in Science. Philosophical Essays on Modelling and Idealization*, London and New York: Routledge, 158-178.

Sugden, Robert (2000), "Credible Worlds: The Status of Theoretical Models in Economics", *Journal of Economic Methodology* 7 (1):1–31.

Toon, Adam (2009), "Models as Make-Believe", in Roman Frigg and Matthew Hunter (eds.), *Beyond Mimesis and Convention: Representation in Art and Science*, Berlin Springer.

Vaihinger, Hans (1911), *The Philosophy of 'As If': A System of the Theoretical, Practical, and Religious Fictions of Mankind*. English Translation London: Kegan Paul 1924.

Walton, Kendal L. (1990), *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge/MA: Harvard University Press.

Weisberg, Michael (2009), "Models for Modelling", *Unpublished Manuscript*.

Winsberg, Eric (2009), "A Function for Fictions: Expanding the Scope of Science", in Mauricio Suárez (ed.), *Fictions in Science. Philosophical Essays on Modelling and Idealization*, London and New York, 179-189.

Young, Hugh D., and Roger Freedman (2000), *University Physics with Modern Physics*. 10th ed. San Francisco and Reading (MA): Addison Wesley.