Models as fictions
A pretence theory of scientific modelling

Harm Faber

18 June 2012

Master's Thesis
Cognitive Artificial Intelligence
Utrecht University
30 ECTS

1st supervisor and reviewer	: dr. Janneke van Lith
2nd supervisor and reviewer	: prof. dr. Albert Visser
3rd reviewer	: dr. Stefan van der Stigchel

Universiteit Utrecht

Abstract

Scientific models are often not physical objects. They are simplifications of the real world, they are inaccurate, inconsistent and incomplete. This raises questions: what sort of entities are models, how do they represent their target system, and how can we learn from them?

The aim of this thesis is to investigate the idea that scientific models share a lot of aspects with fictions. An account of scientific modelling that views models as fictions could well provide answers to the questions above. In particular, Kendall Walton's *pretence theory* will be explored in order to formulate a theory of scientific modelling in analogy with his theory of fiction.

The main idea of Walton's theory is that objects (props) can induce and coordinate our imaginings. If we use props this way, we are engaging in a game of make-believe. What is true in a game of make-believe is a fictional truth. Fictional truths do not depend on whether or not they are imagined, it is the prop together with appropriate principles of generation that prescribe what is true in the game. By carefully constructing a model, we can use it as a prop in a game of make-believe. The model and the principles of generation then determine what is true in the world of the model. I will argue that this is a viable account of scientific modelling, but that there are some major challenges that are in need of more attention.

Preface

While thinking about topics for a thesis, I came across an overview of possible topics of which one was about the nature of models. Questions as 'how do models relate to scientific theories?' and 'what is the nature of scientific representation?' were given as example to get an idea of what direction a thesis on models could be heading. Since modelling is one of the core activities of the master's programme *Cognitive Artificial Intelligence*, I reckoned it could be a very interesting topic.

After a few meetings with Janneke van Lith, the outline of the thesis became clear. A short time later, Albert Visser joined as second supervisor. I would like to thank both of them for their advice and feedback. And I thank Stefan van der Stigchel for reviewing this thesis.

Thanks to my friends and family for their support. I especially want to thank my brother Herbert, for kicking my butt now and then.

Enjoy the read.

Contents

1. Introduction

Designing a model is often the first step when dealing with a scientific problem. We need models, because in most cases the real phenomenon or system is too complex. To overcome these complexity issues, we create a simplified version of the problem. We neglect air resistance or the fact that a sphere isn't a perfect, homogeneous sphere at all. We assume that predators and prey interact according to simple laws and in economics the exchange of goods is carried out by perfectly rational agents, without transaction costs, and dealings are done in no time.

These simplifications of the real world raise a lot of questions about the nature, the function and the workings of models. What sort of entities are models? What is truth in a model? Models are inaccurate, inconsistent and incomplete, how can they represent empirical phenomena?

1.1 Models

There are physical models, abstract models, scale models and mathematical models. A model can be a description, a graph, a diagram or an equation. But what is a model or what does the act of modelling amount to? According to Weisberg "modelling involves indirect representation and analysis of real-world phenomena via the mediation of models"[1]. If the model is sufficiently similar to the phenomenon that is being investigated, then the analysis of the model is indirectly an analysis of the properties of the real-world phenomenon.

1 Weisberg (2007, page 209).

Bailer-Jones defines a model as "an interpretative description of a phenomenon (object or process) that facilitates perceptual as well as intellectual access to that phenomenon"[2]. This description may focus on certain specific aspects of the phenomenon, or even deliberately disregard others. The model thus does not cover all the aspects of the concrete phenomenon it is applied to. It may even be the case that the model contains propositions that are known not to be true about the world.

For the purpose of modelling, the selection of aspects of a phenomenon is a widely accepted strategy. Thomson-Jones calls any passage in a scientific textbook, a journal or a classroom lecture, which has the following features: "(i) it has the surface appearance of an accurate description of an actual, concrete system (or kind of system) from the domain of inquiry, but (ii) there are no actual, concrete systems in the world around us which fit the description it contains, and (iii) that fact is recognized from the outset by competent practitioners of the scientific discipline in question"[3], as *a description of a missing system*. Nancy Cartwright calls a model a *prepared description* (Toon, 2010a), whereas an unprepared (or full) description would describe the phenomenon under investigation fully in all its details.

A simple and much used example of a *description of a missing system* or *a prepared description* is that of the harmonic oscillator (Giere, 2004; Toon, 2010a). A simple pendulum with a small amplitude or a mass hanging from a spring could both be modelled as $F = -kx$, where x is the displacement from the equilibrium position. In doing so we make a number of assumptions: we take the mass hanging from the spring to be a point mass which is subject to a uniform gravitational field and a linear restoring force exerted by a massless frictionless spring with spring constant k attached to a rigid surface. It is obvious that this does not perfectly reflect a real world situation where an actual mass hangs from an actual spring. Still, when we solve the equation in order to predict the position of the mass at any time after its release, we find that our predictions may be rather accurate.

It is clear that models are often simplifications, idealizations or approximations, but despite that they can instruct us about the nature of their target system. We assume that they represent the phenomenon we investigate, but in virtue of what is a model a representation of something else? How should we interpret them?

1.2 Representation

There is a vast variety of representational devices in science and they all represent their target system in some way. And also the target systems differ greatly in nature. A building can be represented by a scale model or by the drawings of the architect. In

2 Bailer-Jones (2003, page 61).
3 Thomson-Jones (2010, page 284).

both cases the target system is a concrete physical object and so are the representations. The billiard-ball model of gases can be described as a physical system, the target system is a state of nature (Suarez, 2003). And in the case of the mass hanging from the spring, the target system is a physical system and its representation a mathematical entity. The question is then, what do these different kind of models have in common that allow them to represent?

An intuitive answer to this question could be that the model is in some important aspects similar to its target system. It is typically assumed that model and target system are similar if they share a subset of their properties. How appealing this view may seem, it has some strong arguments against it.

1.2.1 Arguments against similarity

First of all, an account of representation solely based on similarity cannot account for the representational relationship of all different kinds of models. Of course a model is never completely similar to its target system, so we are talking about partial similarity here. A scale model of a building is in many ways similar to the actual building. But what about the mathematical model of the harmonic oscillator? The equation represents the physical phenomenon, but is not similar to it in any relevant respect.

A second argument against similarity is that, while similarity is reflexive and symmetric, representation is not. The scale model represents the building, but it would be wrong to claim that the building also represents the scale model.

Another problem for similarity is that of mistargeting. Although an account of similarity can deal with the fact that a model can be inaccurate or incomplete, it cannot deal with the possibility of being similar to a target system it does not represent. Similarity would then warrant representation.

If all logically possible properties would be permitted, any object is similar to any other object in an infinite number of ways (e.g. “not being black” or “being on the northern hemisphere”). So there is the need for a restriction on what relevant properties would be accepted for a particular representation. This is problematic, because this restriction would consist of the properties that pertain to the representation. The outcome of this would be something like: *A* represents *B* if and only if *A* and *B* are similar in those respects in which *A* represents *B* (Suarez, 2003). Hence, a circular analysis of representation.

These examples show that similarity is not necessary and not sufficient for representation. A mathematical model couldn't be more dissimilar from the phenomenon it represents, hence similarity is not a necessity. Neither is it sufficient; a model that represents its target system leads to its target, similarity alone cannot account for that. Of two similar objects, there is not one that leads to the other, unless they are in a representational relationship.

1.2.2 Arguments against isomorphism

Another view on how representation works is 'isomorphism'. The main idea of isomorphism is that models are structures. A structure $S = \langle U, O, R \rangle$ is a composite entity consisting of (i) a non-empty set U of individuals called the domain of the structure S, (ii) an indexed set O of operations on U, and (iii) a non-empty indexed set R of relations on U (Frigg, 2006). The model represents its target system if and only if the structure of the model is isomorphic to the structure of the target system. This means that there is a mapping between the structure of the model and the structure of the target system that shows a relationship between their properties.

Isomorphism can be understood as a form of similarity. If A and B are isomorphic, they are similar because they share at least one property, namely their relational framework. Two objects which structures are isomorphic are structurally similar. An example of such a similarity of structure could be a graph of a bridge. The graph is isomorphic to the bridge it represents with respect to the geometric shape and the the proportions between the different points (Suarez, 2003).

Some of the objections against similarity as a theory of scientific representation also apply to isomorphism. Just as similarity, isomorphism is reflexive and symmetric, while scientific representation is not. Furthermore, isomorphism is not sufficient for representation. Two copies of a photograph are isomorphic to each other, but neither is a representation of the other (Frigg, 2006). And the same structure can be instantiated by different systems. Hence, isomorphism is too inclusive.

Another argument against isomorphism is that a target system does not have one unique structure. Different conceptualizations may lead to different structures. The structure of a molecule for example, could be defined as the set of vertices being the individuals and the edges being the relations. But this is not the only natural possibility. We can also define the edges as the objects and the vertices as the relation between them (Frigg, 2006).

If *intention* is thrown into the picture, most objections to similarity and isomorphism are cleared. But the addition of a clause stating that the scientist must intend that the model represents the target system, comes with other difficulties. First of all, we need to understand how a scientist comes to use a model as a representation of a target system. But more important, it seems that the appeal to intention would do all the work and make similarity and isomorphism irrelevant in explaining why a model represents a target system. Thus, the notion of *intention* doesn't help those who defend an account of *similarity* or *isomorphism* as scientific representation. However, it would be prematurely to conclude that *intention* does not have any role in scientific representation.

It is not the contention of this thesis to argue that *similarity* and *isomorphism* are not viable theories of scientific representation. This short overview of difficulties for these two theories is given in order to justify the exploration of a different, promising approach. An approach which core is a theory of fiction. In fact, I will argue that *similarity* and a form of *intention* do have a role in such an account. However, it is not the actual graph or equation that has to be similar to the target system, it is what we imagine about by virtue of the graph or equation that has similarities with the target system. And *intention* is not needed to account for the representational relationship between model and target system, but to identify for what purpose the model is created.

1.3 Research question

The objective of this thesis is to investigate the idea that scientific models share a lot of aspects with fictions. Models are often not physical objects, what sort of entities are models then? It is not uncommon for a model to be inconsistent or incomplete, in other words, in many cases models are false of any real system and cannot actually exist. How do they represent their target system and how can we learn from them? An account of scientific modelling that views models as fictions could well provide answers to those questions.

In particular, Kendall Walton's *pretence theory*, as described in his book '*Mimesis as make-believe: on the foundations of the representational arts*' (1990), will be explored, as it is one of the most influential works on fiction and pretence. Moreover, the potential of his theory with respect to scientific modelling has already been recognized by philosophers as Roman Frigg (2010a) and Adam Toon (2010a).

From this we can distil the main research question of this thesis:

Is Kendall Walton's *pretence theory* viable as a framework from which to formulate a theory of scientific modelling?

1.4 Outline

Chapter two consists of an introduction to Kendall Walton's theory of *pretence*. I will discuss notions as *games of make-believe*, *prescriptions to imagine*, *fictional truth*, and *principles of generation*. Because the objective of this thesis is to formulate of theory of scientific modelling along the lines of Walton's theory of pretence, chapter two will primarily be an outline of this theory and not so much a critical review. By first giving an overview of Walton's theory instead of a review, the foundation for the main part of this thesis will be as clear as possible. In the following chapters I will, of course, develop a more critical stance towards Walton's theory.

In chapter three I will explore this theory of fiction to find analogies with scientific modelling. I will discuss to what extent we can use Walton's ideas in the context of modelling, and on the points where we have to leave Walton I will, if necessary, propose alternative solutions. By the end of the chapter we will have an account of scientific modelling analogues to fiction.

In chapter four I discuss two examples of models to see what the consequences are of this theory of scientific modelling. The first model on which I will test the findings of chapter three is a fairly recent *motor control* model. By applying the theory to this model I will show how the proposed account works in practice. Moreover, I will use this example model to discuss some details of the theory that remain controversial. The second model is the much older, long since disproven geocentric model of Ptolemy. I chose this model to investigate how the theory deals with models that turn out to be false. We will discover that this second example forces us to make different choices as we made in the first example.

Chapter five will be a general discussion and conclusion. I will again briefly discuss the benefits of an account of scientific modelling along the lines of fiction and I'll point out on which points there remain difficulties. And, of course, I will give an answer to the main question of this thesis.

2. A theory of pretence

In this chapter, a theory of pretence developed by Kendall Walton in his book '*Mimesis as make-believe: on the foundations of the representational arts*' (1990), is introduced. This theory is the basis on which we will explore the opportunities and difficulties of a fictional approach on scientific models.

2.1 Games of make-believe

Walton starts off by stating that, in order to understand representational works of art, we need to regard the activities in which those works are embedded as we would regard children's games of make-believe. Children playing with a toy truck imagine of the toy that it is a real truck. The toy coordinates their imaginings. But also an object that does not have the explicit function to do so, can prompt imaginings. A stump in the forest could induce imaginings about a bear, especially if the stump is somewhat shaped as a bear. Objects that induce imaginings are referred to as *props*. Someone who is encouraged by a prop to imagine things, is engaged in a game of make-believe and is *pretending*. Participation in such a game can thus be described as *pretence.*

Props can be used to coordinate collective imaginative activities. Children playing in the woods could agree that every stump they encounter is a bear in their game. With this agreement, they can all participate in a collective game of make-believe. By constructing an artificial prompter we can, in the same way, share our imaginative thoughts with others.

The things that a person imagines about are *objects* of his imagining. In the case of the stump for instance, it is imagined of the stump itself that it is a bear. The stump is the *object* that, as it were, carries the bear. So props can be objects of imagination, but in other cases the prop is not an object of the imagination. The images of a film are props, but they are not objects of imaginings.

2.2 Fictional truth

If something is *true in a game of make-believe*, for example that there is a bear ten meters to the left (where in fact is a stump), this is a *fictional truth*. Or more general: whatever is the case in a fictional world – in a game of make-believe, in a dream or in a representational work of art – is *fictional*. But what exactly is fictionality? What is the relation between fictionality and imagination?

We may not identify fictionality with imagination. What is fictional need not be imagined and what is imagined need not be fictional. In the game where all stumps are bears, two players encounter a stump and imagine it to be a bear. Further investigation reveals that, in fact, there is no stump. It was a moss covered boulder. They were, in the world of the game, mistaken in thinking that there was a bear. They imagined that there was a bear, but it was never fictional that there was a bear.

Without the two players realizing, there is a stump right next to the boulder, hidden in the thicket. They don't notice the stump and thus do not imagine that there is a bear. But fictionally there is a bear lurking in the thicket. It is the stump that makes it fictional that there is a bear. Props generate fictional truths, by virtue of their existence they make propositions fictional.

However, props cannot generate fictional truths entirely on their own. It is because of the convention or agreement in the game of make-believe that a stump makes it fictional that there is a bear. This is called a *principle of generation.* A principle of generation can be established by explicit stipulation, such as the agreement that a stump is a bear, but in many cases they are implicit. It is, for instance, implicit that bears could be dangerous, or even that they have muscles and blood vessels.

What is left in this account of fictionality is the connection between imagination and fictional truths. The key to the notion of fictional truth lies in the realization that imagining is constrained. Some imaginings are appropriate in a certain context, others are not. "A fictional truth consists in there being a prescription or mandate in some context to imagine something."[4] Someone who imagines something that isn't appropriate breaks the rules of the game and plays improperly. Furthermore, fictional properties are properties that are to be imagined, whether or not they are in fact imagined.

The role of props in generating fictional truths is not to be underestimated. It is not just the essential role they play in generating fictional truths, they give fictional worlds

4 Walton (1990, page 39).

and their contents objectivity. Because of the role props play, fictional worlds have an independence from the players of the game. “It's not thinking that makes it so; the prop does.”[5]

2.3 Representations

Not only stumps or toy trucks can be props in games of make-believe. Most representational works of art are props; novels, paintings, plays, etc. What stumps and novels have in common is that they prescribe imaginings and generate fictional truths. But there are differences between the two. Assuming that the *bear game* was invented on the spot, the stump is an *ad hoc* prop, its function agreed upon for a single game of make-believe on a single occasion. A toy truck and a novel are *designed* props which are made specifically for a certain purpose and for certain kinds of games of make-believe[6]. The toy truck is designed to count as a truck in the kind of games that are authorized for it.

Representations can now be defined as *things whose function it is to be props*. *Function* can be thought of as a set of meta-rules, rules about what sorts of games are to be played with the work, and must be regarded in a loose, not very restrictive sense. Something may have the function to serve as a prop, regardless of the intentions of the maker, if things of that kind are typically meant by their makers to serve as a prop. Or if it is common practice or convention in a social group to use it as a prop. Functions are thus society relative.

Of objects that are recognized as props, or representations, in games of certain sort, the principles of generation that apply do not have to be set up every time a game is played. The principles that apply are likely to seem natural and are accepted automatically. Creators of props use this to direct peoples imaginings by designing their props appropriately.

2.4 Fictional worlds

What is it for a proposition to be true in a given fictional world? First of all, we should not confuse fictional worlds with possible worlds. Fictional worlds can be impossible worlds and are usually incomplete, whereas possible worlds are necessarily both possible and complete.

---

5 Walton (1990, page 42).

6 Walton sets *ad hoc* props against *designed* props. This is not a good contrast. If it is agreed that stumps are always bears, every time the children are in the woods, then we can't continue to call the stumps *ad hoc* props any more. But they aren't *designed* props either.

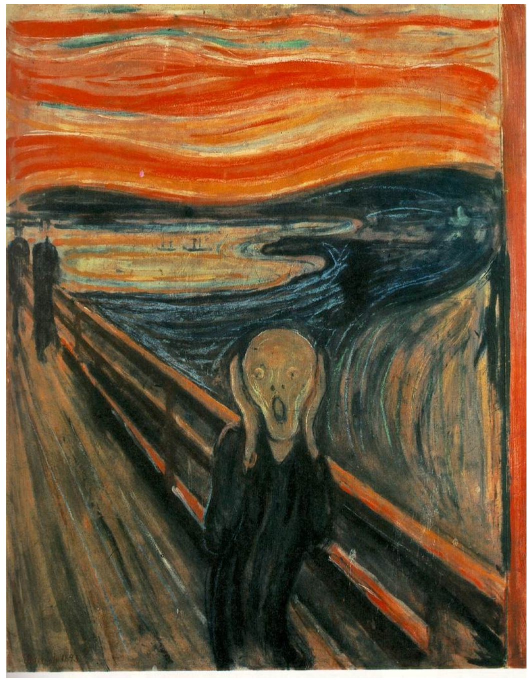

Figure 1: The Scream, Munch

We have seen that fictionality can be explained in terms of prescriptions to imagine. If someone is looking at Edvard Munch's painting *The Scream* (figure 1) and contemplates on it, there is the world of his game, the *game world*. But, there is also the world of the work itself, the *work world*[7]. Propositions that are true in the world of the work are those which are generated by the work alone and which are fictional in any game that is played with the work.

The *game world* and the *work world* are easily confused, mainly because of the fact that they share many fictional truths. The principles and the prop prescribe that it is to be imagined of *The Scream* that there are three people walking along a fence, one of which is screaming. In the game world as well as in the work world it is true that there are three people walking along a fence and that one of them screams. However, the viewer of the painting, who imagines that *he* sees the screaming man, is part of the game world, but not of the work world. The viewer is not part of the painting itself, he is not among the characters he is looking at. The distinction between the two worlds is necessary, otherwise we could never decide which of the games that different appreciators play is to be identified with the world of the work.

Which fictional truths belong to the work world? People can play all sorts of games with a given work. It is the function of the prop that makes the distinction. The function of the prop is to serve as a prop in certain sorts of games involving principles of generation which result in the fictionality of particular propositions. Thus,

7 It is questionable if this distinction between worlds is necessary for a theory of scientific modelling; we will discuss this in the next chapter.

propositions that are fictional in unauthorized games are certainly not part of the work world. However, this does not mean that all propositions that are fictional in authorized games are. Remember, the viewer of Edvard Munch's painting, who imagines that he sees a man screaming alongside a fence with the Oslofjord in Norway in the background, plays an authorized game, but it is not true in the world of the work that *he* sees that man screaming. What is true in the world of the work must be true in any game in which it is the function of the work to serve as a prop, and whose fictionality in such games is generated by the work alone.

In other words, propositions are fictional if there are prescriptions to the effect that they are to be imagined; which world a proposition is fictional in is a matter of who is subject to the prescription. What is fictional in a work is what appreciators of it (qua appreciators of it) are to imagine. Different kinds of contexts and their associated worlds are distinguished by what is responsible for the prescriptions.

Worlds of make-believe have a lot of advantages over reality. They are much more malleable and if we let others construct them, we can enjoy the benefits of their special talents or insights. We can share games of make-believe and play them together, without compromising the objectivity of the fictional world or the spontaneity of our imaginings. It sounds almost to good to be true.

2.5 Objects

Ordinarily, to represent something means that there is a *thing* that is being represented. *A Thousand Splendid Suns*, *a* novel about the recent history of Afghanistan seen from the perspective of two women, makes it fictional of Afghanistan that it exists, that the people suffered from a devastating war against the Soviets and that when the Taliban took over the situation for women took a turn for the worse[8]. Generally, a thing is an object of a given representation if there are propositions about it which the representation makes fictional. This means that a thing is represented *as* something; to represent someone as being tall is to make it fictional that he is tall. It is a matter of what propositions about an object a work makes fictional.

We have to be careful not to confuse representing with matching. Matching is *complete* correspondence between a representation and the world. Informally, a picture of a man matches a man if that man is in *every* detail exactly like the man in the picture (if we assume that there are such things as people in pictures). A work can be a representation of something it does not match, or can match something it does not represent. The first case is called *misrepresentation*. Propositions that are made fictional of an object need not to be true to be about it. If a story makes it fictional that

8 Facts are *shared* between the real world and the fictional world.

Obama lost the 2008 presidential elections to McCain, the story represents Obama nonetheless.

But what about matching without representation? Suppose that Tom Sawyer has a real world double, with the same name and who was and did everything that the Tom Sawyer described in Mark Twain's novel was and did. Mark Twain did not know of this real Tom Sawyer when he wrote the novel. The real Tom Sawyer and the one from the novel match, but the real Tom Sawyer is not an object of the story. The story is not about him, readers are not expected to have imaginings about the actual Tom Sawyer.

Then what is it that determines whether a work represents anything or not? This depends on the principles of generation that are in effect. We will take a closer look at these principles in the next section, but in the mean time we look at some aspects that appear to have, in different contexts, part in determining the objects of representations. When using the name of a well-known person ("Napoleon") in a literary work, there is a linguistic reference to the object. Perhaps the person should be reasonably well-known to the intended audience, for it would be otherwise difficult to expect there being prescribed imaginings about him. And the author should knowingly use the name; accidental uses of a well-known name cannot lead to representation. The same goes for pictures that represent by means of a title, a title should be deliberately chosen. An accidental use of an existing name or location does not do the trick. "It will suffice for now that, typically, some combination of titles and like signs, artist's intentions, and other causal relations, together perhaps with a certain degree of correspondence, serves to establish the relation of representing."[9]

Objects can be used for practical purposes, for instance to create a certain setting for the fictional world. Locating a story in the city of London, for example, provides a background of a large, modern, industrial city in a nation with a democratic tradition and with certain social structures, without having to specify all those (and many more) details which would distract from the story. But a representation can also contain statements or assertions about real things. Using objects this way, they can help deepen one's understanding of the thing that is represented by prescribing imaginings about it.

Representations can be their own objects. The toy truck that was mentioned earlier for example, not only directs players to imagine a truck, but to imagine the toy truck itself to be a truck. It generates fictional truths about itself. Literary fictions in the form of letters or journals are also reflexive, they make it fictional of themselves that they are in fact a letter or journal. Other representations do not have an object at all; a picture of a unicorn for example. The notion of having objects should not be taken as basic in explaining representation, to represent something is not to denote something. Representation is best explained in terms of games of make-believe and the generation of fictional truths. Whether we read Cervantes's *Don Quixote* or we're looking at

9 Walton (1990, page 112).

Shakespeare's *Julius Ceasar*, in both cases we are to imagine about a person, although in case of the latter we actually know about a particular person and in the other one we do not. It does not especially matter whether or not a representation has an actual object. The difference consists in whether we are to imagine about an actual person, but in either case we are to imagine seeing someone.

2.6 Principles of generation

Imaginings are mandated by props, in accordance with applicable principles of generation. This raises the question what those principles of generation are exactly. What principles of generation are applicable to a given work?

In most cases, when we read a novel or look at a painting, it just comes to us that such and such is fictional. In case we are conscious about how we come to these conclusions, which most of the time is not the case, the considerations at work seem somehow reasonable in that specific case. But are there relatively simple, systematic principles that implicitly govern the generation of fictional truths?

Fictional truths can be generated directly (primary) or indirectly (implied). Implied fictional truths are ones that depend on other fictional truths, primary fictional truths do not. Remember the aforementioned painting by Edvard Munch, it is a primary fictional truth that there is a screaming man alongside a fence with the Oslofjord in Norway in the background. We may assume that this man has blood in his veins, even if this is not portrayed[10]. It is an implied fictional truth, just as it is implied that he has to eat and sleep, that he has ambitions and disappointments, that he lives on a planet that circles the sun. All this is implied, in the absence of contrary indications, by the fact that fictionally it is a human being. This is not to say that in general, appreciators just *infer* implied fictional truths from those on which they are based. It can be the indirectness of the generation of a fictional truth that can give it prominence.

The investigation to the way by which fictional truths are generated will have to take account of principles of two different kinds: principles of direct generation, those which explicitly say that if a work contains certain words or marks or whatever, such and such propositions will be fictional, and principles of implication, those which specify what fictional truths are implied given the core of primary ones. The latter raises the question what it is that determines which fictional truths imply which others. Are there principles of implication that, given a core of primary fictional truths, determine what fictional truths are implied? Two such principles will be discussed.

10 Default reasoning; propositions may be assumed by lack of evidence of the contrary.

2.6.1 The Reality Principle

The basic idea of the Reality Principle is to make a fictional world as similar to the real one as the core of primary fictional truths permits. If the primary fictional truths do not indicate otherwise, it is implied that fictional characters have blood in their veins, because that is the case in the real world. In other words, a fictional truth is legitimately implied if it follows the lines of what would be a legitimate inference in the real world. Or, more formal:

"If $p_1,...,p_n$ are the propositions whose fictionality a representation generates directly, another proposition, q, is fictional in it if, and only if, were it the case that $p_1,...,p_n$, it would be the case that q."[11]

It is not necessary for an interpreter of a work to make a list of all primary fictional truths and then trace all implications it generates. For one thing, there is no assurance that the number of primary fictional truths in a work is finite or finitely specifiable. And it may not always be clear whether a particular fictional truth belongs to the list of primary ones. The interpreter focuses on a selection of fictional truths that seem to be unquestionably generated and note their prima facie implications, without worrying whether they are directly or indirectly generated.

There are some problems for the Reality Principle. Works of art, novels, paintings, can contain contradictions. Fictional worlds shown in the works of M.C. Escher for instance, are often impossible. Never-ending stairs that people ascent while not gaining height or water that streams upwards and downwards at the same time, are among the things he sketched. But at the moment that two contradicting propositions are true in a fictional world, it can be concluded that everything is fictional in that world, since a proposition counterfactually implies whatever it entails and a contradiction entails everything (principle of explosion)[12]. If we look at such a work of Escher, we want to accept the primary generated truths and the implicit ones generated by the former truths, but it seems unreasonable to suppose that it is also fictional that there are gnomes on Mars watching television or that the South Pole was the number one holiday destination for the Dutch in 2011.

A solution could be sought in construing counterfactuals in such a way that not all of those with contradictory antecedents turn out to be true. Or we could define the Reality Principle in such a way that unwanted implications are blocked. But also fictional worlds that are not contradictory have the problem of excessive amounts of propositions that are true in them. If a fictional world should be as much as possible as the real world, than most fictional worlds will include almost everything that is true in the real world[13].

---

11 Walton (1990, page 145).
12 Counterfactuals need a complete world; the fictional world is not complete.
13 Why not think of the fictional world as one of which we construct an image from the information that is available to us as opposed to one that is build of propositions?

A better, more effective solution would be just to ignore all this clutter. Besides presenting us with a collection of fictional truths, representations can also order and arrange them for us, focusing on some more than on others. Unimportant details for the story of the work, how important they may be in the history of our Earth, have their position in the deep background of the fictional world and will not bother us as long as we leave them alone.

2.6.2 The Mutual Belief Principle

If implied fictional truths depend on facts of nature, as is the case with the Reality Principle, then the artist's control over them loosens. For a storyteller in a culture where the universal belief and agreement holds that the earth is flat and that to venture too far out to sea is to risk falling off, there is no reason to mention the shape of the earth explicitly. But according to the Reality Principle, it is fictional that the earth is spherical and that there is no danger of falling off. It seems more fitting to go along with the misconception about the shape of the earth, in order to appreciate the story as it was intended, than to force it out of context with an interpretation that neither the author nor his intended listeners could have imagined[14].

This suggests that there are principles at work that give artists a more reliable control over what is fictional than the Reality Principle does. In the example of the flat earth, it seems that the storyteller bases his story on what is commonly believed (in the community of the storyteller), rather than on the real world as it actually is. This leads to principles of implication based on mutual belief[15]; the Mutual Belief Principle:

“If $p_1,...,p_n$ are the propositions whose fictionality a representation generates directly, another proposition, q, is fictional in it if and only if it is mutually believed in the artist's society that were it the case that $p_1,...,p_n$, it would be the case that q.”[16]

Beliefs need to be mutual, because if they weren't, fictional worlds would be insufficiently accessible to appreciators and it would take fictional worlds out of the artist's control, insofar as he must guess what others believe. So the Mutual Belief Principle maximizes the similarities between the fictional world and the real world not as it actually is but as it is or was mutually believed in the artist's society.

---

14 It could be argued that the Reality Principle is still to prefer, even in this case. Although the author believed otherwise, there is in fact no danger of falling off the earth. With the Reality Principle in effect it is not fictional in the story that there is a risk of falling off the earth. The author and his audience just have false believes about that.

15 Something is 'mutually believed' in a society if, roughly, most members of the society believe it, most of them believe that most of them believe it, most believe that most believe that most believe it, and so on.

16 Walton (1990, page 151).

Is the Mutual Belief Principle an improvement over the Reality Principle then? Certainly it gives the artist more control over what is fictional and it's often easier for appreciators to determine what is mutually believed in a society than what is true. But there are situations or types of representations where the Reality Principle is to be preferred. Natural representations, like stumps, are not designed by artists in order to direct the imaginings of appreciators. But this could be resolved by accepting that the mutual beliefs of the players of the game are at work.

It becomes more difficult for the Mutual Belief Principle when fictional truths having to do with morality come into the picture. It is one thing to accept for the sake of the story that it is fictional that the earth is flat, but it is quite another to accept that slavery is just and that the mixing of the races is evil and its repression a moral necessity. It is by no means evident that we should set aside our own moral convictions in order to be able to understand and appreciate the story.

When participating in a game of make-believe, using representations as props, the appreciator makes it fictional that in his game he investigates reality in certain ways. The Reality Principle seems the most suitable for this kind of participation. But when we treat works from a different culture than our own, we feel the need to replace the Reality Principle with the Mutual Belief Principle; then we take people's beliefs about reality in order to trace implications.

In sum, the Reality Principle makes for richer and more natural participation in games of make-believe, the Mutual Belief Principle gives the artist better control over what is fictional, his immediate audience better access to it and it better facilitates the use of representations by artists to direct the imaginations of appreciators.

Whichever principle is preferred, there are counterexamples that can be thought of. Moreover, there are examples in which it is evident that certain fictional truths are implied, but where the implication is not sanctioned by the Reality Principle or the Mutual Belief Principle. There are even cases where implications that seem not to hold, ought to hold according to one or both of the principles.

An example of the first kind could be a line drawing of a horselike animal with a single horn in the middle of its forehead. It is obvious that the drawing depicts a unicorn and that it is white, because unicorns are white. Of course we are aware of the fact that unicorns are mythical beasts, but if they would be real, they would most likely not be white, as most horselike animals in the real world aren't white. Neither the Reality Principle nor the Mutual Belief Principle sanction the implication of the animal being white, but still it is unquestionably fictional that it is white.

In *'Allo, 'Allo!*, a British television series set up in a small town café in Nazi-occupied France, every character, French, German or British, speaks English. However, this doesn't hold, the viewer is perfectly aware that it is fictional that the Germans speak German and the French speak French. But the Reality Principle as well as the Mutual Belief Principle sanction the fictional truth that everyone speaks English. Perhaps it could be argued that the use of accents imply that a certain language is spoken, but this line of reasoning would not work for similar situations in literary works of fiction. Charles Dickens' *A Tale of Two Cities* takes place in England and France, but both the English and the French use the English language in their conversations. It is understandable why Dickens chose to not to use French in his novel, it would make the story unreadable for his intended audience, but with the Reality Principle or the Mutual Believe Principle the consequence is that it is made fictional that the French characters in the story speak English.

It turns out that it is difficult to find a simple or systematic set of principles that govern the indirect generation of fictional truths. It rather seems to be a complicated and shifting array of understandings, precedents and local conventions. Divergent principles, answering to different needs, seem to be at work in different cases, and it seems unlikely that there are general or systematic meta-principles for determining which is applicable when. The appreciator needs his knowledge of the arts, of society and of the world, in order to feel his way in the game.

2.6.3 Direct generation

Are the principles of direct generation any more straightforward or systematic than the implicit ones then? It certainly looks like they do at first; literary works just *say* what is fictional, paintings *show* it. But there is a lot more to it than one would think at first.

We must resist identifying directly generated fictional truths with what the work makes *explicit*. The words of a novel do not necessarily generate fictional truths. It could be fictional that a narrator utters the words of a text, and if that narrator is reliable, this would *imply* the fictionality of what the words express. Does a picture always generate directly the fictionality of what is shown in them? "Rousseau's painting *The Dream* (figure 2) shows an elephant, a pair of tigers, and a snake charmer in a jungle scene, in the midst of which the dreamer sleeps. But it is fictional not that there is an elephant and so on, but merely that the dreamer dreams that there is. And it is fictional neither that she sleeps in a jungle nor that she dreams of doing so, though it would seem that this state of affairs is 'shown'."[17]

Artists are very creative in finding original ways of getting fictional truths generated. Some of them are more or less traditional, others are ad hoc. Some of them leave no doubt about what is fictional, others keep us guessing. Some require familiarity with the genre to be understood, others don't. Often we need just common sense to understand which fictional truths are generated and which are not. Is it fictional that the soundtrack of a film is heard by the characters in that film? If a band is shown that appears to be playing, it probably is. But if the scene is a lone rider in an empty

17 Walton (1990, page 170).

Figure 2: The Dream, Rousseau

desert, it probably isn't.

Does a halo around the head of a saint in a painting mean that fictionally a ring of light hovers around the saint's head which in turn implies that it is a saint? Or does the yellow ellipse on the canvas directly generates the fictional truth of the character in the painting being a saint? It appears that the generation of primary fictional truths isn't any more structured or orderly than that of the implied ones.

What are the consequences of this disorderly behaviour of the machinery of generation? It seems somehow that there must be more regularity; how can we ever learn to apply these rules if they are as complex and unsystematic as they seem to be? How can we justify considering an attribution of a fictional truth to be right or wrong? If fictional truths are generated in such different manners, then is fictionality even a *natural kind*[18] on which a theory can be built? Fictionality consists in prescriptions to imagine, it is not *defined* by principles of generation. It is the establishment of these prescriptions that happens in various ways.

But how can we master such complex rules? An answer can be given along the lines of Wittgenstein. "It is an inescapable fact that many concepts are applied without the aid of rules or formulas; otherwise language would be impossible. How do we decide whether something is *sweet smelling* or *red*? We sniff or look, and it just seems to be so or not."[19]

So the generation of fictional truths by a work is just something we recognize when experiencing it in its context. We should consider the principles of generation not as guides for the generation of fictional truths, but rather as reconstructions of our judgements.

18 It probably isn't even necessary for fictionality to be a *natural kind.*
19 Walton (1990, page 185).

2.7 The barrier between worlds

When we read *A Thousand Splendid Suns*, we cannot rescue Mariam and Laila, the two women, from their abusive husband Rasheed, or do anything to make their lives any more tolerable. And we cannot interact in any way with the screaming man in Edvard Munch's painting. Fictional worlds seem isolated from the real world, separated from it by a logical or metaphysical barrier.

On the other hand, we seem to be in *psychological* contact with Mariam and Laila. We worry about their well-being. When we look at *The Scream*, we feel to a certain degree the character's anxiety. We learn about the fictional world, we have epistemological access to it, and we respond to this knowledge in many of the ways in which we respond to what we know about the real world.

Then is there a selective barrier that allows psychological links between appreciators and characters, while preventing physical interaction? Physical and psychological relations are so closely intertwined that such a selective barrier would be problematic.

Suppose a very naive chess player is one turn away from being checkmated. He is on the verge of surrendering his king to the enemy, but at the last moment inspiration strikes him and he makes a lifesaving move that saves his king. He is mistaken, of course, that he can save the piece. It is not in danger. But the king it portrays on the battlefield is in danger and does need saving. Can the chess player save his king, despite the fact that he does not live in his world?

Real world novels, plays, paintings and so forth determine what happens in fictional worlds. By affecting those props in the real world, we affect what is the case in the fictional world. The physical isolation of fictional worlds from the real world seems to have vanished. But if it is so easy to interfere in the fictional world, why don't we do it more often?

There are two questions to be asked: Is it true that the chess player saves the king? Is it fictional that the chess player saves the king? The answer to both questions is no. It is not true that the king is in danger or even that he exists, so he cannot be saved. And it is not fictional that the chess player exists, hence it is not fictional that he saves anyone. But it *is* true that the chess player makes it fictional that the king survives, by arranging things in such a way that this fictional truth is generated. But this is neither really nor fictionally *saving* the king. A person can only save another if they live in the same world. Cross-world saving, interaction between worlds, is excluded.

But what about the psychological bonds that we feel to fictions? To allow fictions as objects of our psychological attitudes while disallowing the possibility of physical interaction severs the normal links between the physical and the psychological. The question arises if we actually have psychological attitudes towards fictions.

Let's consider a moviegoer watching a horror movie. It is not unthinkable that the moviegoer confesses at the end that he was terrified. But *was* he terrified? He was, assuming he is not naive, fully aware of the fact that the images he saw were not real.

He was not actually scared, for that would take a genuine believe of the movie being real. But still his hart rate could have been considerably higher than normal and his palms could be sweaty.

The moviegoer participated psychologically in a game of make-believe. His fears were fictional, in quite the same way as children who are afraid of stumps in the forest, imagining them being bears. Although there are a lot more interesting things to say on this topic, that goes far beyond the scope of this thesis.

2.8 Participating versus observing

Participants in a game of make-believe need to consider themselves constrained to imagine the propositions that are fictional in the game. They acknowledge that the rules or principles of generation apply to themselves. These are minimal conditions for participation. Mere onlookers are not subject to the rules of the game, they do not consider the fictionality of a proposition a reason to imagine it. But this does not mean that they can't have a great interest in the game. Onlookers, critics or historians of the arts for example, are perfectly able to study a game and its props and find out what is fictional and what principles of generation apply. They can analyse and explain the game without taking part in it.

Besides being objects, participants in games of make-believe are also props. It is by virtue of the fact that the children playing in the woods, imagining stumps to be bears, that fictionally they are walking through a forest full of bears. Props are almost always reflexive. The existence of the stump makes it fictional that it is a bear. And the various properties the stump has almost automatically make it fictional that the bear is of a certain sort. A big stump means fictionally a big bear. If the stump is on a hill, then fictionally the bear is on a hill. The hill is also a reflexive prop, it is fictional that the bear is on *it*.

Participants as well as mere onlookers, are thus enlisted in the game as reflexive props. If the mother of one of the children, not participating in the game, stands near a stump, it is fictional in the game that she stands dangerously close to a bear.

Participants in games of make-believe are props, objects and imaginers at the same time. They prescribe imaginings, imaginings about themselves by virtue of the fact that they themselves do the prescribing, and it is to themselves that they issue the prescriptions.

2.9 Unofficial games

Besides the games we play that are authorized for certain works of fiction, we often make up our own games or modify authorized games such that the principles by which they generate fictional truths are altered. Walton calls these sort of games *unofficial* games of make-believe. Saying of an actress playing Julia every night that 'she already died 6 times this week' is an example of a statement made in pretence in an unofficial game. 'The crowd decapitated Saddam Hussein' is another example of such a statement, when it is said referring to the decapitation of a statue of Hussein.

For our purpose, the most interesting kind of unofficial games are those where comparisons are made between different works of fiction. 'Sherlock Holmes is more brilliant than Hercule Poirot' is such a statement. According to Walton, Holmes and Poirot are both characters in the world of the combination game and each brings to it the degree of brilliance he exhibits in his home world.

What is someone asserting making such a statement and under what conditions is his assertion true? He is asserting something true if and only if it is fictional in the game of the implied sort that he speaks truly. What he is asserting is that the situation is such that to pretend in the way exemplified or indicated is fictionally to speak truly in that game.

2.10 Ontology of fictitious entities

Do fictitious entities exist? If we are asked whether there are such characters as Tom Sawyer or Don Quixote, the first answer that comes to mind is that indeed there are. But at the same time we are perfectly aware that they are just stories and that they never were. If we are asked about unicorns and fairies, we tend to deny that there ever were any, but if we are reading a text we allow that there are, in fiction, unicorns and fairies.

These conflicting intuitions make us uncertain about the ontological status of fictional entities. Do we have to make some sort of distinction between *existence* and *being real*? A distinction where Tom Sawyer and Don Quixote do not *exist*, yet *are*? Or vice versa? Such a distinction is unappealing, it seems to be nothing more than a trick to camouflage a contradiction.

On the other hand, to deny fictions any sort of existence or being, while insisting that nevertheless we can refer to them and talk about them, also poses problems. The ordinary claim that there is such a character as Don Quixote could then easily be rejected. Moreover, it would imply that causal theories of reference would be unavailable, since nonentities cannot be causes. How would a theory of reference be construed?

It is clear that both sides have problems. Those who reject fictional entities any form of existence must account for the apparent references to fictions in ordinary discourse. If there is no Don Quixote and we know it, what are we saying when we say that he had a horse named Rocinante? But those who allow fictional entities to exist, must account for the fact that we, in certain contexts, deny that there is a Don Quixote or that there are unicorns and fairies. Moreover, they must explain what sorts of things these are, what properties they have.

It is the notion of make-believe which is the key to these semantic and metaphysical questions. It seems obvious that, whatever the specifics of the theory, some such notion must have a prominent place in any adequate account of the institution of fiction. We *pretend* to refer to fictions, combined with a serious interest in this pretence. And when we assert that Don Quixote has a horse named Rocinante, we pretend to describe the real world rather than actually describing a fictional one[20]. This is not to be mistaken for genuine ontological commitment. "We are so deeply immersed in make-believe that it infects even theorizing itself. Our job is to extricate ourselves enough to be able to see how pervasive it is".[21]

> *"Objectivity, control, the possibility of joint participation, spontaneity, all on top of a certain freedom from the cares of the real world: it looks as though make-believe has everything."*
>
> \- Kendall Walton

20 Walton introduced the notion of *fictional truth*. Along that line we could introduce *fictional existence*; we then have a fictional world in which things exist.

21 Walton (1990, page 390).

3. Models and Fiction

3.1 The appeal of fiction

What is the appeal of a theory of fiction that makes it worthwhile to investigate it as the basis of an account of scientific modelling? In chapter 1 we established that scientific models are inaccurate, inconsistent and incomplete. Thus, in their efforts to get a better understanding of the empirical world, scientists consider things that are known not to be parts of that world. Things that do not exist, but would have been concrete, physical things if they did. Those imaginary things are investigated, sometimes collaboratively in a community of scientists, and surprising properties might be uncovered. There seem to be analogies to fiction (Godfrey-Smith, 2009).

Scientists do not talk about the Newtonian model system of sun and earth as something purely mathematical or structural. They talk about it as, if it were real, it would consist of two spherical bodies with mass, colour, density and other concrete properties. Properties that structures do not have. A model in evolutionary biology is best understood as describing hypothetical physical populations, not as some abstract structure. Thus, it seems that hypothetical systems are an important part of the theoretical apparatus we employ in scientific modelling and that they therefore have to be included in an analysis of how scientific modelling works (Frigg, 2010a).

If we use a model to learn about the target system, we compare properties of the model with the target system. The model and the target system can both be physical systems, a scale model of an aeroplane in a wind tunnel gives us information about the air resistance of a real aeroplane. But comparisons can also be made between two fictional systems or between a fiction and a physical system. We are perfectly capable to compare the world of Tolkien's *Middle Earth* with the world of Malory's *Le Morte* *d'Arthur*. And we can compare the events in Orwell's *Animal Farm* to those in Russia in the first part of the twentieth century (Godfrey-Smith, 2009). In each case we compare properties of one system with that of the other. In the last two examples it is a comparison between features *attributed* to things in fiction, or properties that the system *would have* if it were actual, with properties of the target system. It seems natural to treat the hypothetical systems or fictions in a similar way as we treat the physical ones.

3.2 Finding analogies and disanalogies

In this section I will first discuss analogies between model systems and a pretence theory of fiction. Then I will focus on some parts of Walton's theory of which I do not think that the analogy between model systems and his theory is valid.

Finding analogies between model systems and fictions according to Kendall Walton's *pretence theory*, starts with the recognition that models do not describe *actual* systems, just as novels are not a true description of anything in the world. We can make claims about a model, beyond what is explicitly stated, just as is the case with fiction, and we can learn about this *extra content* using rules of inference. And we can compare features of a model system with that of the target system, just as we can compare what we read in a novel with what is the case in the real world (Frigg, 2010a).

Walton begins his theory of pretence by stating that objects can induce and coordinate our imaginings. A child that plays with a toy truck imagines that it is a real truck, the toy truck induces and coordinates these imaginings. The child is engaged in a *game of make-believe.* Objects that induce imaginings are called *props*, and a prop is a representation if it is a prop in an authorized game. The analogy here is that the model is the prop, the model induces imaginings about it. Scientists that are using the model are engaged in a game of make-believe, they are pretending.

At this point we have to make a distinction between the model that is a prop which induces imaginings, e.g. a scale model, an equation, a diagram or a descriptive text in a book, and the world of the model we are imagining about. Both are representations of the target system. I will call the *prop* type of model the *p-model* and the world of the model the *f-model* ('f' for fiction). If we talk about 'the model' of some target system, we talk about the p-model, the f-model and the relation between them as a whole. But they are distinct components of the model and need to be discussed individually.

A p-model is an artificial prop, designed for a certain purpose. By constructing a prop, we can share our imaginative thoughts with one another in the same way as children who play a game in which every stump they encounter is a bear. The stumps are used to coordinate collective imaginative activities. By accepting the rules of the

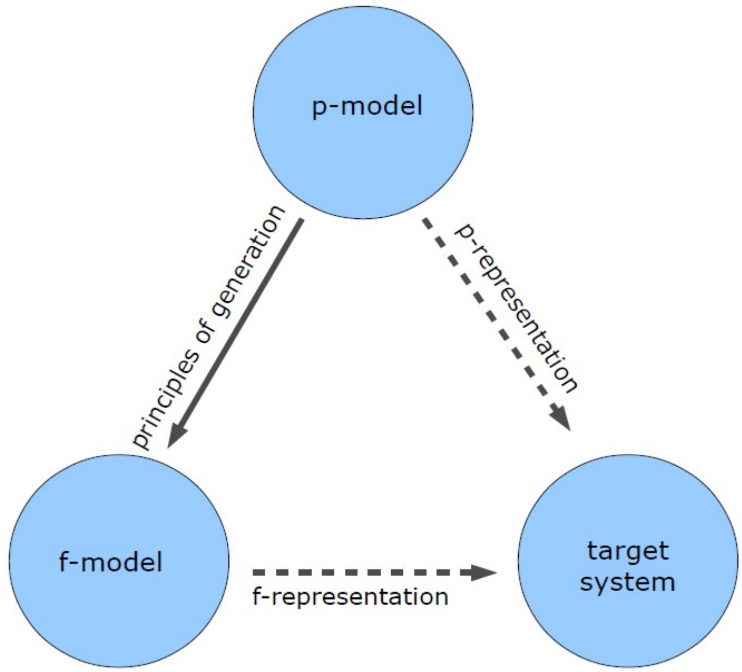

Figure 3: The components of a pretence theory of scientific modelling (1).

game, we can participate in collective games of make-believe. A community of scientists can use the p-model to collectively participate in a game of make-believe. This does not have to take place in a group setting, where everyone is present at the same time. Different runs of the game can be played at different times and different places. Because the p-model and the principles of generation prescribe what is to be imagined, every run of the game will lead to the same f-model, unless of course the game is not played properly.

What is true in a game of make-believe is a fictional truth. Fictional truths do not depend on whether or not they are imagined, it is the prop together with the principles of generation that prescribe what is true in the game. Thus imaginings are constrained, a fictional truth consists in there being a prescription or mandate to imagine something. So the p-model and the appropriate principles of generation determine what is true in the world of the model (see figure 3). The p-model itself gives its world, the f-model, a kind of objectivity, an independence from the scientists who are playing the game. The principles of generation that apply are accepted automatically, they seem natural and do not have to be set up every time a game is played; a feature that is used to direct peoples imaginings when designing a p-model. So when we let others construct a p-model, we can enjoy their talents and insights.

A p-model is a representation of a real world phenomenon, or target system, if there are propositions about that phenomenon which the p-model makes fictional. I will call this kind of representation *p-representation*, as is shown in figure 3. Suppose we have on our desk a spring system with a mass hanging from it. If we pull the mass down, we can see it bouncing up and down in a certain manner. We can model the behavior of the mass and spring with the equation $F = -kx$. This is a p-representation of the system. The p-model (the equation), and the appropriate principles of generation (e.g. Newtons second law), prescribe what is fictional in the model. Suppose that we have a mass that weighs 7 grams, and that when we attach it to the spring it stretches 4.4 cm. We can now find the spring constant *k.* Newton's second law tells us that $F = ma$ and we know that $k = |F|/x$. So it follows that k = 0.007*9.8/0.044 = 1.56 N/m. In other words, the p-model together with the principles of generation determine that it is a fictional truth, that it is true in the f-model, that the spring constant is 1.56 N/m. The model *f-represents* the target system (figure 3), so we conclude that the spring constant of the spring on our desk is 1.56 N/m.

On what basis does the world of the model f-represent the target system? First of all, it does so by virtue of the fact that the p-model is a p-representation of the target system. Moreover, although the p-model in the example above is very different from its target system, the f-model is not. We imagine about the mass hanging from the spring and we imagine, by virtue of the p-model and the principles of generation, that it has a spring constant of 1.56 N/m. So, the f-model is in many respects similar to the target system, whereas the p-model is not.

According to Walton's theory of pretence, representations can be their own objects; they can be reflexive. Walton argues that, for example, a doll not just prescribes to imagine a baby, but to imagine the doll itself to be a baby. Following this line of reasoning, we should then imagine of a scale model of a building that it is the building. But is this true? In the first chapter we argued that representation isn't reflexive. A representation is a representation of something else. The scale model is a p-model that p-represents the actual building and that prescribes imaginings which f-represents the actual building. We can indeed *imagine* that the scale model is the actual building, but nevertheless both the scale model and the imaginings are representations of the actual building. There is no reflexive representation here[22].

What happens if we want to represent something of which we are not sure if it actually exists? Just as it is possible to imagine about a non-actual character such as *Don Quixote*, it is possible to have a model of, for example, the *Higgs particle*. We presume that the *Higgs particle* exists and if it exists, we have an idea of what its properties are. This is enough to create a p-model which in turn generates an f-model. Now suppose it turns out that the *Higgs particle* does not exist, then we have a p-model and an f-model of the particle, but there fails to be a target system. Hence there is no representational relation of either sort.

22 In Walton's example of a game with a doll there is perhaps not a particular baby which the doll represents, but I would argue that it represents some indefinite, but real baby. The imaginings the doll induces are, as Walton argues, about the doll, but they represent the target system. So there is no *reflexive* representation here either.

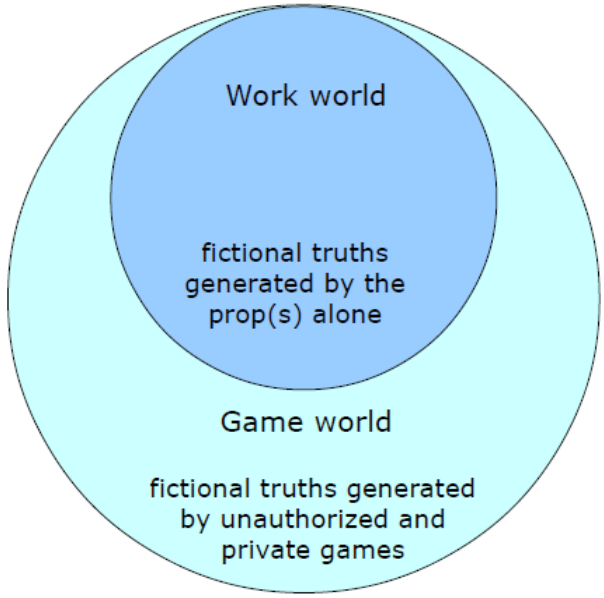

Figure 4: The distinction between work world and game world.

We have discussed earlier that models are often inconsistent, inaccurate and incomplete. This is not a problem for Walton's theory. Fictional worlds can contain inconsistencies, can be inaccurate and are almost always incomplete, seen from the real world that is. A novel can have inconsistencies in its story, and the works of M.C. Escher are often full of impossibilities. This is one of the great benefits of this theory. Note that fictional worlds, seen from within, are not inconsistent, inaccurate and incomplete. They are the way they are, just as the real world is as it is.

Following Walton's theory, we have to distinguish the world of the model, the f-model, from the world of the game that the scientist is playing (figure 4). Remember, an appreciator that is looking at a painting, imagines that *he* sees the characters that are portrayed in it. But he is not part of the painting, so the fictional truth that *he* sees the characters is only part of the world of his private game.

But is this distinction between worlds necessary for our purpose? The scientist that uses the p-model as a prop in a game of make-believe, doesn't imagine himself looking at the p-model like an appreciator of a painting imagines himself looking at what is portrayed. The scientist just imagines about the p-model, about the fictional truths it generates. Only if someone would use the p-model as a prop in an unauthorized game of make-believe, to imagine about it in a way that is not appropriate, then we would need the game world of that someone to account for those imaginings. Consider the mass hanging from the spring again; if the scientist performs the wrong calculations, or applies the wrong laws, his conclusions are part of his game world. So in that sense there is indeed a game world. It is a private domain, where one can do as he pleases, and in which fictional truths are generated due to playing inappropriate games.

In a community of scientists there has to be certainty about what world it is that is under investigation. Private game worlds cannot have a place in a theory which core is playing collective games of make-belief. The world of the model has to be objective and can only be generated from the p-model together with the appropriate principles of generation, otherwise it would be impossible to use a p-model in a group setting and share thoughts and ideas. On this point we have to leave Walton.

In line with the analogies between models systems and Walton's theory of fiction discussed above, we can describe the *pretence theory* in terms of model systems: (1) a p-model has the social function of serving as a prop in a game of make-believe, (2) the p-model prescribes by virtue of the principles of generation what is mandated to imagine and (3) insights that are reached while following the principles of generation, are truths in the world of the model.

3.3 Principles of generation in scientific fictions

Defining the principles of generation that apply in a pretence theory of scientific modelling, is crucial for such a theory. Studying a model is finding out what fictional truths it generates. But *what* fictional truths does it generate? According to what principles?

The generation of fictional truths starts with the designing of a model (the p-model). By designing a model a certain way, the scientist can direct the imaginings of others that are using the model. The model, possibly with an accompanying description, generates the primary fictional truths. The implied (indirectly generated) fictional truths depend on these primary (directly generated) fictional truths. They are, as it were, built on the structure of primary fictional truths. Therefore, designing the model has to be done carefully, because it determines what fictional truths will be generated. Fictional truths that are based on a sloppy design, aren't reliable. On the other hand, if the model is well designed it will generate valuable fictional truths. So, the model and its accompanying description have to form a solid basis of directly generated fictional truths, on which a structure of implied fictional truths can be built.

The question that immediately comes to mind is *which* fictional truths are implied by the primary ones. Roman Frigg (2010b) argues that it is plausible to assume that different disciplines have different rules and therefore we should not expect a single unified answer. We should study what rules apply to what branch of science. In the Newtonian model of the solar system for example, the appropriate laws and principles assumed to hold would be the laws of classical mechanics, the law of gravity, and some general assumptions about physical objects.

According to Arnon Levy (2011), the *relevant* principles of generation specify what participants in the game are to imagine. And Adam Toon argues that "the principles of generation by which models prescribe imaginings will vary from case to case"[23]. He uses an example of a scale model of a bridge. In one case the model is used to carry out structural tests and so it is implied that the real bridge will be built from the same material. But in another case, the scale model is built for a museum display and the principle may not hold. So these three authors all argue that there isn't one general principle that is to be applied in all cases of scientific modelling.

But what about the Reality Principle or the Mutual Belief Principle that Walton discusses in his book? Especially the Reality Principle would seem to be a good candidate to take up this role. After all, science is about finding out how things work and what things are. So it is only natural to use reality as the guiding principle in a pretence theory of scientific modelling. Remember, according to the Reality Principle a fictional truth is legitimately implied if it follows the lines of what would be a legitimate inference in the real world[24].

As the above mentioned scale model example shows, however, the Reality Principle alone is not enough. The Reality Principle on its own cannot distinguish between the case where the model is used for structural tests and the case where it is a model only for display. So what is missing is some form of *intention*, or an understanding of what purpose the model is made for. In many cases this will be implicit knowledge within a group of scientists, but in other cases this should be made explicit in a model description. The Reality Principle together with the implicit or explicit knowledge about the purpose of the model is a serious candidate for the function of *principle of implication* (figure 5).

There are also cases where one p-model can lead to different target systems. We saw for example that the equation $F = -kx$ can be used to model the behaviour of a mass hanging from a spring. But this p-model can also be used to model the behaviour of a simple pendulum with a small amplitude. Again, the p-model and the appropriate principles of generation are not enough. Also in this case we need to have implicit or explicit knowledge about the purpose of the model.

The other candidate for the function of *principle of implication* discussed here, the Mutual Belief Principle, works quite similar in a scientific context. Scientists will make the same inferences using the Mutual Belief Principle as they would if they used the Reality Principle, because what is mutually believed are the same rules and laws as the rules and laws they would use with the Reality Principle to come to conclusions. The difference is that fictional "truths" generated with the Reality Principle could turn out to be false, because of our inadequate knowledge of reality at the time, while fictional truths generated by the Mutual Belief Principle continue to be fictional truths. For the same reasons as with the Reality Principle, the Mutual Belief Principle needs *intention* in order to be suitable as *principle of implication.*

23 Toon (2010b, page 14).
24 Assuming our inference-making capacities are infallible.

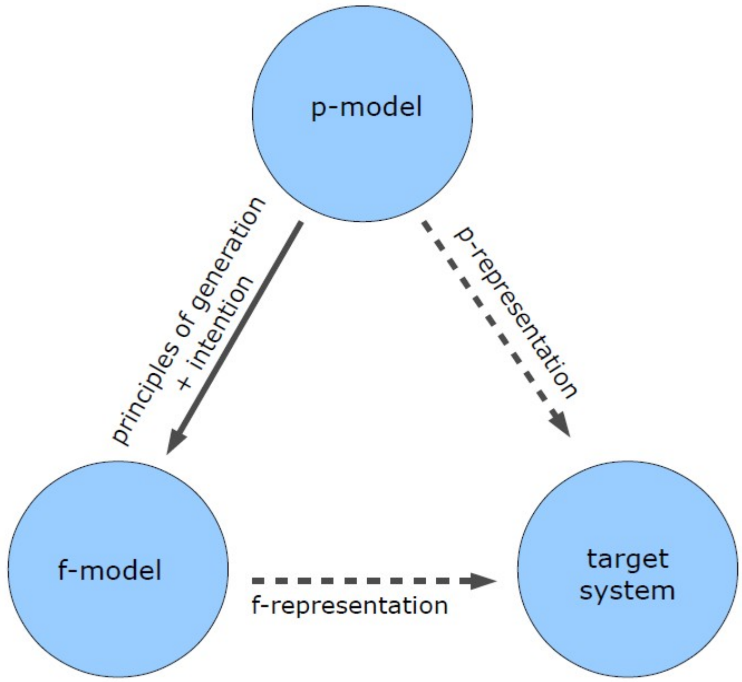

Figure 5: The components of a pretence theory of scientific modelling (2).

Another aspect of scientific modelling that could perhaps shed some light on this matter, is sharing your model with colleagues, for example by getting an article published. It is a common practice that before a scientific article gets published, it is peer reviewed. Again, the model itself (the p-model) and the accompanying description will generate the primary fictional truths. Normally, all the new and controversial stuff will be explicitly stated in the model. In a community of scientists it is to be expected that they share more or less the same set of beliefs about what is implicit knowledge in their field. So, when they review the work of their colleagues, they will comment mainly on what is specifically stated. But, of course, there could be flaws in the model due to false implicit assumptions. A discussion about this will be won on the basis of what is actually mutually believed in the community. If there continues to be controversy, it is obviously something that is not mutually believed and it should be explicitly stated in an updated version of the model.

Which of the two principles of implication is to be preferred depends on a fundamental choice: do we want an account of scientific modelling that produces fictional truths on the basis of reality or on the basis of what we mutually believe in a community? Assuming that we can have wrong ideas about reality, perhaps the Mutual Belief Principle is the preferred option. The consequence is then that if we gain new insights about reality and adjust our beliefs accordingly, we must play a new game of make-believe and generate new fictional truths. The old fictional truths remain fictional truths in the old game, just as it remains a fictional truth that the world is flat in a story written in a time when that was mutually believed.

However, it could be argued that in a scientific context, where the search for truth is so important, there should be no place for a principle that allows for fictional truths to be generated that are not in line with what is actually true. The advantage of the Reality Principle is that is a more objective principle and that a reached conclusion on the basis of inadequate knowledge of reality was never fictionally true and as soon as we discover that, our new insights can lead us to new, better results.

In the next chapter, we will take a look at two examples of models and investigate how the Reality Principle and Mutual Belief Principle behave and what the consequences are if we choose one or the other.

Of course Frigg (2010b), Levy (2011) and Toon (2010b) are right when they argue that different disciplines use different rules or laws to generate fictional truths. Fictional truths generated from a linguistic model for instance, aren't affected by the law of gravity. And an economic law of supply and demand does not have a direct influence on a model of our solar system. But it is not necessary to make explicit what rules or laws of nature should be taken into account for a particular model. The inferences we make from a model are the inferences we would make if the model was real. As long as we do not consider gravity as a factor in the structure of language, we do not take gravity into account when studying a linguistic model. But we do take it into account when studying a model of our solar system.

Perhaps it is easier to decide which rules or laws we should not use with a particular model than to decide which ones we do need to use. But there is no harm done if we do not consider all the rules and laws that are working in a certain target system in its corresponding model system. Science is about discovering step by step how things work, which factors are of influence on the target system and which are not. Results can be improved as soon as we obtain more knowledge.

3.4 Metaphysical Commitment

Walton's theory is anti-realist, i.e., fictional characters, places or events do not exist. The question then arises what that means for a theory of model systems understood in analogy to fiction, what metaphysical commitments do we incur?

It is widely accepted among philosophers that names in ordinary contexts are *directly referential:* the semantic content of a name is its referent, rather than a descriptive sense (Friend, 2007). So if a name lacks a referent, as is the case with fictional entities, the sentence cannot express a complete proposition and is therefore neither true nor false. But at the same time an assertion as 'Don Quixote has a horse named Rocinante' seems to be true when we understand that it is implicitly prefixed by a story operator such as 'according to the fiction' (Friend, 2007)[25].

According to Frigg (2010a) a story operator is a solution for such metafictional propositions. He argues that a claim *p* within a work of fiction (intrafictional) is fictional if the prop together with the principles of generation prescribe *p* to be imagined. Formally: *p* is fictional in work *w* iff the *w*-prop together with the *w*-principles of generation prescribe *p* to be imagined. So the connection between truth and fictionality can be formulated as follows: *p* is fictional in work *w* iff $F_w(p)$ is true (Frigg, 2010a).

If a metafictional claim as 'Don Quixote has a horse named Rocinante' is in fact an ellipse for 'in *The Ingenious Gentleman Don Quixote of La Mancha*, Don Quixote has a horse named Rocinante', we make an assertion of the form 'in work *w*, *p*'. Since *w* is a work of fiction, this assertion is equivalent to '*p* is fictional in work *w*', or $F_w(p)$. Then formally: $F_w(p)$ 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 prescribes *p* to be imagined. From this we can derive that *p* uttered as a metafictional claim is true if and only if *p* is fictional when uttered as an intrafictional claim (Frigg, 2010a). In the context of model systems, we can read *p as* a claim about the model, *w* as the world of the model, the *w*-prop as the p-model and the *w*-principles of generation as the appropriate principles of generation.

A bigger problem arises for Frigg when he has to deal with the fact that there are also statements that cannot be prefixed by a story operator, statements where we make comparisons between different works of fiction or between a fiction and the real world (transfictional); the kind of statements that we want to make in a scientific context, e.g., 'the model aeroplane is sixteen times smaller than the real one', or 'the period of oscillation of the mass in the model is within 10% of the period of the mass in the system'. Frigg acknowledges that these kind of statements seem to pose a problem, but he argues that in a scientific context we do not have to deal with this problem 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."[26]. So, transfictional statements about model systems should be prefixed with a clause that states what the relevant aspects of the comparison are. According to Frigg we now have replaced the problematic comparison with a nonexistent object with an unproblematic comparison of properties.

With Garcia-Carpintero I have concerns about this account: "what is the justification for the claim that the transfictional statements in model-based science *compare features of the model system with features of the target system*?"[27]. First of all, it seems to be a strange move to allow the comparison of features, while denying the existence of the objects that should possess those features. Secondly, it is

25 Being prefixed with a story operator does not give a fictional entity a referent.
26 Frigg (2010a, page 263).
27 Garcia-Carpintero (2010, page 161).

questionable if all transfictional statements make such explicit comparisons that can be justified because they contain prefixes. Godfrey-Smith writes, in an attempt to describe model-based science from the perspective of biology, that models are often not mathematical in nature and "many of the models instead proceed by describing an idealized, schematic causal mechanism, noting how it will and will not behave, and exploring plausible evolutionary paths from one situation to another"[28]. This does not suggest that such models are in any way prefixed (Garcia-Carpintero, 2010).

Toon (2010a) offers an account of scientific modelling that at first seems quite similar to that of Frigg, but differs on one important aspect. Toon argues that our imaginings are about the actual mass hanging from a spring instead of imagining an idealized bouncing spring system which in turn represents the actual system. So where Frigg proposes an indirect view on scientific modelling, Toon proposes a direct account. The p-model, its accompanying description and the relevant principles of generation prescribe us to imagine, about the actual mass, for example that the mass oscillates sinusoidally. Thus, an utterance of the claim 'the mass oscillates sinusoidally' is fictional, without depending on the existence of any object that really does oscillate sinusoidally. Even in cases as 'the point mass oscillates sinusoidally', where we appear to refer to a non-entity, Toon argues that it is the mass on the spring that is being represented and that the model makes it fictional that the mass oscillates sinusoidally, without the reference to some fictional entity.

If we have imaginings about the actual mass on the spring instead of imaginings about a fictional one, comparative statements as 'the period of oscillation of the mass in the model is within 10% of the period of the mass in the system' seem to pose a problem. There is only one mass on a spring. But also in this case Toon doesn't need the help of any fictional or abstract object, "we are simply comparing what our model asks us to imagine with what is true of the system"[29].

He uses Walton's notion of *unofficial* games to explain such comparative statements. The statement invokes an unofficial game in which it is fictional that both the mass and an entity called 'the model mass', that fictionally has all the properties attributed to it by the model, exist[30]. When we now say 'the period of oscillation of the mass in the model is within 10% of the period of the mass in the system', we indicate that pretending in the way that we do is appropriate in this unofficial game, and what we assert is true if and only if it is fictional in the game that we speak truly.

A problem with this account is that it is "intended to apply only to prepared descriptions and theoretical laws scientists formulate when they model some actual system"[31], as Toon himself notices. So models of entities that we now belief not to exist

28 Godfrey-Smith (2006, page 732).
29 Toon (2010a, page 313).
30 Despite his intentions to do without, here Toon resorts to some abstract entity called 'the model mass'.
31 Toon (2010a, page 308).

or that could turn out not to exist, e.g. a model of the Higgs boson, are not covered with this account.

Ronald Giere (2010) agrees that ontologically we should think of abstract scientific models as works of imagination, that the imaginative processes at work when we produce a work of fiction are similar as those when producing a scientific model and that they have no ontological status beyond being imaginations. He thinks it unlikely that the ontological status of abstract models in science will become any clearer and he concludes by saying: "It is enough for the purpose of understanding scientific practice to know that constructing abstract objects is a natural capacity of creatures possessing language"[32].

According to Saul Kripke, "the fictional character can be regarded as an abstract entity which exists in virtue of the activities of human beings, in the same way that nations are abstract entities which exist in virtue of the activities of human beings and their interrelations"[33]. Thus, a fictional character or entity exists if human beings created them, for example by writing a book or making a painting. If we say 'Don Quixote has a horse named Rocinante', we are referring to a person that really does exist and really has a horse named Rocinante. Only if we speak outside the story we can say that no such person as Don Quixote exists (Kripke, 2011).

Following Kripke, I suggest to introduce the notion of *fictional existence*[34], alongside that of *fictional truth*. Within the world of the model, things really do exist. So when a prop, together with the principles of generation, prescribe us to imagine something, what we imagine is fictionally existent and it possesses *fictional properties*. As Kripke notices, nations are abstract entities that exist in virtue of the activities of human beings. We can compare properties of those abstract entities, e.g. the Netherlands has a total area of 41,526 square kilometres, Germany has a total area of 357,021 square kilometres, so Germany is 8.598 times as big as the Netherlands. The comparison of the period of oscillation of the actual mass with the period of oscillation of the fictional mass happens in the same way; we compare period T_{actual} with the fictionally existent period T_{model} and determine whether or not T_{actual} is within 10% of T_{model}.

3.5 Summary

Let's summarize the theory of scientific modelling as developed in this chapter, before we go on to the next chapter to take a look at two models on which we will test this theory.

32 Giere (2010, page 279).
33 Kripke (2011, page 63).
34 Here I need to thank Janneke van Lith, who brought the term *fictional existence* up during one of our meetings.

Scientists design models of real world phenomena. These models can be scale models, mathematical models, physical models, graphs, descriptive texts, and so on. We call these models *p-models*. A p-model is a representation of a real world phenomenon if there are propositions about that phenomenon which the p-model makes fictional, we call this *p-representation*. A p-model functions as a prop in a game of make-believe; by virtue of the appropriate *principles of generation* and implicit or explicit knowledge about the purpose of the model, the p-model generates *fictional truths*, truths in the world of the model (the *f-model*).

There are *primary* fictional truths and *implied* fictional truths. The primary fictional truths are those that are directly generated by the p-model, by everything that is made explicit in the p-model. The primary fictional truths are the structure on which the implied fictional truths depend. Which implied fictional truths are generated also depends on what *principles of implication* are in effect. Two such principles are the *Reality Principle* and the *Mutual Belief Principle*. According to the Reality Principle a fictional truth is legitimately implied if it follows the lines of what would be a legitimate inference in the real world. According to the Mutual Belief Principle a fictional truth is legitimately implied if it follows the lines of what is mutually believed in a society to be a legitimate inference.

What the p-model, together with the principles of generation, prescribes us to imagine, is *fictionally existent* and has *fictional properties*. Within the game of make-believe we can compare properties of these abstract entities with properties of the target system.

The question remains what principle of implication is the best candidate as principle for our theory of scientific modelling. In the next chapter I will try to get an answer to that question by testing them both on two examples in order to see what the consequences of both of them are.

> *"To represent something is to prescribe imaginings about it; and engaging in imaginings about something is a good way of deepen one's understanding of it"*
>
> \- Kendall Walton

4. For example...

The aim of this chapter is to discuss two scientific models to test our theory of pretence. The first model is directly related to A.I., it comes from the field of neuroscience and is a *motor control* model. The second model is a geocentric model of Ptolemy.

4.1 Motor control model

Wolpert et al. (1995) carried out an experiment in which participants had to make arm movements in the dark, under three different experimental conditions: with assistive force, resistive force and without an external force on their arm. By asking the participants to localize visually the position of their hand after the movement, they assessed the participants' internal estimate of the location of their hand. The aim of the experiment was to find a model that would explain the information-processing mechanisms that the nervous system uses to obtain an estimate of the current state (position, velocity) of the hand.

Beforehand they distinguished three possibilities. The system could use *sensory inflow* (information available from one's own perception of the orientation of their limbs in space (proprioception)), *integrated motor outflow* (the motor commands sent to the arm), or a combination of these two sources by using a forward model[35]. Wolpert et al. found that the participants consistently, for all experimental conditions, overestimated the distance their hands moved. The bias increased with the assistive force and decreased with the resistive force, but the variance was unaffected. Moreover, the bias

35 A forward model takes as input the initial state of the body and a motor command and has as output a predicted position of the body.

showed two distinct phases as a function of movement duration. As the duration of the arm movement increased from .5 second (shortest movement duration) to 1 second, the magnitude of the overestimation increased. After that, the magnitude of overestimation decreased.

Wolpert et al. reasoned that neither *sensory inflow* nor *integrated motor outflow* alone could account for these findings, but the combination of these two could. They established this conclusion with a Kalman filter model. To be able to fully understand their model, I will first discuss Rick Grush' *emulation theory of representation* (Grush, 2004), in order to clarify the general idea of Kalman filters. Grush' theory is an information-processing strategy that is essentially identical to the model proposed by Wolpert et al. (1995). Grush however argues that this framework not only applies to *motor control*, but also to *motor imagery*, *visual imagery*, *visual perception*, and perhaps to other psychological capacities such as *reasoning*, *theory of mind*, and *language* as well.

4.1.1 Kalman filters

Figure 6 shows a standard version of a Kalman filtering scheme (within the outer dotted-line box), with the addition of a feedback-loop. The general idea is that when one system (e.g. a brain, a ship's crew) interacts with another system (a body, a ship), the state of the second system is not entirely predictable. How the body (or ship) functions in general is known to the brain (or the ship's crew), but *process noise* (internal or external disturbance of the limbs, unpredictable currents), and *sensor noise* prevent to give a perfect estimate of the state of the system. A model of the process – maintained by specialized circuits in the brain or a navigation team on the ship – provides a prediction of what the state of the system will be and uses this prediction in combination with sensor information to maintain a good estimate of the actual state of the system (Grush, 2004).

Let's consider the scheme in figure 6 in terms of a (part of the) brain which has the goal of moving an arm in order to reach an object. This goal triggers the brain (*controller*) to send out a control signal. The upper dashed-line box represents the plant, in this case the body, or the arm in particular. The *process,* the arm and its location, is at any time a function of its location at the previous time, plus the *driving force* (force by which the arm is moved), plus the *process noise* (unpredictable external influence). Now the *real signal I(t)* is a measurement of states of the process. Unpredictable noise (imperfect sensors) in this measurement process is represented by the *sensor noise* and all this together results in the actual output, the *observed signal S(t)*. The task is now to filter the *sensor noise* from the observed signal to determine the real, noise-free signal *I(t)* (Grush, 2004).

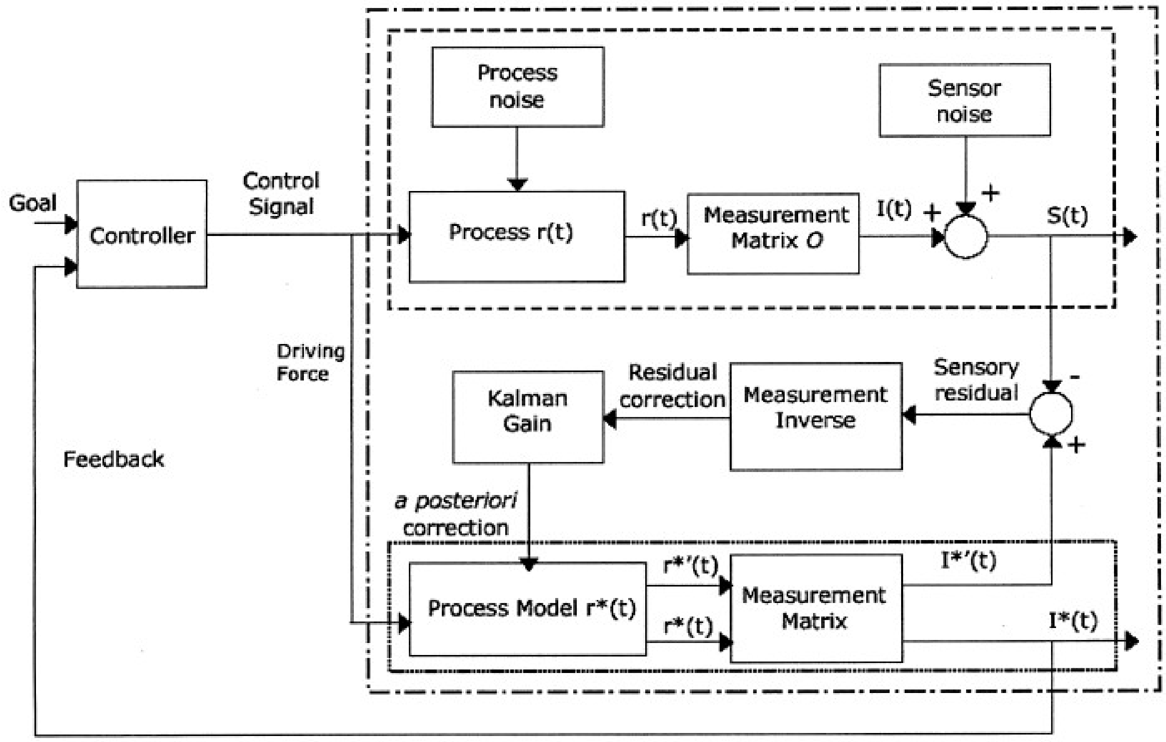

Figure 6: A control scheme consisting of a Kalman filter additioned with a feedback-loop (Grush, 2004).

In order to obtain an optimal estimate of the real process' state, the controller sends an efference copy[36] of the control signal to the process model. This signal undergoes the same measurement update as happens in the actual plant, only without the sources of noise. The result is an *a priori signal estimate I*'(t)*. This *a priori signal estimate* is compared with the *observed signal S(t)*, of which the difference is called the *sensory residual*. By applying the inverse of the measurement update to this signal, the *residual correction* is found. This is done to find how much the *a priori estimate* would have to be changed to eliminate the residual altogether. The *Kalman Gain* then is the determination how much of this correction should actually be applied, because not all of the *sensory residual* is a result of the *a priori signal estimate*, but also from the *sensor noise*. The *sensor noise* could in fact be the major factor that determines the residual, while *I*'(t)* is highly accurate (Grush, 2004).

The *Kalman Gain* is found by comparing the expectation of the signal to what the signal actually is and on the basis of that difference an adjustment is made of the estimation of the state the real process is in. If the sensors are expected to be unreliable, the *a priori expectation* is given more weight, but as sensor information is

36 An *efference copy* is an internal copy created with a motor command of its predicted movement and its resulting sensations.

expected to be reliable, then the expectation is weight less. The result is an *a posteriori estimate* of the process state, which is a function of both the initial expectation and the actual observation. This estimate is run through the measurement update in order to get the *final estimate* $I^*(t)$ of the noise-free signal $I(t)$. The *final estimate* is sent back to the *controller* (brain) to function as input for the next cycle (Grush, 2004).

Thus, the location of the arm with respect to the goal state is estimated by the combination of the actual, but noisy observation of the body itself and the prediction of the location of the arm based on the process model.

4.1.2 Kalman filter model for motor control task

Figure 7 shows the Kalman filter model proposed by Wolpert et al. (1995) to account for the findings of their motor control experiment. The model in figure 7 is essentially identical to the earlier described Kalman filtering scheme in figure 6 (Grush, 2004).

The upper half of the model consists of a forward model of the combination of the *sensory inflow* and the *integrated motor outflow* that Wolpert et al. described. So it uses the current state estimate and the motor command to predict the next state. The lower half of the model consists of a model of the *sensory output process* that predicts the sensory feedback signal. The difference between that prediction and the *actual sensory feedback* gives the *sensory error* that is used to correct the state estimate resulting from the forward model. The *Kalman Gain* modulates the relative weighting of the internal simulation and the sensory correction process to find the optimal state estimate (Wolpert et al., 1995).

Wolpert et al. found that the magnitude by which the participants overestimated the distance that their hand moved, increased for movement durations between .5 and 1 second. For movements that took longer than 1 second, the magnitude of the overestimation decreased. Overall, this bias increased in case of the assistive force and decreased with the resistive force. This can be explained as follows. At the beginning of

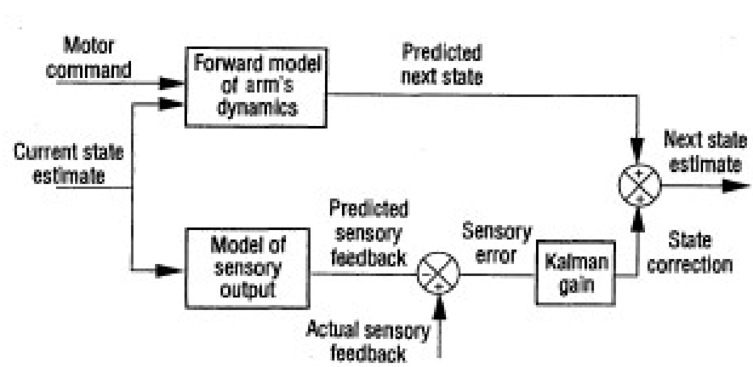

Figure 7: Kalman filter model (Wolpert et al., 1995).

the task, the state estimate is assumed to be accurate. Therefore the Kalman filter weights the contribution of the forward model heavily. Because of overestimation of the force performed on the arm in the forward model, the magnitude of the initial overestimation of the distance that the hand traveled is relatively big. The assistive force increases this overestimation, the resistive force decreases it. During the later stages of movement, the current state estimate is less accurate and thus the relative weighting in the Kalman filter shifts toward the *sensory feedback* and then remains approximately constant. Hence the decreasing magnitude of the overestimation after approximately 1 second. So in case of the assistive force condition, the overestimation is already larger at the moment the relative weighting in the Kalman filter shifts, and smaller in case of the resistive force condition (Wolpert et al., 1995).

4.1.3 Pretence theory applied to the motor control model

The first task is to identify what the model and its corresponding target system are. The aim of the experiment was to find a model that would explain the information-processing mechanisms used by the nervous system to obtain an estimate of the current state of the hand. So the target system is the relevant part of the human nervous system, from the brain to the neurons in the hand. The model system is the Kalman filter model shown in figure 7 together with the accompanying description provided by Daniel Wolpert and his colleagues (Wolpert et al, 1995).

The Kalman filter model, including its description, is the prop (p-model) that induces imaginings in a game of make-believe. It is prescribed to imagine the relevant parts of the nervous system, the structures, cells and molecules, communicating with each other following the causal connections as given in the p-model. The processes that are happening within the individual cells, the behaviour of the molecules involved, are among the things that are implied by the principles of implication.

If we apply the Reality Principle as principle of implication, those processes happen in the fictional world as they happen in reality, and we imagine about it to the best of our knowledge. In case we gain new, conflicting insights about, for example, the processes going on inside the individual neurons, that means our imaginings about them were false, what we pretended about it was never fictionally true. With new insights we can update our imaginings and, in some cases, gain new insights about the target system. In this case the exact workings of the individual cells do not seem to be of much influence on the whole process.

If we would use the Mutual Belief Principle as principle of implication, the initial imaginings that the model induces are the same as in case of the Reality Principle. The difference is that what is imagined is fictionally true, even if it turns out that new insights are conflicting with our knowledge at the time of the imagining. The solution would be to "reset" the original imaginings and do it over according to the new insights.

The pretence induced by the model enables us to use the compactly described model system to reason about the process and highlight particular aspects of the process that have explanatory power (Levy, 2011). The designers of the p-model direct our imaginings in such a way that we can access their ideas and insights. Directly generated fictional truths depend on the design of the p-model, implied fictional truths on the principles of implication. The directly generated fictional truths are mainly about the information-processing mechanisms and causal links of which Wolpert et al. think are present in the target system.

Who are qualified to take part in this collective game? In general, everyone who is willing to accept the rules of the game is. But with accepting the rules comes the necessity of understanding them. Not only the explicit rules, which are directly prescribed by the p-model, need to be properly understood, but also the implicit rules. In this case one must have, to a certain degree at least, a familiarity with the field of neuroscience and an understanding of the workings of the Kalman filter. So the model is accessible to the community of scientists in the field of neuroscience, to some scientists in closely related fields and probably also to neuroscience students. Depending on the complexity of the model, or on the degree of specific knowledge of the field that is needed to understand the rules, certain groups of people will or will not be able to take part in the game.

Which principle of implication is to prefer in the case of this motor control model? I think that the Reality Principle gives the most plausible explanation of what happens. All the implicit processes in the individual cells for example, happen in the model just as they actually happen. In this case, the precise workings of the individual cells do not have a direct influence on the model, but indirectly they do. If we gain new insights about the inner workings of cells, it is as much a discovery about the real world as it is about the world of the model, and we adjust our understanding of the world of the model accordingly. As long as our new insights do not conflict with the fictional truths the model generates directly, we can keep using the model.

The Mutual Belief Principle does not seem to have much advantages in the case of this model. The inner workings of the cells would be as is mutually believed at the time the model was created, and if our understanding about cells changes what is fictionally true in the model isn't in line any more with our beliefs, so we need to update our imaginings. Moreover, it just seems more natural in a context of science, to have reality as the generating principle.

However, the actual consequences of a choice between the Reality Principle and the Mutual Belief Principle are hard to make visible with an example of a fairly recent model. As indicated earlier, at the moment the model is designed, the scientist of course follows what he thinks comes closest to reality. What is mutually believed is what is believed to be according to reality. So, it does not really become apparent what the differences are until the moment a model is superseded. In an attempt to show that, I will continue with a canonical example, namely that of the geocentric model.

4.2 The geocentric model

The geocentric model is the theory that places the Earth at the centre of the universe, with all other heavenly bodies orbiting around it. It is not hard to imagine why such a theory is so appealing. The Earth does not seem to move, it feels solid and stable and all other planets and stars appear to revolve around it; the sun comes up every morning and sets every evening. Although the geocentric model was present in ancient Greek astronomy from as early as the 6^{th} century BC, with Plato and Aristotle as two of the most influential philosophers to hold these views, it was Ptolemy in the 2^{nd} century AD who devised a model, described in *The Almagest*, a series of 13 books, that would remain dominant for almost 1500 years.

4.2.1 The Ptolemaic system

One of the biggest challenges for the geocentric model was the observation, from a terrestrial perspective, that the outer planets stop, move backwards in a *retrograde* motion for a while, before continuing to move forwards again. Moreover, there seemed to be changes in the distances of the planets from Earth.

In the Ptolemaic system, the planets move in a small circle called an *epicycle*, which in turn moves along a larger circle called a *deferent* (figure 8). The centre point of the deferent is not the Earth, but a point halfway between the Earth and another point called the *equant*. By placing the earth not in the centre of the deferent, but slightly of centre, Ptolemy could account for the Sun's varying motion through the zodiac. He combined this eccentricity with the epicyclic model to explain the way the planets appeared to move. Because one half of an epicycle runs counter to the general motion of the deferent path, the combined motion will sometimes appear to be slower or even to be reversed (in retrograde motion). To a hypothetical observer placed at the equant point, the centre of the epicycle appears to move at a constant speed.

This system accounted fairly well for planetary motion, which explains its success for so long. It took until 1543, with the publication of Copernicus' *De revolutionibus orbium coelestium*, for the geocentric model to be seriously challenged.

In the Ptolemaic system, the order of spheres from Earth outward is Moon, Mercury, Venus, Sun, Mars, Jupiter, Saturn and then the stars.

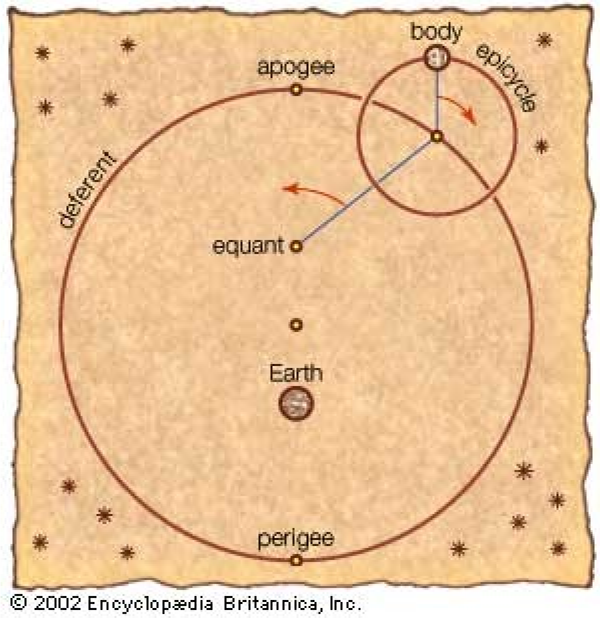

Figure 8: The basic elements of the Ptolemaic system (Encyclopedia Britannica, 2012).

4.2.2 Pretence theory applied to the geocentric model

The aim of Ptolemy was to find a model of the universe, that could explain the phenomena that they observed. So, the target system is the universe, and the model system the geocentric system described in *The Almagest*. The book itself is the prop that, together with the principles of generation, induces imaginings in a game of make-believe. It is prescribed to imagine the earth at the centre of the universe, with all planets, stars, the sun and the moon orbiting around it in the manner described.

The interesting part of this example is to find something that was implicit knowledge in Ptolemy's time, but of which we now know that it was false. Of course we know now that the earth orbits the sun and that the earth is the third rock from the sun instead of the centre of the universe. But all that is explicitly stated in the model, those are directly generated fictional truths, so that can't be used for a discussion about the Reality Principle versus the Mutual Belief Principle.

I think that it is safe to assume that between the 2nd and 16th century AD no one thought of the possibility that the universe expands. So if we accept that assumption, the mutual belief was that the universe is static, possibly infinite, but not expanding nor contracting. This implies that, were we to accept the Mutual Belief Principle, it is fictional in the model that all the stars are at a fixed distance from the Earth. The Reality Principle, however, would prescribe to imagine that the distance between Earth and the stars increases over time.

This example clearly shows the consequences of the choice between the Reality

Principle and the Mutual Belief Principle. Although the implications of the Reality Principle (that the distance between Earth and the stars increases over time) are a little bit more in line with what we now belief to be true, they violate the ideas and intentions of Ptolemy. If we argue that in Ptolemy's model, because of the Reality Principle, it is fictionally true that the universe expands, then we don't do justice to his model, his beliefs and ideas.

The Mutual Belief Principle is better able to preserve the model the way it was intended by the creator. It enables us to see the model in its context, in this case the beliefs and insights of the society Ptolemy lived in in the 2nd century AD, whereas the Reality Principle brutally rips it out of that context. Therefore, in contrast with the *motor control model*, I would argue in this case in favour of the Mutual Belief Principle.

5. Discussion

The aim of this thesis was to formulate a theory scientific modelling along the lines of Kendall Walton's *pretence theory*, to discuss the benefits of a fiction view on models, and to see if and where difficulties lie for this approach. In particular, the precise research question to be answered with this thesis was:

Is Kendall Walton's *pretence theory* viable as a framework from which to formulate a theory of scientific modelling?

Walton summarizes his theory as follows:

"Representations, I have said, are things possessing the social function of serving as props in games of make-believe, although they also prompt imaginings and are sometimes objects of them as well. A prop is something which, by virtue of conditional principles of generation, mandates imaginings. Propositions whose imaginings are mandated are fictional, and the fact that a given proposition is fictional is a fictional truth. Fictional worlds are associated with collections of fictional truths; what is fictional is fictional in a given world[37] – the world of a game of make-believe, for example, or that of a representational work of art. This, in brief outline, is the skeleton of my theory."[38]

37 More precise nomenclature: "[...] what is fictional is fictional *with respect to* a given world [...]."

38 Walton (1990, page 69).

Following this theory, we can regard scientific models as representations that serve as props in games of make-believe. The model and the appropriate principles of generation prescribe what is true in the world of the model. By carefully constructing a p-model, scientists can direct our imaginings and share their insights.

This view on scientific modelling has some great benefits. First of all, it offers an intuitive way of thinking about models. As noted earlier, we do not think of the Newtonian model system of Sun and Earth as something purely mathematical or structural, but as, if it were real, a system that consists of two spherical bodies with a certain mass, colour and density and other concrete properties. Moreover, it provides an explanation for the fact that models are inaccurate, inconsistent and incomplete. With this fiction view on models, we are able to deal with complexity issues by making simplifications, we can assume that the spherical bodies in the Newtonian model system are perfect, homogeneous spheres, or that the mass on the spring is a point mass which is subject to a uniform gravitational field and a linear restoring force exerted by a massless frictionless spring attached to a rigid surface.

Secondly, it provides us with an explanation of how different sorts of models can be representations of a target system. A scale model of a building for example, has the same target system as the drawings of the architect of that building. They both prescribe imaginings about the building and could in principle prescribe the exact same imaginings, perhaps by virtue of an accompanying description. Two models are identical if and only if the worlds of the two models are identical (Frigg, 2010a).

A situation which we did not discuss yet is what happens if a result is reached unexpectedly, or even by mistake. A famous example is the error Satyendra Nath Bose made while presenting a lecture on the theory of radiation. He made the error in applying the theory while trying to show his students that the contemporary theory was inadequate. It predicted results which were not in accordance with experimental results. His error unexpectedly gave a prediction that did agree with the experiment. His mistake was simple, similar to arguing that flipping two fair coins produces two heads one third of the time. Bose's mistake produced accurate results because photons are indistinguishable from each other; any two photons having equal energy cannot be treated as two distinct identifiable photons. By analogy, if the two coins would behave as photons, the probability of producing two heads would indeed be one third. Bose's error eventually led to the *Bose-Einstein statistics*.

How does this example fit in the fiction view of scientific modelling discussed in this thesis? In short, we have a model of which it is believed that it is inadequate, but it unexpectedly turns out to be adequate. The p-model together with the principles of generation prescribe what is to be imagined. But there is no guarantee that one's imaginings are always in accordance with what is prescribed to imagine. In that case, an unauthorized game is played, a game that is not appropriate to play with the p-model. Bose discovered by coincidence that the way the game was played before was inappropriate; he discovered the appropriate rules.

The biggest challenge for this theory of scientific modelling lies with the fictional truths that are implied. On what grounds can we formulate a principle of implication? In chapter three we saw that Frigg, Levy and Toon argue that different disciplines use different laws and rules to generate fictional truths, that they vary from case to case. This seems unsatisfying and unnecessary. Walton offers two possibilities: the Reality Principle and the Mutual Belief Principle. We have seen in chapter four that both principles have advantages and drawbacks.

The Reality Principle is objective and doesn't generate fictional truths if it is not in line with reality. That seems natural in a scientific context. With the Reality Principle, we could use a model as a search tool, we can explore the world of the model and discover things we didn't know of or of which we were wrong. But what happens if, for example, it turns out the the *Higgs particle* does not exist? We have a model of it, but the Reality Principle was in that case never able to generate implied fictional truths about it. This seems to be a big problem for the Reality Principle.

The Mutual Belief Principle is better able to preserve the intentions of the creator of the model. If a model is superseded, it seems wrong to claim that all implied fictional truths are according to reality, even if that contradicts with the model. Moreover, we are not omniscient; we make inferences on the basis of what we belief to be true. That is perhaps the most important argument to make a choice in favour of the Mutual Belief Principle.

Another controversial topic is regarding the ontology of fictional entities; the realist versus anti-realist debate. Walton's theory is anti-realist, fictional entities do not exist. The question is then, how can we refer to them? According to Kripke (2011), fictional entities do exist. They are abstract entities brought into existence by human beings, for example because a book is written or a painting is made. By introducing the notion of fictional existence, we should be able to make comparisons between those abstract entities and the real world.

Of course I am aware of the fact that the realist versus anti-realist debate is a whole topic on its own, which deserves a lot more attention than I was able to provide in this thesis. I have tried to give an overview of the difficulties that one encounters and an answer to the question of what metaphysical commitments we incur if we formulate a theory of scientific modelling in analogy to a theory of fiction. This could perhaps be a topic of a next thesis.

To conclude with an answer to the main question of this thesis:

Kendall Walton's *pretence theory* indeed offers a very elegant framework from which to formulate a theory of scientific modelling. It provides us with an intuitive notion of what models are and how we use them, and it clarifies how we can use different models for one target system. At the same time it has to be said that there is still work to be done regarding the principles of implication and the ontological status of fictional entities.


UGH...
WHAT HAPPENED?
WHERE AM I?
HELP!
SOMEONE
FWOOOOOOOSH
HUH. LOOKS LIKE PHYSICS PROFESSORS DON'T LIKE WORKING IN FRICTIONLESS VACUUMS AFTER ALL.
THEY'RE SUCH LIARS.

xkcd (2009)

6. Bibliography

Bailer-Jones, D.M. (2003). When scientific models represent. *International Studies in the Philosophy of Science, 17*, 59-74.

Encyclopedia Britannica (2012). Ptolemaic System. http://www.britannica.com/EBchecked/topic/482079/Ptolemaic-system (last visited: 03-05-2012).

Friend, S. (2007). Fictional Characters. *Philosophy Compass, 2, 141-156.*

Frigg, R. (2006). Scientific representation and the semantic view of theories. *Theoria, 55*, 49-65.

Frigg, R. (2010a). Models and fiction. *Synthese 172,* 251-268.

Frigg, R. (2010b). Fiction in Science. In Woods (ed.), *Fictions and Models.* Munich: Philosophia.

Garcia-Carpintero, M. (2010). Fictional Entities, Theoretical Models and Figurative Truth. In Frigg, Hunter (eds.), *Beyond Mimesis and Convention, Boston Studies in the Philosophy of Science, 262*, 139-168.

Giere, R.N. (2004). How models are used to represent reality. *Philosophy of Science, 71,* 742-752.

Giere, R.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.

Godfrey-Smith, P. (2009). Models and fictions in science. *Philosophical studies, 143*, 101-116.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain sciences, 27*, 377-442.

Kripke, S.A. (2011). Vacuous Names and Fictional Entities. In Kripke (ed.), *Philosophical Troubles: Collected Papers, Volume I.* Oxford Univerity Press, 52-74.

Levy, A. (2011). Information in Biology: A Fictionalist Account. *Noûs, 45,* 640-657.

Suarez, M. (2003). Scientific Representation: against similarity and isomorphism. *International Studies in the Philosophy of Science, 17,* 225-244.

Thomson-Jones, M. (2010). Missing systems and the face value practice. *Synthese, 172,* 283-299.

Toon, A. (2010a). The ontology of theoretical modelling: models as make-believe. *Synthese 172,* 301-315.

Toon, A. (2010b). Models as Make-Believe. In Frigg, Hunter (eds.), *Beyond Mimesis and Convention, Boston Studies in the Philosophy of Science, 262,* 71-96.

Walton, K.L. (1990). *Mimesis as make-believe: on the foundations of the representational arts.* Cambridge/MA: Harvard University Press.

Weisberg, M. (2007). Who is a modeler? *British Journal for Philosophy of Science, 58,* 207-233.

Wolpert, D.M., Ghahramani, Z., Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science, 269,* 1880-1882.

xkcd (2009). Experiment. http://xkcd.com/669/ (last visited: 09-05-2012).