L&PS – Logic & Philosophy of Science
Vol. XII, No. 1, 2014, pp. 41-69

On the Relation Between Models and Hypotheses and the Role of Heuristic Hypotheses in the Construction of Scientific Models

Tjerk Gauderis
Center for Logic and Philosophy of Science
Ghent University
e-mail: tjerk.gauderis@ugent.be

ABSTRACT. In our understanding of model-based scientific practice, it has become unclear what the role of hypotheses is. Many take models and hypotheses to be more or less on the same footing; others take hypotheses to be claims about the intended representational features of models; some have even argued against the use of hypotheses in model-based science. In this paper, I argue that the first and third of these positions are untenable, while the second position applies only to a subclass of the many hypotheses actually employed in model-based scientific practice, which I call fully interpretable hypotheses. Next, I show, based on some case studies from astronomy, that many scientific hypotheses are in fact of a different type, which I call heuristic hypotheses. Therefore, I argue for a fourth position which complements the second position to provide an account of the role of these two kinds of hypotheses in model-based scientific practice.

KEYWORDS: Models, Hypotheses, Scientific Practice, Model-Based Reasoning, Heuristic Reasoning.

Published by L&PS – Logic and Philosophy of Science
http://www2.units.it/~episteme – ISSN: 1826-1043

Imagination creates events.
(Giovanni Francesco Sagredo,
letter to Galileo, 1612)

1. Introduction

As a result of a shift in focus in the philosophy of science from dealing largely with issues of scientific confirmation towards studying actual scientific practices and the questions they invoke, philosophical interest in the use of *models in science* has steadily increased in recent decades. Although early interest was mostly fueled by adherents of the so-called *semantic and structuralist views of theories*,[1] who tried to tailor their formal analyses towards the type of models actually used by scientists, it is now recognized that the structural set-theoretical meaning of models is best not equivocated with the actual practices of *model-based science*,[2] which elicit many ontological and epistemological questions in their own right. The study of these questions in relation to many actual scientific cases has led many, nowadays, to appreciate that much of science can be adequately described as model-based science, which should not be seen so much as a division between the various disciplines, but rather as a strategy that any discipline can employ to address theoretical scientific research (Godfrey-Smith 2006). The construction, manipulation and refinement of models are also now generally considered to be key scientific practices (Frigg and Hartmann 2012).

The recognition of the use of models in science has elicited a substantial amount of research to clarify the relation between this rather new addition to the jargon of scientific methodology and older inhabitants of this conceptual

[1] Until the 1970s, the received view of theories (also called the *syntactic view*) maintained the Euclidean or Aristotelian ideal of a theory as a set of axioms and a suitable logic to infer all true sentences in an ideal scientific language, supplemented with a set of correspondence rules to link theoretical terms to empirical observations. The heavy language-dependency of this view has led various scholars to develop the so-called *semantic view* of theories, in which theories are equated with a class of models, abstract mathematical structures for which the theory was true (Suppes 1960; Suppe 1977, 1989; Van Fraassen 1980). A related, *structuralist view* of scientific theories was developed by, among others, Balzer, Moulines and Sneed (1987). For a recent paper incorporating these structuralist ideas, see Leuridan (2013).

[2] The two main arguments for this distinction are that many models from actual scientific practice cannot be accommodated within the set-theoretical view of models (Downes 1992) and that, while the semantic view aims to analyze all of science in terms of models, not all actual scientific practice relies on the manipulation of models (Godfrey-Smith 2006, Weisberg 2007).

jungle, such as the relation between models and theories (e.g. the semantic view, Giere 1988), models and discovery (Redhead 1980, Morrison and Morgan 1999), models and laws of nature (Cartwright 1983, Giere 1999a) and models and data (Suppes 1962, Harris 2003). Yet, no substantial attention has been paid so far to the relation between models and hypotheses in science. The main reason why this relation has been left unattended may have to do with the fact that models and hypotheses are generally considered to belong to the jargon of two mutually exclusive conceptions of the scientific method, i.e. the inductive (model-based) view and the hypothetico-deductive view.[3]

In this paper, I investigate how hypotheses and models relate in actual model-based scientific practice, show that both are necessary concepts in understanding this practice, and show that they are mutually supportive. Apart from touching upon recent debates in the literature on models, such as those concerning the nature of their representational function and their construction, this research will reinstate a modernized concept of a scientific hypothesis, in line with model-based scientific practice, by shrugging off some of the unrealistic intuitions with which it has been burdened by the old Popperian hypothetico-deductive view.

After delineating my precise usage of the main concepts of this paper (Section 2), I will identify four stances on the relation between hypotheses and models by examining the scattered remarks that have been made in the literature and consider what objections might threaten these stances (Section 3). Then, I will look into actual scientific practice and present three case studies to expose the nature of the interplay between models and hypotheses (Section 4). This will allow me to develop my own account of how hypotheses and their role should be understood in the context of model-based science (Sections 5 and 6).

2. Some Conceptual Issues

Before I present the main arguments of this paper, some preliminaries about its scope and topic are in order. More specifically, as 'model' and certainly 'hypothesis' are often used as umbrella terms and as their meaning is often thought to be more or less self-evident, I need to specify more precisely the kind of hypotheses and models this paper will deal with. Unavoidably, this is a trade-off between catching as much as possible of the actual usage of these concepts

[3] This position is advanced, for instance, in the article of Glass and Hall (2008) that I discuss in Section 3.3.

in scientific practice and defining sufficiently coherent concepts to allow for analysis.

> **Scientific hypotheses**. I take *scientific hypotheses* to be (1) statements (2) about the empirical world (3) that have an unknown or underdetermined truth status and (4) are advanced as a tentative answer to a particular research question.

Let me expand on each part of this characterization. First, scientific hypotheses are linguistic statements or propositions, by virtue of which it always makes sense to talk about their truth status.

Second, this paper focuses only on hypotheses that make a reference to the empirical world. This excludes, along with mathematical conjectures and metaphysical claims, also hypotheses that refer exclusively to parts of and relations within a particular model. Studying the internal properties of scientific models is an important aspect of theoretical science, but conjectures of this kind are generally not what scientists refer to with the notion 'scientific hypothesis'.[4]

Third, although it makes sense to speak (typically in retrospect) of confirmed hypotheses, it is assumed that hypotheses are not known to be true. Yet this does not exclude that scientists can have a firm and even justified belief in them, certainly in later stages of research. Also, I do not assume that hypotheses are fully determined or have an unambiguous reference. As the case studies in this paper show, many actual hypotheses in early stages of research unavoidably have ambiguous or vague references. It is only afterwards, when the conceptual apparatus, requisite models and governing conditions have been developed in subsequent stages of research, that the intended hypothesis can be formulated unambiguously.

Finally, scientific hypotheses are not mere conjectural statements; they are advanced in an attempt to answer particular research questions. In other words, they are *truth-purposive*. Scientists advance them with the purpose of finding the answer to a research question by trying to determine the suggested hypotheses' truth value, even if they know that any particular hypothesis can be rejected or refined later on. Importantly, it is not required that hypotheses be compatible with the agent's background knowledge: many valuable truth-purposive

[4] This relates to Contessa's (2007) distinction between external sentences (e.g. "The emission spectrum of hydrogen can be calculated with the Bohr model") and internal sentences (e.g. "In the Bohr model of the atom, electrons orbit around the nucleus in well-defined orbits"). I consider scientific hypotheses to be external, while internal sentences belong to the model itself or a description of it.

hypotheses presented in history firmly contradicted large portions of the adopted set of beliefs or (assumed) knowledge of those who suggested them. In such cases, the agent thought that pursuing the truth value of the hypothesis he had in mind might anyway lead to certain answers to his research question, even when he was well aware that parts of his background knowledge would need revision if this particular hypothesis turned out to be true.

With this final condition, I have excluded a large class of hypotheses from my characterization of scientific hypotheses: explicit counterfactuals and belief-negating hypotheses. Although these *truth-denying hypotheses* have their role in science by virtue of, for instance, thought experiments (De Mey 2006), I consider them fundamentally different from the *truth-purposive hypotheses* this paper deals with, as (a) their truth value is explicitly known or believed to be false and (b) they neither provide a direct answer to any particular question, nor is it their suggestor's aim to determine the hypothesis' truth value (as he already assumes it to be false). Their purpose is generally to set up a line of reasoning that can lead to certain sought-for answers via a detour, such as a thought experiment or a *reductio ad absurdum* argument.[5]

> **Scientific models**. I take *scientific models* to be (1) abstract or concrete artifacts (2) purposefully created in order to be manipulated to perform particular scientific tasks (such as prediction or explanation) by exploiting certain representational relations.

Although this characterization is in line with much of the actual usage of the notion '(scientific) model' by scientists and in the contemporary literature on models,[6] I have made some restricting choices.

[5] My distinction between truth-purposive and truth-denying hypotheses relates to Rescher's (1964) classic distinction between hypotheses with an unknown truth status, on the one hand, and belief-negating hypotheses and counterfactuals, on the other. However, there is one caveat: Rescher operates in a logical framework (which assumes logical omniscience). Therefore, for Rescher, it makes no difference whether the agent explicitly believes (or knows) that the hypothesis is false, or that this is only a consequence of his set of beliefs (or knowledge). For my purposes, this distinction does matter. A hypothesis is only truth-denying if the agent explicitly believes (or knows) that it is false. When the agent thinks that it might be true, it is truth-purposive, even if it is in contradiction with his set of beliefs (or knowledge). This situation actually occurs frequently in science: as many problems are overdetermined, scientists are often willing to accept that part of their set of beliefs (or assumed knowledge) is wrong in advancing a new hypothesis.

[6] This characterization is inspired by, amongst others, the views of Giere (2004, 2010), Hughes (1997), Teller (2001), Bailer-Jones (2003), Nersessian (2008) and Knuuttila (2011) and fits accounts of actual scientists reporting on their use of models (Bailer-Jones 2002).

First, ontologically, I consider models to be either concrete or abstract models, yet my focus will be on the abstract type. It is commonly accepted that the human imagination can create such things as abstract objects and that many scientific models, such as the ideal pendulum or the Bohr model of the atom, should be understood as such.[7] As it is my purpose to determine the relation between models and hypotheses, my analysis will unavoidably focus on abstract models. In principle, this would exclude from the analysis any tangible model, such as plastic models, diagrams, descriptive texts or annotated drawings. But this should not unduly concern us, as we can straightforwardly interpret most such tangible concrete models as representations of a particular abstract model,[8] while tangible models used for the direct representation of real target phenomena, such as a wooden bridge model, are not a part of our concern here.

Second, functionally, I take models to be used to represent some target system in the real (or empirical) world.[9] This is what Giere (1999b) has called the *representational conception* of models, as opposed to the *instantial conception* of models used in the semantic and structuralist analysis of theories. According to the representational conception, the intended representational relations can be exploited for predictive or explanatory purposes by manipulating the model.

Finally, the target system the representation of which the model is used for can also be a set of data points or measurements. Such models of data (Suppes 1962) or phenomenological models, which are generally constructed via statistical methods of data analysis, are sometimes seen as temporary models requiring further explanation by deeper explanatory or constitutive models (the 1885 Balmer formula for the hydrogen emission spectrum lines, for example,

---

[7] For the current debates on how this should be understood, see, amongst others, Godfrey-Smith (2009), Giere (2009) and Contessa (2010) but also Teller (2001) or French (2010) for an alternative position.

[8] Interpreting concrete or tangible models as representations of abstract models only makes sense if one adopts a three-place analysis of the representation relation: representation is not purely a relation between a model M and a target T, but a relation of an agent S who uses a model M to represent a target T for some purpose (Giere 2004). As such, concrete tangible models, such as a double helix made from cardboard, can be used in two ways: either to represent directly a target phenomenon (actual DNA), or to represent an abstract model (the Crick and Watson double-helix model), which is itself used to represent that same initial target (actual DNA). Although the particular form in which an abstract model is represented does influence the scientist's actual manipulations (Knuuttila 2011, Vorms 2011), I will pay no further attention to individual (tangible) models in the present paper.

[9] This is a choice. Although this characterization fits large classes of models in science, it does not fit all models (Downes 2011).

was explained by the 1913 Bohr model of the hydrogen atom). Yet, this type of model is often employed in actual scientific practice, especially for predictive purposes (consider, for instance, the importance of the discipline of data analysis), and is highly esteemed by scientists with a strongly inductivist mindset (see e.g. Glass and Hall in Section 3.3). Therefore, it is important that our analysis of the relation between models and hypotheses should apply to this type of model as well.

3. Four Stances on the Relation Between Models and Hypotheses

In this section I review four stances that can be found in the literature. However, it should be kept in mind that none of the authors I will associate with these stances was explicitly concerned to specify the relation between hypotheses and models. In each case, the characterization was embedded in a broader research goal.

3.1. Models Are (a Particular Form of) Hypotheses and the Concepts Can Be Used Interchangeably

Although the stance that models are just a form of hypotheses is never explicitly articulated in the current literature, it is often implied by the fact that the terms 'model' and 'hypothesis' are sometimes used more or less interchangeably. The idea is that models are just a particular form of hypotheses: they are a bit more elaborate, and often have some figurative elements, but in essence they are just hypothetical suggestions which can be tested to confirm whether they conform to reality. This view is particularly appealing to people focused on explanatory and mechanistic models, as for this kind of models it is intended that the parts of the model should have an accurate one-to-one correspondence relation with the parts of the target system.

This stance, however, neglects to take into account the important and currently hot issue of the representational relation between models and the world.[10] The representational relation between hypotheses and the world is rather straightforward to specify: hypotheses are linguistic entities. Therefore,

[10] See also footnote 8. For a discussion about the representational relation between models and the world see Van Fraassen (1980, 2008), Giere (1989, 2010), Knuuttila (2010) and Downes (2011).

whether they represent the world can be indicated by stating whether they are true or false. But models are not linguistic entities.[11] Therefore, one cannot determine whether a model is literally true or false. When a model is called true (or false), this attribution normally has to be understood in a *metaphorical* or *pragmatic sense*: it indicates that the model meets the purpose for which it was designed, such as accurate prediction or explanatory power, not that it consists of literally true sentences.[12] Even if one replaces truth with a gradual notion such as accuracy, it makes for some models no sense to assess whether or not they are accurate, because they were never intended to be so because of their use of idealizations, simplifications and fictional entities.

Finally, one might suggest that, although models are maybe not linguistic in nature and hypotheses are, they might still be interchangeable if every model would have a full characterization that is purely defined in linguistic terms (and which could, hence, act as the hypothesis of this model). After all, many models in science are known purely from a textual description, and Craver (2006) has introduced in the mechanism literature the notion of the *ideally complete description of a mechanism* as the ideal for a mechanistic model. Let us grant this for a moment, and assume that there exists for each model in science an ideal fully characterizing and fully linguistic description. Such a description of a model would indeed have a truth value. But it would be true only by reference to the model itself. If we were to determine its truth value by reference to the world, it would always be false. Models include fictitious entities (e.g. point masses or frictionless planes) or describe unreal and simplified conditions (e.g. no air resistance or uniform mass density) and even if a model is very descriptive, as are particular mechanism models in biology, its (ideally) full description would be false by reference to the world because of the simplifications and abstractions it incorporates.[13] For instance, the description

[11] There is a minority position that does take models literally as linguistic entities. This view, which is embedded in a syntactic view of theories, takes models (just like theories) to be a set of statements about a target system, simplified or idealized for certain purposes (Achinstein 1968, Redhead 1980). This position, however, has to cope with similar concerns as the syntactic view of theories. Moreover, it faces the obvious objection that there can be many different linguistic descriptions of the same model. How should the canonical description be determined? As a matter of fact, I have found no recent adherents of this position.

[12] Mäki (2011) has, however, tried to define a literal truth relation for models (see also Perini 2005 on the possibility of such a truth relation for pictorial representations), but, in essence, Mäki's proposal boils down to defining the truth of a model as the truth of the assertion that the driving mechanism of the model is the same as its target mechanism (which makes him rather fit the stance discussed in Section 3.2).

[13] See also Niiniluoto (2012, 2013) about the verisimilitude of models.

might state that one part is directly adjacent to another part, while in reality there are blood vessels, tissues and fat cells in between. Or, turning the argument around, if the ideally full description of a model were to be completely true with respect to the world, there would be no model defined, as the description would be just a direct description of this part of the world. We can conclude that if such a thing as the (ideally) full description of a model existed, it would be literally false with respect to the world and, hence, counterfactual.[14] Therefore, if we were to use this construction to call models hypotheses, they would be truth-denying hypothesis and not truth-purposive hypotheses, as their creators had likely intended.[15]

3.2. Hypotheses Are Statements about the Relation Between Fully Interpreted Models and their Target Systems

This is the idea Giere has been arguing for since his book *Explaining Science: A Cognitive Approach* (1988). According to him, (theoretical) hypotheses (which, he claims, overlap considerably with the use of the notion by scientists themselves) are assertions of some sort of relationship between a model and the system it is intended to represent. In his more recent work (2004, 2008, 2010), Giere specifies this notion of hypotheses further, holding that hypotheses are claims that a fully specified and interpreted model (a model of which each element is provided with a physical interpretation) fits a particular real system more or less well, or any generalization of such claims.

If one has come to appreciate that the relation between models and the world is not simply a matter of truth (or falsehood), but may include a plenitude of possible representational relations depending on the purposes of the agent, it is quite natural to understand scientific hypotheses as specifications of the nature and fit of these representational relations. For instance, many hypotheses state that the values calculated using a particular model fit particular measurements of the target system of the model (within certain error margins), or that the mechanism represented by a particular simplified and idealized model is the same mechanism driving a real target system. Perhaps because it is

[14] This analysis relates to the analysis of the falsehood of models by Cartwright (1983) and Wimsatt (2007[1987]).

[15] An exception to the general idea that modelers aim to be truth-purposive might be toy models, which are purposefully built not to represent much but rather to experiment with the theoretical tools themselves. Toy models could also be characterized as counterfactuals, and allow, therefore, for analysis both as models and as thought experiments.

natural to understand hypotheses in this bridging role, I have found no dissenting voices on this issue amongst scholars working on scientific models.

However, although this analysis is compelling and very suitable to account for a number of hypotheses used in actual scientific practice, it does not fit the majority of hypotheses advanced and defended in this practice. The reason for this is actually straightforward. Giere's characterization of a hypothesis depends on the existence of a model that can be fully interpreted. This means that this kind of hypotheses can be stated only once a fully interpretable model has been developed, which is typically only in the closing stages of the discovery process. Giere is not to blame for this. His project is to analyze how accomplished science is structured—the starting point of his 1988 investigation was a mechanics text book. But if we want to understand the role of hypotheses in scientific practice, we should take into account that hypotheses are much more closely linked to the discovery process than to the presentation of well-established science. In the process of scientific discovery, advanced hypotheses are seldom well-specified and fully interpretable (as the case studies in the next section show).

Therefore, although we can use Giere's account for a subclass of scientific hypotheses, i.e. the fully interpretable hypotheses (see section 5), we must complement it with an account of hypotheses used in the actual process of scientific discovery.

3.3. Radical Inductionism: Hypotheses Should be Avoided in Model Construction

Recently, Glass and Hall (2008) launched a well-argued attack in the top-ranked journal *Cell* on the use of hypotheses in scientific practice. The use of hypotheses, they argue, is a relic from the old hypothetico-deductive perspective on science, which denied induction as a valid form of reasoning. According to them, the latest articulation of this obsolete view, Popper's Critical Rationalism, has been successfully challenged in the second half of the 20^{th} century by, amongst others, Kuhn, while probability and Bayesianism gave the inductivist better tools to defend his position.

Apart from summarizing the main historical and philosophical positions in this well-known debate, Glass and Hall also argue on a pragmatic level that scientists would do better to replace top-down hypothesis testing with bottom-up inductive model-building. Framing research by hypotheses adds severe biases. Not only are negatives less valued than positives (confirmation bias), but also researchers are rendered blind to alternative routes, as negatives are not differentiated (categorization bias). Furthermore, not all interesting research (or research proposals) can be framed by a hypothesis. A telling example was the Human Genome Project, of which, when pressed to state a research hypothesis, J. Craig Venter, a major player in the project, stated that "It is our hypothesis that this approach will be successful" (Glass 2006, p. 18).

Therefore, Glass and Hall suggest that research (and research proposals) would better start by asking an open research question, after which data collection could begin. From this data, which is more and more abundant and elaborated in this Era of Big Data, the methods of statistical data analysis might extract a first model, which would lead to new questions, further data gathering and model refinement. Nowhere should one, according to this view, have to introduce unproven premises or hypotheses.

Glass and Hall's argument has the merit that it points out to scientists and funding organizations the danger of bias if research hypotheses are given too much weight. In fact, their suggestion to frame research proposals by open research questions instead of hypotheses (as is sometimes required by funding agencies) is an interesting one, but, philosophically, their suggestion to literally eradicate all hypotheses from scientific practice in favor of model-building cannot be taken seriously.

First, hypotheses are (implicitly) present at all stages of inductive model-building. Even when the research project is framed by a research question, choices will have to be made as to which variables should be tested for in obtaining the first data set. And such choices rely on (hidden) assumptions about which variables are plausible and which are not. For instance, if one is looking for the causal factors and catalysts of a particular disease, the data set will probably contain variables such as air quality or the diet or medical history of the test subjects, but not whether they are left- or right-handed or what their favorite ice cream topping is. These decisions as to which variables to include rely on initial hypotheses concerning what might plausibly be factors in the investigated disease.

Further, inductive model-building or statistical data analysis is a discipline crucially dependent on the introduction of assumptions to mold vast data sets into models that can be manipulated for scientific purposes. The discipline has been described as being "more an art, or even a bag of tricks, than science" (Good 1983). An often cited and telling example is the curve-fitting problem: given the simplest data set of only two variables, there are already an infinity of fitting mathematical functions. Data analysts constantly have to make decisions (based on assumptions) on how to handle outliers, on the tradeoff between simplicity and data fitting, on how the data is best represented (as this influences model construction), on how the variable is spread in the population (is it normally distributed or not?), and so on.

Finally, Glass and Hall's analysis is very focused on scientific experimentation, and their generalization is based on the old inductive idea that the whole process of scientific discovery can be reduced to inferences from data. It was precisely against this view that Nickles (1980) and other philosophers of scientific discovery have argued: discovery, they hold, is not separate from theoretical considerations and choices. As the examples in the next section will show, many models originate from theoretical considerations. Only later on, when sufficient detail is attained, can they be compared with experimental data or models of data.

However, Glass and Hall's analysis is not completely without value as their analysis does (largely) apply to the models of data and phenomenological models mentioned above. It does, however, not apply to explanatory or constitutive models about which the literature on models in science generally talks. To them is the burden to argue how this latter kind of models could be constructed without hypotheses.

3.4. Heuristic View: Hypotheses are Necessary Guidelines in Model Construction

A view opposite to the previous stance is that hypotheses somehow have a heuristic and methodological role in the process of model construction. Although this idea is sometimes mentioned (e.g. Nola and Sankey 2007, p. 25), it is more often implicitly assumed. In the remainder of this paper, I will give an explicit account of this stance, which could then complement the second position to give a full analysis of the relation between models and hypotheses.

In my view, heuristic hypotheses are direct attempts to initially answer the research question, but, precisely because the research still needs to be done, they unavoidably contain vague filler terms or black boxes and can do little more than hint at a particular direction of research. Yet, by this hinting they sketch an outline or rough blueprint, or even maybe just identify the type of the model(s) needed to substantiate the initial hypothesis. As such, they reduce the initial research problem to the more specific problem of filling in the black boxes of the model outline, resulting finally in an adequate model, of which a fully specified and interpreted hypothesis (in Giere's sense, see section 3.2), if confirmed, can provide an answer to the initial research question.

Before I give a detailed account of this position in Sections 5 and 6, I will first present in Section 4 three case studies that will allow me to benchmark this analysis.

4. Three Cases from Astronomy

In this section, I introduce three historical cases that illustrate my analysis of the role of hypotheses in model-based science. Due to space restrictions, only the first case will be fully elaborated; for the other two cases, only the key steps in my analysis will be indicated, together with further references to the literature.

4.1. The Energy Source of the Stars (1920-1930s)[16]

Around 1920, the source of stellar energy was still a mystery. By that time, Eddington had crafted the basic structural model of a (stable) star, largely confirmed by the observations at the time. His model represented stars as spheres of gas in which, at each internal point, there was an equilibrium between the inward gravitational pressure and the outward gas and radiation pressure, resulting in concentric layers of increasingly lower pressures and temperatures towards the surface.

But the source of the stellar radiant energy was still a mystery. Clearly, it could not be the result of a chemical reaction, such as exothermic oxidation (fire). Even if the Sun would be totally composed of carbon, its mass would be barely enough to radiate the Sun's current luminosity for a few thousand years. To solve this problem, von Helmholtz and Kelvin had defended in the 19th century what was later referred to as the *contraction hypothesis*, which was in turn inspired by the *nebular hypothesis* for the origin of our solar system by Kant and Laplace.[17] Von Helmholtz and Kelvin took as the source of stellar energy the inward gravitational energy provided, at first, during the accretion of the star and, after it has started to radiate, by the contraction of the star as it cools down. Using this model, Kelvin estimated the age of the solar system to be on

[16] For a thorough and detailed version of this history, see Shiaviv (2010). For a good introduction see Bahcall (2000) or Mazumdar (2005).

[17] This hypothesis situates the origin of our solar system in the gravitational collapse of a gaseous nebula (Kant 2012[1755]).

the order of ten million years—in contradiction to estimates based on the biological and geological record. For instance, Darwin suggested in the *Origin of Species*, based on some geological calculations, that the Earth was at least three hundred million years old, the time he thought to be necessary for the evolution of our current biodiversity.[18] As a matter of fact, this whole situation led to a public controversy between these two leading scientists.

At the dawn of the 20th century, better geological observations and the discovery of radioactivity quickly discredited the contraction model. The Earth (and, hence, the Sun) must be older than Kelvin's estimate. Therefore, the contraction model could not supply the requisite energy. Many looked at the new physics that was emerging, hoping it could provide an answer. Rutherford and the young Eddington suggested that radioactive elements might be the source of stellar energy, and Jeans, upon learning of Einstein's $E = mc^2$, suggested that in the extremely hot interior of stars, protons and electrons might annihilate each other, turning their mass into energy.

The experimental breakthrough that led to Eddington's initial suggestion of nuclear fusion was Ashton's measurements of the mass of *He* and *H* nuclei, finding that the mass of a *He* nucleus was only 99,3% of the combined mass of the four hydrogen nuclei it contained. This led Eddington to the hypothesis of *nuclear fusion*:

> Now mass cannot be annihilated, and the deficit can only represent the mass of the electrical energy set free in the transmutation. [...] If 5 per cent of a star's mass consists initially of hydrogen atoms, which are gradually being combined to form more complex elements, the total heat liberated will more than suffice for our demands, and we need look no further for the source of a star's energy. (Eddington 1920, p. 353)

This suggestion, although defended fiercely, is clearly just a hypothesis. Apart from Ashton's measurements, he had little or no evidence to back it up, nor did he understand how and when such a fusion process might occur. After all, one should not forget that at the time, neither the neutron nor any nucleus of atomic mass 2 or 3 had yet been discovered. Quantum mechanics had not yet been developed and the amount of hydrogen in the Sun was not yet determined. So, Eddington's hypothesis suggested that somehow four protons and two electrons (which it was thought, at the time, the *He* nucleus consisted of) come to-

[18] At the moment, it is widely accepted that the age of our solar system is approximately 4.6 billion years old, while the earliest evidence of life on Earth is about 3.5 billion years old.

gether at one position at a given time, something which Eddington knew was probabilistically nearly impossible, as is illustrated by the following quote:

> Indeed the formation of helium is necessarily so mysterious that we distrust all predictions as to the conditions required. [...] How the necessary materials of 4 mutually repelling protons and 2 electrons can be gathered together in one spot, baffles imagination. (Eddington 1926, p. 301)

Therefore, it is understandable that throughout the 1920s his hypothesis still met with competitors: Jeans kept defending a proton-electron annihilation, while Bohr even thought that in stars the conservation of energy was violated.[19] It was only after numerous contributions of the likes of Gamov, Houterman, Atkinson and Weizsäcker that Bethe (1939) finally put forward a model of stellar energy production in satisfactory agreement with the observational record, which consisted of two well-described processes that converted hydrogen into helium: the p-p chain and the CNO cycle (the latter occurring only in stars more massive than the Sun).

Let us review the various characteristics of Eddington's hypothesis of nuclear fusion. Clearly, it fits our characterization: it is a claim about the world with an unknown truth value in answer to a particular research question. In fact, it would be better to state that its truth value is underdetermined. Eddington had no idea how energy could be liberated by combining atoms. There are many possible models—some even totally different from Bethe's model with completely different concepts, elements and forces—that could still be seen as a specification of Eddington's hypothesis.[20]

Still, the credit that Eddington received for this suggestion is justified, as his suggestion was immensely important in redirecting research. In a sense, it simplified the problem of what the source of stellar energy was to the question of how hydrogen nuclei can combine so as to form helium nuclei, a process involving entities that could also be studied in laboratories on Earth. This sim-

[19] Bohr's suggestion (1986[1929]) to renounce energy conservation must be linked primordially with the problem of the continuous β spectrum (Gauderis 2014), but the way in which Bohr combined it with this problem of astrophysics, a field to which he has not contributed at all, shows how pressing the problem of stellar energy still was around 1930.

[20] Consider, for instance, also the briefly mentioned nebular hypothesis. Our current model of the origin of our solar system differs completely from what Kant had in mind (Kant 2012[1755], Palmquist 1987). Still, our current model for the origin of our solar system can be seen as a specification of Kant's severely underdetermined original hypothesis.

plification is achieved by providing an initial answer to the question of stellar energy, using a sketchy outline of a stellar model containing a black box process that somehow turns present hydrogen into helium. This is why his idea was so hugely important and why he kept on defending it and urging research in that direction for twenty years, until, finally, Bethe was able to crack open the black box.

So what is the nature of the relation here between model and hypothesis? Eddington's model was largely a black box or at most a rough outline, so Giere's characterization of hypotheses does not apply to his hypothesis, because Eddington's model could not be fully specified or provided with a physical interpretation. His hypothesis was heuristic in nature. Only once Bethe's model was available could one say that Eddington's hypothesis, refined by stating that the "combination of hydrogen atoms" has to occur according to Bethe's model, is a hypothesis in Giere's sense: a claim that a fully interpreted model fits a target system.

4.2. The Nice Model (2000s)

In 2001, simulations of the model specified by the nebular hypothesis (describing the origin of our solar system), with reasonable assumptions for the initial conditions, confirmed the idea raised a few years earlier that Neptune could not have become such a large planet at such a great distance from the Sun (Stewart & Levison 1998, Levison & Stewart 2001)—a research problem that triggered, amongst other possible solutions, the hypothesis that Neptune initially formed nearer to the Sun and then migrated out (Thommes *et al.* 1999). Yet this hypothesis was nearly meaningless, as no available model showed how such a migration could have occurred. In 2005, in a series of three papers in *Nature*, the Nice model[21] was presented (Tsiganis *et al.* 2005, Morbidelli *et al.* 2005, Gomes *et al.* 2005). This model postulates that four billion years ago there was a period in which Jupiter and Saturn were in 2:1 orbital resonance.[22] This led to a global gravitational instability in our solar system that caused the outer pla-

[21] This model, named after the French Mediterranean city where the research was conducted, is generally represented and explored via computer simulations. It is yet an open debate how models and simulations relate. See among others Humphreys (2004), Frigg and Reiss (2009), Winsberg (2010).

[22] This means that one orbit of Saturn takes exactly as long as two orbits of Jupiter. Hence, the direction where they line up (with respect to the Sun) and coerce their combined gravitational pull on the rest of the solar system, remained the same for several thousands of years.

nets to move from orbits much nearer to the Sun outwards to their current trajectories. Furthermore, simulations of this model showed that it also explained many other curious features of our solar system, such as the Late Heavy Bombardment (that caused the many lunar craters), the heavy eccentricities of the outer planets' orbits, and the Trojan satellites locked in Jupiter's orbit. In subsequent years, improved simulations and new explanations of further features of the solar system, such as the characteristics of the Kuiper belt, have made the Nice Model generally accepted (Crida 2009).

The Nice model is clearly a very different type of model than the stellar model from Section 4.1. Where the stellar model was mainly a very general theoretical model applicable to any star, the Nice model is an applied model tailored to our solar system, established by numerous computer simulations, in which mainly the initial conditions were sought that, given the principles of a well-known theory (Newtonian dynamics), could result in the observed specificities of our solar system.

Still, we find here the same type of relation between the model and the heuristic hypothesis that led to its development. The initial suggestion, i.e. that Neptune formed closer to the Sun and then migrated out due to gravitational forces in our solar system, provided a first tentative but direct answer to the research question of why Neptune was so massive. Yet, this suggestion was largely vacuous without an exact model or initial conditions to specify how such a migration might have occurred. On the other hand, it was precisely the persuasive plausibility of this initial heuristic hypothesis that motivated and coordinated a large research effort to conduct the numerous computer simulations that led to the substantiation of this claim by explicating the unknown mechanism of Neptune's migration. Only now that this model has been built can we reformulate the hypothesis as a fully interpretable hypothesis in Giere's sense: Neptune formed closer to the Sun and then migrated out according to the conditions and the mechanism described by the Nice model.

4.3. Dark Matter (1930s-Present)[23]

Notwithstanding some earlier references to dark stars or matter, the start of the modern search for dark matter is to be found in Zwicky (2009[1933]). Having found that the galaxies in the Coma Cluster rotate way too high around their center to be explained by the gravitational forces of the visible stars, he sug-

[23] Classic histories of dark matter are Trimble (1987), Van den Bergh (1999), Rubin (2003).

gested that dynamical models of galaxies should incorporate the presence of non-visible dark matter to explain the observed rotational speeds. In the following decades, the problem was largely cast aside, although a growing number of studies for different galaxies confirmed the high rotational speeds. Gradually, more galactic models incorporating dark matter were advanced, attributing more and more features to it. For instance, Ostriker and Peebles (1973) calculated that, in contrast with visible matter which is mostly found in the galactic disk, dark matter is mostly present in the galactic halo. The enumeration of the various indications of its existence in a highly-influential review paper of Faber and Galagher (1979) convinced most astrophysicists of its existence by 1980. In subsequent decades, we saw an enormous increase in the number of suggestions to characterize dark matter, while some possibilities, such as neutrinos or brown dwarfs and other massive dark astronomical bodies (so-called MACHOs), could already be ruled out. At the same time, other hypotheses have been raised to address the initial problem of the galactic rotation curves (e.g. the MOND hypothesis proposed a modification of Newtonian dynamics), but we also saw an increase in the use of the concept 'dark matter' in models that explain other features of our galaxy, such as gravitational lensing or fluctuations in the cosmic background radiation. Nowadays, the fact that the concept is incorporated in virtually any successful galactic or cosmological model is considered by almost everyone to be sufficient proof of its existence. On the other hand, although some possibilities have already been ruled out and some characteristics have already been determined, there is still no satisfactory account of the nature of dark matter. The best guess at present is that it consists of unknown weakly interacting massive particles (so-called WIMPs).

This final case, about a not yet specified hypothetical entity, might seem different from the other two cases. Yet, also here we can find the same interplay between hypotheses and models, the only difference being that, in this case, most of our present models cannot be fully interpreted and specified (in Giere's sense), as dark matter is not yet fully understood. Zwicky's initial heuristic hypothesis, i.e. that there exists a large amount of dark matter in galaxies, has, despite its neglect at the time it was proposed, redirected much research toward specifying the nature of this unknown type of matter and supplementing this claim with suitable models. But, although galactic and cosmological models including dark matter have been substantially refined over the years and have become the only widely accepted models, and even if these models can be operationalized for some explanatory or predictive purposes, the notion 'dark matter' still remains something of a black box in these models.

5. Heuristic and Fully Interpretable Hypotheses

Before turning to the relation between models and hypotheses in model-based scientific practice, let me first draw more precisely the distinction I have been hinting at between two types of hypotheses: *heuristic hypotheses* and *fully interpretable hypotheses*, a distinction that draws on Craver's (2006) distinction between *mechanism sketches* and *ideally complete descriptions of mechanisms*.[24]

A *fully interpretable hypothesis* is a hypothesis the meaning of which (or any part of which) leaves no room for vagueness or ambiguity. In other words, expressions of such hypotheses do not contain any unexplained *filler terms*, terms such as 'process,' 'to interact,' or 'entity' that have a broad and generic meaning covering up some uncertainty, imprecision or unknown details.[25] Hence, these hypotheses are fully expressed in terms with a precise meaning, which is provided either by the conceptual framework of the field the researcher is working in, or by the researcher himself by means of suitable models. *Heuristic hypotheses*, on the other hand, do contain such unspecified and generic filler terms.[26]

The main idea is that heuristic hypotheses are both unavoidable and useful in the early stages of scientific discovery, as they sketch an early blueprint or incomplete model without committing one to too much (yet unknown) detail. A heuristic hypothesis suggests that research should proceed in a particular direction, i.e. that it aims to fill gaps in the incomplete model instead of trying to address the general research question directly. Fully interpretable hypotheses, on the other hand, can be put forward only after the construction of a full model that specifies how the hypothesis (which is a claim about a part of reality) should be interpreted precisely and under what conditions it should hold. Therefore, in principle, it is possible to design a conclusive experiment to verify whether a fully interpretable hypothesis holds, while heuristic hypotheses can seldom be tested conclusively due to their vagueness and ambiguity. Experiments in this case mostly aim to refine the model and reduce the vagueness and ambiguity.

[24] The notion of a mechanism sketch had already been introduced in the seminal paper on mechanisms by Machamer, Darden and Craver (2000).

[25] This does not mean that the hypothesis cannot contain any approximations or abstractions.

[26] Heuristic hypotheses are, however, still real truth-conductive hypotheses, which aim to provide directly answers to particular research questions.

Before I add some further remarks and consider some examples, it is useful first to explain how these two types of hypotheses relate. As the main criterion that distinguishes these two types is the amount of precision in the expression of the hypotheses, the two distinguished types are actually the extremes of a continuum. Moreover, as it is an unwieldy (if even possible) task to specify all relevant conditions for a particular hypothesis, it is clear that the idea of a fully interpretable hypothesis is actually an idealization (as Craver could only speak of ideally complete descriptions of mechanisms). Therefore, at first sight, it seems as if there exist only heuristic hypotheses, interpretable to a greater or lesser extent. In scientific practice, however, some hypotheses are clearly considered to be sufficiently unambiguous and interpretable, allowing them to be tested conclusively. Therefore, for our purposes, we can evade this conclusion by allowing for a pragmatic or social epistemological threshold of precision sufficient for full interpretability. A hypothesis can be considered *sufficiently fully interpretable* if it invokes no disagreement in the research community as to which is its meaning. Yet the flip side of adopting this social epistemological criterion is that a single researcher cannot himself decide whether a hypothesis is fully interpretable. Also, that a particular hypothesis is considered to be fully interpretable at a certain point in time does not warrant that it will remain so in the indefinite future.

A few further remarks are in order concerning the concept of filler terms, including some examples. First, what counts as a filler term is topic dependent. For instance, the phrase ‘exerting a force’ has a precise meaning in physics, while in economics this would be a filler term for an unspecified process of influence. Having said this, the fact that so many words in various fields can be considered to have a precise meaning is precisely because of the cumulative processes of abstraction and concept formation in these sciences. Therefore, whether a phrase counts as a filler term or whether it has a precise meaning (in a particular reference framework) is dependent on the stage of development in the field. Let me return to the examples presented in Section 4. When Eddington in 1920 spoke of “the combination of hydrogen atoms” and somewhat later even used the term ‘nuclear energy,’ these concepts were certainly filler terms. Despite having good arguments why focusing on a possible transition from nuclear mass to energy could possibly solve the problem of stellar energy, he did not have any account of how this energy could be released from the nucleus and why this process occurred in stars. It was only after the acceptance of Bethe’s 1939 nuclear fusion models for the pp-cycle and CNO-cycle that the term ‘nuclear energy’ received a precise meaning in astrophysics.

Also, filler terms generally gain precision only gradually. For instance, while the concept 'dark matter' was at first a pure filler term to indicate the possibility of unobserved but present matter, the term has gained some precision and delineation over the past decades. It is now accepted that dark matter mostly resides in galactic halos, that there is at least five times more dark matter than regular matter, that it consists of weakly interacting massive unknown particles (WIMPs), which move at relatively slow speeds (with respect to the speed of light) and which are electrically neutral, etc. Yet no astronomer at present would claim that the concept of dark matter is fully understood and precisely defined.

Finally, the given examples might suggest that in the discovery process filler terms themselves always gain a more precise meaning. This happens, such as in the case of 'nuclear energy' or 'dark matter', but more often vague filler terms are replaced with more meaningful descriptions, names or acronyms, such as 'nuclear fusion' or 'WIMP'.

So how do these hypotheses relate to models? For fully interpretable hypotheses, as indicated in Section 3.2, I follow Giere in the sense that such hypotheses are claims that a fully specified model provided with a physical interpretation fits a target system more or less well. This idea can now be extended to heuristic hypotheses. Heuristic hypotheses are also claims that a particular model or model type fits a target system more or less well, but in this case, as the models are just bare model sketches containing black boxes labeled by filler terms, this claim should be understood as the weaker claim that a full specification and interpretation of the model sketch that would fit the target system is possible. But in providing such a model sketch, the initial research question is already partially answered, while at the same time the direction is shown for future research, i.e. to fill in the black boxes.

6. The Role of Hypotheses in Model-Based Scientific Practice

Let me now spell out the role of these two types of hypotheses in the process of scientific discovery in model-based science. This view will incorporate the two theses I defended above, i.e. that hypotheses are necessary in the process of model construction and that hypotheses that are not fully interpretable are valuable and even needed in this process.

In general, research aimed at constructing models is triggered by a research question or trigger. In her monograph on abductive reasoning (the inference from observations to explanatory hypotheses), Aliseda (2006) distinguishes between anomalies and novelties as the two types of observational triggers for abductive reasoning. This classification can be adopted for our current purposes if we keep in mind the main criticism developed by Nickles (1980) and other scholars of scientific discovery against the idea that abductive reasoning could be the logic of scientific discovery (as suggested by Hanson 1958), i.e. that abductive reasoning neglects the triggering role of theory in scientific discovery. Much research is fueled by theoretical considerations, but also here we can distinguish between questions triggered by contradictions (related to experimental anomalies) and questions triggered by lacunas (related to experimental novelties). Therefore, I conceive of four triggers for research aiming at the construction of models: *experimental* (or *observational*) *novelties*, *experimental* (or *observational*) *anomalies*, *theoretical gaps* or *lacunas* and *theoretical contradictions*.

In model-based discovery, these triggers or research questions are answered at the end of the research process by proposing a model and claiming that its similarities with the target system can be exploited to sufficiently address the research question, or, in other words, by stating a (sufficiently) interpretable hypothesis whose claim is sufficiently verified.

As the model is only linked to the trigger or research question through a hypothesis claiming its fit, such a linking hypothesis, constituting the (partial) answer to the research question, must be present through all stages of model construction; though in the early stages it will heuristic in nature, not fully interpretable.

Now we have to investigate what the role of these heuristic hypotheses is in the research process itself. If we take a constraints-based view of scientific discovery, the view Nickles (1978) developed in the tradition of scientific research as problem solving, we can conceive of a scientific problem (or research question) as a set of constraints. Progressing on a problem consists in manipulating these constraints such that the problem turns into a simpler problem or a problem that is easier to solve.

In the case of suggesting a heuristic hypothesis as an initial partial answer to a research problem, one deliberately adds a constraint: however vague a heuristic hypothesis might be, it excludes particular solutions and direct research in a particular direction. As such, one progresses on the problem by reducing it to a simpler problem, though always at the risk that one will not find a solution along these lines (if the heuristic hypothesis turns out to have been a wrong path from the start). After reducing the initial research problem to the simpler problem of finding a suitable model to fill in the filler terms, the heuristic hypothesis remains important as the link between the reduced problem and the initial research question, as it shows how the latter can be answered by means of the answer to the reduced problem.

Let me illustrate this role of heuristic hypotheses with some of the cases of Section 4. Eddington reduced the open problem of stellar energy (a theoretical gap) to the more restricted problem of how hydrogen could combine so as to form helium. After the problem was reduced to finding a suitable model for this combination, Eddington's hypothesis remained the link that allowed the answer to this reduced problem, namely Bethe's model of hydrogen fusion, to be used to answer the initial research question of where stellar energy originated. By the time Bethe's model was developed, Eddington's hypothesis could be considered a fully interpretable hypothesis.

Similarly, the research question of the improbable accretion of Neptune (an observational anomaly) was reduced by the initial heuristic hypothesis to the more straightforward problem of constructing a model and determining the initial conditions for an outward-directed gravitational slingshot of a planet within our solar system. Only when such a model—the Nice model—was constructed through numerous computer simulations could the original hypothesis that Neptune initially formed much closer to the Sun and migrated outwards be considered as the fully interpretable answer to the initial research question or trigger.

A final thing to address is the fact that many research triggers have the form of an anomaly or a contradiction. Heuristic hypotheses addressing such research questions unavoidably sometimes contradict major parts of the agent's (assumed) background knowledge. Yet as history shows, this clearly does not prevent scientists from coming up with heuristic hypotheses for such overdetermined problems. In such cases, scientists reason according to what Rescher (1960) has called belief-negating (or even knowledge-negating) hypothetical reasoning: they assume the hypothesis while retaining all beliefs from their belief set that are compatible with it, and suspending judgment on beliefs that are contradictory to it. For instance, in the case of Neptune, researchers had at first to suspend judgment on the idea that the planets in our solar system were formed where we observe them today, while retaining acceptance of full Newtonian dynamics. The beliefs compatible with the heuristic hypothesis then become the basis for solving the reduced problem of the construction of a suitable model to interpret this heuristic hypothesis. Only once the model is verified and the research question answered, the initially incompatible beliefs on which judgment was suspended can then be revised.

7. Conclusion

In this article I have addressed the relation between models and hypotheses in model-based science. After reviewing and pointing out the shortcomings of various stances in the literature, I presented my own view on the matter.

First, a distinction has to be made between heuristic hypotheses and fully interpretable hypotheses. Heuristic hypotheses are initial and partial answers to research questions that contain necessarily vague filler terms, yet sketch the outline for the type of model that might be needed to answer the research question. Fully interpretable hypotheses, on the other hand, are claims concerning how a fully constructed model can be used to provide an answer to the research question.

Next, I have shown in this article, by examining three cases from astronomy, how initial heuristic hypotheses fuel the process of model construction and how, once the requisite models are built, they gradually evolve into fully interpretable hypotheses that can, if verified, serve as answers to the initial research questions.

REFERENCES

ACHINSTEIN, P. (1968). *Concepts of Science. A Philosophical Analysis*. Baltimore: Johns Hopkins Press.

ALISEDA, A. (2006). *Abductive Reasoning. Logical Investigation into Discovery and Explanation*. Dordrecht: Springer.

BAHCALL, J. (2000). How the Sun Shines. Retreived October 15th, 2013 from www.nobelprize.org/nobel_prizes/themes/physics/fusion/.

BAILER-JONES, D. (2002). Scientists' Thoughts on Scientific Models. *Perspectives on Science*, **10**, 275-301.

——— (2003). When Models Represent. *International Studies in the Philosophy of Science*, **17**, 59-74.

BALZER, W., ULISES MOULINES, C., & SNEED, J. (1987). *An Architectonic for Science. The Structuralist Program*. Dordrect: Reidel.

BETHE, H. (1939). Energy Production in Stars. *Physical Review*, **55**, 434-456.

BOHR, N. (1986[1929]). β Ray Spectra and Energy Conservation. In R. Peierls (Ed.), *Niels Bohr Collected Works. Vol. 9: Nuclear Physics (1929-1952)* (pp. 85-89). Amsterdam: North Holland Physics. (Original unpublished manuscript written 1929.)

CARTWRIGHT, N. (1983). *How the Laws of Physics Lie*. New York: Oxford University Press.

CONTESSA, G. (2010). Scientific Models and Fictional Objects. *Synthese*, **172**, 215-229.

CRAVER, C. (2006). When Mechanistic Models Explain. *Synthese*, **153**, 355-376.

CRIDA, A. (2009). Solar System Formation. Retrieved from arXiv:0903.3008v1 [astro-ph.EP]

DE MEY, T. (2006). Imagination's Grip on Science. *Metaphilosophy*, **37**, 222-239.

DOWNES, S. (1992). The Importance of Models in Theorizing: A Deflationary Semantic View. In D. Hull, M. Forbes, & K. Okruhlik (Eds.), *PSA 1992, vol. 1* (pp. 142-153). East Lansing, MI: Philosophy of Science Association.

——— (2011). Scientific Models. *Philosophy Compass*, **6**, 757-764.

EDDINGTON, A. (1920). Presidential Address to section A of the British Association at Cardiff (24th Augustus 1920). *Observatory*, **43**, 353.

——— (1926). *The Internal Constitution of Stars*. Cambridge: At the University Press.

FABER, S., & GALLAGHER, J. (1979). Masses and Mass-to-Light Ratios of Galaxies. *Annual Review of Astronomy and Astrophysics*, **17**, 135-187.

FRENCH, S. (2010). Keeping Quiet on the Ontology of Models. *Synthese*, **172**, 231-249.

FRIGG, R., & REISS, J. (2009). The Philosophy of Simulations: Hot New Issues or Same Old Stew? *Synthese*, **169**, 593-613.

FRIGG, R., & HARTMANN, S. (2012). Models in Science. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2012 edition)*. Retrieved from http://plato.stanford.edu/archives/fall2012/entries/models-science/.

GAUDERIS, T. (2014). To Envision a New Particle or Change an Existing Law? Hypothesis Formation and Anomaly Resolution for the Curious Case of the β Decay Spectrum. *Studies in History and Philosophy of Modern Physics*, **45**, 27-45.

GLASS, D. (2006). *Experimental Design for Biologists*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

GLASS, D., & HALL, N. (2008). A Brief History of the Hypothesis. *Cell*, **134**, 378-381.

GIERE, R. (1988). *Explaining Science. A Cognitive Approach*. Chicago: University of Chicago Press.

——— (1999a). *Science withouth Laws*. Chicago: University of Chicago Press.

——— (1999b). Using Models to Represent Reality. In: L. Magnani, N. Nersessian & P. Thagard (Eds.), *Model-Based Reasoning in Scientific Discovery* (pp. 41-57). New York: Kluwer/Plenum.

——— (2004). How Models are Used to Represent Reality. *Philosophy of Science*, **71**, 742-752.

——— (2008). Models, Metaphysics and Methodology. In: S. Hartmann, C. Hoefer & L. Bovens (Eds.), *Nancy Cartwright's Philosophy of Science* (pp. 123-133). New York: Routledge.

——— (2009). Why Scientific Models Should not be Regarded as Works of Fiction. In M. Suárez (Ed.), *Fictions in Science. Philosophical Essays on Modeling and Idealisation* (pp. 248-258). London: Routledge.

——— (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.

——— (2009). Models and Fictions in Science. *Philosophical Studies*, **143**, 101-116.

GOMES, R., LEVISON, H., TSIGANIS, K., & MORBIDELLI, A. (2005). Origin of the Cataclysmic Late Heavy Bombardment Period of the Terrestrial Planets. *Nature*, **435**, 466-469.

GOOD, I. (1983). The Philosophy of Exploratory Data Analysis. *Philosophy of Science*, **50**, 283-295.

HANSON, N. R. (1958). *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press.

HARRIS, T. (2003). Data Models and the Acquisition and Manipulation of Data. *Philosophy of Science*, **70**, 1508-1517.

HUGHES, R. (1997). Models and Representation. *Philosophy of Science*, **64** (supplement), S325-S336.

HUMPHREYS, P. (2004). *Extending Ourselves. Computational Science, Empiricism, and Scientific Method*. Oxford: Oxford University Press.

KANT, I. (2012[1755]). Universal Natural History and Theory of the Heavens. In: E. Watkins (ed.), *Natural Science. The Cambridge Edition of the Works of Immanuel Kant* (pp. 182-308). Cambridge: Cambridge University Press. (Original published in German in 1755.)

KNUUTTILA, T. (2011). Modelling and Representing: An Artefactual Approach to Model-Based Representation. *Studies in the History and Philosophy of Science*, **42**, 262-271.

KROON, F. & VOLTOLINI, A. (2011). Fiction. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*. Retrieved from http://plato.stanford.edu/archives/fall2011/entries/fiction/.

LEURIDAN, B. (2013). The Structure of Scientific Theories, Explanation, and Unification. A Causal-Structural Account. *The British Journal for the Philosophy of Science*, in print. doi:10.1093/bjps/axt015.

LEVISON, H. & STEWART, G. (2001). Remarks on Modeling the Formation of Uranus and Neptune. *Icarus*, **153**, 224-228.

MACHAMER, P., DARDEN, L. & CRAVER, C. (2000). Thinking about Mechanisms. *Philosophy of Science*, **67**, 1-25.

MÄKI, U. (2011). Models and the Locus of their Truth. *Synthese*, **180**, 47-63.

MAZUMDAR, I. (2005). Nucleosynthesis and Energy Production in Stars: Bethe's Crowning Achievement. *Resonance*, **10**, 67-77.

MORBIDELLI, A., LEVISON, H., TSIGANIS, K., & GOMES, R. (2005). Chaotic Capture of Jupiter's Trojan Astreroids in the Early Solar System. *Nature*, **435**, 462-465.

MORRISON, M. & MORGAN, M. (1999). Models as Mediating Instruments. In: M. Morgan & M. Morrison (Eds.), *Models as Mediators. Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

NERSESSIAN, N. (2008). *Creating Scientific Concepts*. MIT Press, Cambridge, Massachussets.

NICKLES, T. (1978). Scientific Problems and Constraints. In: P. Asquith & I. Hacking (Eds.), *PSA 1978. Proceedings of the Biennial Meeting of the Philosophy of Science Association* (pp. 134-148). East Lansing, Michigan: Philosophy of Science Association.

——— (1980). Introductory Essay: Scientific Discovery and the Future of Philosophy of Science. In: T. Nickles (Ed.), *Scientific Discovery, Logic and Rationality* (pp. 173-183). Dordrecht: Reidel.

NIINILUOTO, I. (2012). The Verisimilitude of Economic Models. In: A. Lehtinen, J. Kuorikoski & P. Ylikoski (Eds.), *Economics for Real: Uskali Mäki and the Place of Truth in Economics* (pp. 65-80). London: Routledge.

——— (2013). Models, Simulations and Analogical Inference. In: V. Karakostas and D. Dieks (Eds.), *EPSA11: Perspectives and Foundational Problems in Philosophy of Science* (pp. 19-27). Dordrecht: Springer.

NOLA, R., & SANKEY, H. (2007). *Theories of Scientific Method. An Introduction*. Durham: Acumen.

OSTRIKER, J. & PEEBLES, P. (1973). A Numerical Study of the Stability of Flattened Galaxies: Or, can Cold Galaxies Survive? *The Astrophysical Journal*, **186**, 467-480.

PALMQUIST, S. (1987). Kant's Cosmogony Re-Evaluated. *Studies in History and Philosophy of Science*, **18**, 255-269.

PERINI, L. (2005). The Truth in Pictures. *Philosophy of Science*, **72**, 262-285.

REDHEAD, M. (1980). Models in Physics. *The British Journal for the Philosophy of Science*, **31**, 145-163.

RESCHER, N. (1964). *Hypothetical Reasoning*. Amsterdam: North-Holland.

RUBIN, V. (2003). A Brief History of Dark Matter. In M. Livio (Ed.), *The Dark*

Universe. Matter, Energy and Gravity (pp. 1-13). Cambridge: Cambridge University Press.

RUTHERFORD, E. (1911). The Scattering of α and β Particles by Matter and the Structure of the Atom. *Philosophical Magazine*, Series 6, **21**, 669-688.

SHIAVIV, G. (2010). *The Life of Stars: The Controversial Inception and Emergence of the Theory of Stellar Structure*. Heidelberg: Springer Verlag.

STEWART, G. & LEVISON, H. (1998). On the Formation of Uranus and Neptune. *Proceedings of the 29th Annual Lunar and Planetary Science Conference*, abstract no. 1960.

SUPPES, P. (1960). A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences. *Synthese*, **12**, 287-301.

——— (1962). Models of Data. In: E. Nagel, P. Suppes & A. Tarski (Eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* (pp. 252-261). Stanford: Stanford University Press.

SUPPE, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. Urbana, Illinois: University of Illinois Press.

——— (Ed.) (1977). *The Structure of Scientific Theories* (2nd ed.). Urbana, Illinois: University of Illinois Press.

TELLER, P. (2001). Twilight of the Perfect Model Model. *Erkenntnis*, **55**, 393-415.

THOMMES, E., DUNCAN, M., & LEVISON, H. (1999). The Formation of Uranus and Neptune in the Jupiter-Saturn Region of the Solar System. *Nature*, **402**, 635-638.

TRIMBLE, V. (1987). Existence and Nature of Dark Matter in the Universe. *Annual Review of Astronomy and Astrophysics*, **25**, 425-472.

TSIGANIS, K., GOMES, R., MORBIDELLI, A., & LEVISON, H. (2005). Origin of the Orbital Architecture of the Giant Planets of the Solar System. *Nature*, **435**, 459-461.

VAN DEN BERGH, S. (1999). The Early History of Dark Matter. *Publications of the Astronomical Society of the Pacific*, **111**, 657-660.

VAN FRAASSEN, B. (1980). *The Scientific Image*. Oxford: Oxford University Press.

——— (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press.

VORMS, M. (2011). Representing with Imaginary Models: Format Matters. *Studies in the History and Philosophy of Science*, **42**, 287-295.

WEISBERG, M. (2007). Who is a Modeler? *The British Journal for the Philosophy of Science*, **58**, 207-233.

WIMSATT, W. (2007[1987]). False Models as a Means to Truer Theories. In W. Wimsatt (Ed.), *Re-engineering Philosophy for Limited Beings* (pp. 94-132). Cambridge, Mass.: Harvard University Press. (Original article published in 1987.)

WINSBERG, E. (2010). *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press.

ZWICKY, F. (2009[1933]). The Redshift of Extragalactic Nebulae. *General Relativity and Gravitiation*. **41**, 207-224. (Original article published in German in 1933.)